



OPEN ACCESS

EDITED BY

Çagri Erdem,
University of Oslo, Norway

REVIEWED BY

Nádia Moura,
University of Coimbra, Portugal
Georgios Diapoulis,
University of Gothenburg, Sweden

*CORRESPONDENCE

Jenn Kirby
✉ jenn.kirby@liverpool.ac.uk

RECEIVED 03 February 2025

ACCEPTED 19 September 2025

PUBLISHED 05 November 2025

CITATION

Kirby J (2025) Exploring pose estimation in
instrumental composition: the Body
Fragmented project.
Front. Comput. Sci. 7:1570296.
doi: 10.3389/fcomp.2025.1570296

COPYRIGHT

© 2025 Kirby. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Exploring pose estimation in instrumental composition: the Body Fragmented project

Jenn Kirby*

Department of Music, University of Liverpool, Liverpool, United Kingdom

Introduction: In contemporary and experimental music composition, the integration of physical movement via sensors and digital technology offers new pathways for music composition, performance and interdisciplinary practice. This research explores how pose estimation can be used to generate movement-based notation and support embodied musical expression.

Methods: Through the Body Fragmented composition project, this study introduces a methodology that centres the composer and performer's bodies in instrumental composition using pose estimation technology. The approach supports non-linear collaborative processes between composer and performer, facilitating a movement-led instrumental composition practice. A guiding question "what does that movement express and how could that sound" provided a conceptual anchor and encouraged an openness to the interpretation of the pose estimation outputs, allowing it to suggest sonic material and musical expression.

Results: Findings reveal that pose estimation can effectively capture expressive movement for compositional development, support performer interpretation through visual scores, and enhance collaborative dialogue. The study also identifies limitations in pose estimation's ability to convey nuanced musical gestures, and detailed musical information, prompting the integration of supplementary notation materials.

Discussion: The reflective methodology enabled the collaboration to explore new forms of movement-led methodologies in instrumental composition.

KEYWORDS

pose estimation, instrumental composition, music performance, motion capture, embodied music interaction, human-computer interaction

1 Introduction

In contemporary and experimental composition, the integration of physical movement via sensors and other digital technology has emerged as a transformative practise offering new avenues for music composition, performance and interdisciplinary creative practises, with established practises within digital musical instrument (DMI) and human-computer interaction (HCI) practises (Tanaka, 2011; Erdem et al., 2020; Frid, 2019; Green, 2014; Paredes et al., 2022). Developing in tandem with sensor technologies, HCI, DMI, and Computer Music communities have engaged in motion capture for composition and performance (Bevilacqua et al., 2001; Collins et al., 2010; Bazoge et al., 2019; Wanderley, 2022). Indeed, the boundaries between technology, instrument, body, composition have long been morphed culturally and technically (Harraway, 1991; Tanaka and Donnarumma, 2019; Hsu and Kemper, 2019; Baalman, 2017; Waters, 2021; Richards, 2006). Embodied music interaction (Xambó et al., 2017) often engages with data capture of body-worn sensors as input into a system, with a focus on the body forming part of this input system. There is a long history of the integration

of sensors in experimental composition and performance practises (Medeiros and Wanderley, 2014), capturing gesture of instrumental movement (Kanga, 2016), and extending instruments and instrumental practise (Emmerson, 2011). This integration of sensor, wearable and acoustic instrument allows for a dynamic gestural symbiotic relationship between and an assemblage of the instrument's sound, the processed/manipulated sound, and the performer's body.

For instrumental composition, these wearables have provided ways of extending the traditional instrument through live electronics. Development in instrumental composition have largely focused on augmentation of traditional instrument (Kimura, 2012) and expanding timbral capabilities (Yang and Essl, 2012; Boutard, 2016; Nichols et al., 2025).

Pose estimation models (PEMs), such as OpenPose, PoseNet, and BlazePose, which offer non-invasive, markerless, real-time analysis of body movement, have gained traction in interdisciplinary performance contexts due to their accessibility and expressive potential. Their use in instrumental composition as a notational and collaborative tool is yet to be fully explored. This article aims to address that gap by investigating how pose estimation (PE) technology can support a movement-led approach to instrumental composition and facilitate the emergence of sonic and performance possibilities directly from body movement. Central to this enquiry is the question revisited throughout the process: 'what does that movement express and how could that sound?' This enquiry frames a methodology that centres the composer and performer's bodies in an embodied compositional process, using PE.

This methodology is presented through a case study, the Body Fragmented composition project, commissioned by Larissa O'Grady, which follows a non-linear process that embraces the unpredictable, iterative and reflective nature of creative practise, over fixed linear progressions or outcomes. Redhead (2017) describes a similar non-linearity in creative practise, highlighting the overlapping processes of composing, notating, performing, and listening. Mainsbridge and Beilharz's (2014) body-centric approach supports an evolving process where ideas emerge from bodily experiences and sensations. In this project, starting points are fluid and evolve with ongoing reflection and dialogue generating multiple drafts and versions.

An observational reflective practise, grounded in notetaking, artefact analysis, and embodied experience, affords a way to set aside work that does not serve the current enquiry without devaluing that work. This approach allows for the temporary or permanent discarding of ideas and materials, supporting multiple entry points into the creative process, and maintains openness to the interpretation of material as it emerges.

The aims of the project are to (1) explore embodied methods of composing and performing instrumental composition by utilising PE technology as a means of representing and investigating musical material; (2) implement a system using PE to support the compositional, notation, and collaborative processes; (3) compose a piece for violin and electronics in collaboration with violinist Larissa O'Grady; (4) reflect on the efficacy of the methods employed whilst iteratively developing the composition.

This research adopts an autoethnographic methodology, drawing on reflective practise, performer interviews, and collaborative experimentation. By embedding PE into the creative workflow, this study proposes a framework for movement-led instrumental composition that may be of interest to composers, performers, and

researchers working at the intersection of embodiment, musical expression, technologies for music notation, HCI, motion capture, sensors-based performance and instrumental composition.

To contextualise this approach, the following section outlines the composer's movement-led composition practise, reviews literature on relevant notational strategies and expressive needs, and examines the role of motion capture, and pose estimation technologies within creative practises.

2 Background

Movement-based composition practises have often emerged through collaborative intersections between music and dance. For example, Morales-Manzanares et al. (2001) developed a sensor-based composition system that integrated the expressive qualities of both disciplines to generate music in real-time using compositional grammar rules.

Camera-based motion capture has played a significant role in experimental music practises. Whilst marker-based approaches are often regarded the gold standard for accuracy (Das et al., 2023), their practical application in live performance settings are limited. Recent advancements in PEMs have supported body-oriented performance practises, particularly in interdisciplinary contexts. Within human movement sciences, PE approaches are valued for their non-invasive, cost-effective qualities (Roggio et al., 2024). These qualities are also relevant and applicable to artistic practises. Within dance, Nogueira et al. (2024) highlight how machine learning-based PE acted as a catalyst for creative innovation, whilst also acknowledging limitations and ethical concerns. Lim et al. (2022) made use of hand-tracking via an ML-based PE model for the creation of two browser-based instruments with real-time control over sound parameters. Nichols et al.'s (2025) extensive overview of music composition and performance that incorporate motion capture, shows a wide range of approaches within experimental music practise. Caramiaux and Tanaka (2013) showed ML techniques as becoming popular within the New Instruments for Musical Expression (NIME) community in 2013. An analysis 10-years later analysing the use of ML in NIME shows a continuation of this trend (Jourdan and Caramiaux, 2023). Jourdan and Caramiaux highlight that the "majority of the literature considers ML as a medium, focusing on creating processes where the system adapts and reacts to the sounds, gestures, and other ways of interacting of the user" (Jourdan and Caramiaux, 2023).

2.1 Movement-led composition in electronic music

The rationale for the Body Fragmented project grew out of identifying specific creative challenges in applying the movement-led practise of the composer's electronic composition to an instrumentation composition practises. This electronic music practise integrates body movement with digital musical instruments using sensors, which capture body movement and the resulting movement data is used in the interaction between body and digital musical instruments. This workflow considers the desired input (movement) first, rather than output (sound), integrating soma design methods (Martinez Avila et al., 2020). This process prompts critical questions about the

affordances of the digital instruments, gesture, and expression such as, how do I want to move with this device, what does that movement express, and how could that sound? An enquiry into how movement can shape sound, coupled with an embodied practise, allows the sonic opportunities to emerge from the movement. Focusing on designing with the body enables a holistic embodied process, extending the body to incorporate tools and become part of a compositional system, where the necessary hardware and software components are created as part of a compositional process. The creation and integration of these components inevitably shape the composition, becoming intentionally “entangled” (Waters, 2021). The compositional outputs from this methodology often maintain a flexible state, allowing all integrated components to contribute to and extend the work. This approach results in a system that is both a creative work and an extendable, fluid framework for ongoing artistic exploration.

2.2 Gesture and notation approaches and challenges

Though the gestures produced in movement and sound are linked, they can express different characteristics. The disparity in this nuanced relationship between visual and sonic gesture can enhance the composition and live performance by offering divergent interpretations and translations. The disparity provides opportunity for non-literal “action-sound coupling” (Jensenius, 2022) and mappings of gesture to sound, allowing for greater interpretative freedom and agency for performers. This non-literal coupling is also important because different movements on an acoustic instrument can produce similar sounds, and similar movements can produce different sounds, therefore, it may be reductive to heavily map gestural movements with sonic gestures.

This process facilitates exploration by embracing the fluid and non-linearity of creative development. The movement-led approach effectively supports non-linearity by making it possible to have multiple entry points as interventions into the creative process, e.g., pursuing a sonic idea, a movement idea, or an unexpected gesture in the data. The Body Fragmented project seeks to address the creative challenge of developing a methodology and a system for enabling movement-led instrumental composition, similar to the movement-led electronic composition process outlined. A movement-led approach will inform the composer and performer collaboration and should therefore also support a means of representing and communicating the composition, as some form of notation.

Western staff notation primarily depicts the desired sonic output, leaving the performer to determine the necessary movements. The performer may determine the technique required, performer input is sometimes defined in the notation with extended techniques, for example in the use of multiphonics both desired pitches and fingerings are often described (Stone, 1980, pp. 194). Performer input is not typically indicated for many reasons, not least of all because the performer will likely have significantly more knowledge of the instrument, and in performance, the performer and the instrument are inseparable, with the sound and performance as the product of that relationship. For example, with multiphonics, the fingerings may differ between different performers and their instruments. Regardless, it is the sonic output, not the movement specified, which is the priority. Notation practises in experimental music including text-based, and graphic

scoring continue to evolve (Sauer, 2009; Kojs, 2011). Many experimental music approaches consider notation as part of the creative process (Vear, 2019), and how the acts of notating, composing and performing can be intertwined as embodied practises (Redhead, 2022). Contemporary technology-based notation methods have allowed for interactive approaches, including real-time scores (Kim-Boyle, 2010) and animated scores (Hope, 2017), offering new and alternative ways of expressing musical intentions, and informing the compositional process. Notation systems used by artists are not passive. Whether intentional or not, the tools influence the creative process and resulting output.

Rebello, in considering notation as document, states that the “notated artefact is deconstructed, with its discrete elements being re-contextualised in order to derive meaning” (2010). In this deconstruction process, the movement and expression of ideas can become lost. Notation methods which prioritise specific musical elements, such as pitch, can create a hierarchy and lead to a de-prioritisation of other musical parameters. The process of notating discrete elements results in a deconstruction of the composition, which is to be reconstructed by a musician. A movement-led compositional method requires an alternative approach to the deconstruction of musical parameters by focusing on movement expression of the compositional ideas. There are many dance and movement-based notation systems, notably Labanotation, that depict movement and effort (von Laban and Lawrence, 1974) which considers the expressive attributes of movement. Labanotation with its steep learning curve may not be an intuitive method for composers and instrumentalists, however the concepts around movement qualities and effort are pertinent. Silang Maranan et al. (2014) “believe that because [movement qualities] reveals movement expressiveness, their use has strong potential for movement-based interaction with applications in the arts.” In addition, Hope (2020) argues that “animated notations may guide the facilitation of the body as an instrument in real time. They can engage and determine movement in and through space and time – an area yet to be fully explored.”

The aim through this project is to develop a methodology which can both capture the musical expression of the body and compositional ideas, support collaborative and non-linear approaches, and allow for flexibility of performer’s interpretation. Notation as a creative tool intertwined with composing and performing provides a means to codify and communicate, rather than solely to document. Therefore, the requirements of the system are to facilitate the exploration (not just a depiction) of movement-led compositional processes and communicate (notate) the associated concepts and material between composer and performer. Many composers and performers collaborate during the compositional process, and this collaboration can take many forms (Taylor, 2016; Aslan and Lloyd, 2016). Hope stresses that “[a] scoring format that enables varying degree of openness and acknowledges the contributions of performers – electronic and acoustic – is desperately needed in contemporary practise” (Hope, 2020).

2.3 Motion capture in music composition and performance

Motion capture may provide a useful means to facilitate composer-performer collaboration in a new notation format. The term motion capture or MoCap often refers a method using suits and IR cameras in a specialised environment where the mover has body markers attached

to their suit. This method can capture subtle and nuanced movements and is widely used in animation. Markerless methods for capturing body movement often use three common methods:

- 1 Inertial sensors to track movement of different points on the body (Santos et al., 2021; Nichols et al., 2025)
- 2 Camera-based tracking which could use a range of methods such as blob tracking (Sivarathinabala and Abirami, 2014), frame differencing (Kramer et al., 2012, pp. 75) and machine learning methods which typically use neural networks to analyse input from multiple cameras to reconstruct a 3D representation of motion (Ray et al., 2024).
- 3 Multimodal approaches often combine position and orientation, for example through integrating camera- and sensor-based analysis methods (Medeiros and Wanderley, 2014; Nichols et al., 2025). These can produce accurate spatial positioning and keypoint orientation, however many challenges are still present especially in representing highly nuanced motion (Jin et al., 2024) in the different settings in which the motion capture might occur (Jensenius, 2018).

Whilst there is extensive use of internal biosignal sensing as well as external movement sensing technologies and strategies, there are still many practical barriers restricting the use of motion capture by composers and performers. Alongside DIY approaches, there are many commercial applications and combined approaches. Chen et al. (2024) highlight that motion capture technologies used in film, gaming, sports and medical industries come at a high cost which make them inaccessible to consumers and so propose a consumer-affordable multimodal approach which combines Inertial Measurement Units (IMUs) and computer vision (CV). This multimodal approach is also used in the entertainment industries applications, for example Movella, Xsens MVN (n.d.) and Movella, Xsens Sirius (n.d.) focus on combining IMU-style approaches with 3D positional tracking solutions such as HTC Vive and GPS. Paired solutions such as Xsens and HTC Vive can provide highly precise mocap data, addressing occlusion issues which can occur in camera-only solutions. For instrumental composition the occlusion issues are likely to be encountered by self-occlusion, from turning the body, and object-occlusion, from the instrument blocking parts of the body (Traver et al., 2017; Jürgens et al., 2020). The cost and environmental setup make these combined solutions less accessible for a composer to add to their workflow, and for a performer to engage with on a per-piece basis.

Motion tracking via sensors, including IMUs, as discussed, are widely used in composition and performance. Attached to different points on the body, IMUs are ideal for measuring orientation in physical space, but alone are not capable of providing physical positioning (Jensenius, 2018). When using IMUs it is necessary to determine which points on the body would be most appropriate for motion capture, for example does the wrist and forearm need separate tracking, does head position and distances between limbs need tracking. Adding a lot of sensors to gather more data and later filter to retrieve the required data could be a useful strategy. However, this adds unnecessary cost, introduces significant redundancy which could introduce other unknown issues resulting from a more complex system (e.g., latency and additional setup time), and the wearable nature of the sensors may alter the performer's range of movement. Practical challenges such as time charging batteries and costs of components also increase.

2.4 Pose estimation in music performance

PEMs offer an alternative to bespoke hardware solutions and afford a non-invasive and markerless approach to the analysis of body movement. Their use in music has largely focused on DMI creation and movement analysis in musical performance. The research utilising PEMs for music is often built upon an interest in extending sonic possibilities through intrinsic movement, an engagement with embodied musical interaction, an awareness of musical expression within movement, and the audio-visual context in which we perceive much of musical performance. Bodily movements communicate musical intention and clarify “musical structural features” (Davidson, 2012) in classical instrumental performance. Dahl and Friberg (2007) in their study on the visual perception of expression in musician's movement found that viewers could identify emotional intentions just from body movement. Determining which movement cues (head, torso, etc) are strongest can be emotion and instrument specific. Importantly, it is not just the sound-producing movement which communicates expression, but also the ancillary movement (Goebel et al., 2014).

Lim et al. (2022) made use of MediaPipe to track hand movement to control audio processing and produce MIDI data in real-time for a browser-based instrument. Brown et al.'s (2021) work, however, focuses not on the individual movement, but the interaction of movement between two bodies, where virtual touch produces sound when the positional tracking from both bodies intersect. Smith's (2022) use of PoseNet for generative music produces both a sonic output from the movement and a visualisation tracing the movement analysis to offer a motion description. MediaPipe Pose was introduced in 2019 (Roggio et al., 2024) and has since received many upgrades to new solutions including Face landmark detection, Pose landmark detection and Holistic landmarks detection (Google AI, 2025). It has been used by Moussas et al. (2024) to track face and hand movement of a vocal performer and map it to various audio effects. Tobita and Mima (2024) also made use of MediaPipe, combining it with audio analysis methods to timestamp a scored performance. Their evaluation of a variety of methods showed MediaPipe to be an appropriate solution for the analysis of musical performance. However, they noted that even with the audio and movement data, the analysis may not be sufficient to capture all aspects of the variety and nuance of instrumental performance. For example, with analysing wind instruments where there is a need to distinguish the breath before a note from the note produced, the PEM and audio analysis were not able to differentiate the two events. Jin et al. (2024) also highlight the challenge of capturing “domain-specific actions” and nuanced movement in string performance. This demonstrates that these PEMs and combined approaches as useful for the analysis of musical performance and embodied movement, however it also highlights some limitations of current PEMs due to nuanced movement in musical expression. These limitations and nuances may only come to light when encountered by the performer or composer.

3 Materials and methods

With these considerations in mind, the following requirements and guidelines are proposed:

- 1 The solution must be usable in non-specialised environments and ideally can be used in a living space or home studio.

- 2 Use easily attainable low-cost hardware, and/or use personal hardware such as a smartphone or computer.
- 3 The workflow of the system should be intuitive and not introduce a considerable learning curve for the composer or performer.
- 4 Setup time should be minimal, so to not reduce composition time or collaboration time.
- 5 Design of the system should be artist-driven, where changes are incrementally implemented from a need uncovered in the process.
- 6 The system should facilitate the exploration of movement-led compositional processes.
- 7 The system should facilitate the communication of the composition to the performer and facilitate further collaboration between composer and performer.

In *Body Fragmented*, a PEM is employed to create movement-based notation. This workflow involves body tracking of musical expression representing compositional intentions rendered in video form for interpretation by a human performer.

3.1 Workflow and composer-performer collaboration

The stages of the process include movement capture, score generation, and iterative refinement of materials through collaboration (Figure 1).

The workflow is as follows:

- 1 The composer records a video that depicts body movement, such as through playing an instrument or expressing musical gestures.
- 2 The composer feeds the PEM with this video, which renders a figure of the movement. This output serves as a form of notation that encodes the expressive movement data of the composition.
- 3 The performer interprets the video as a score.

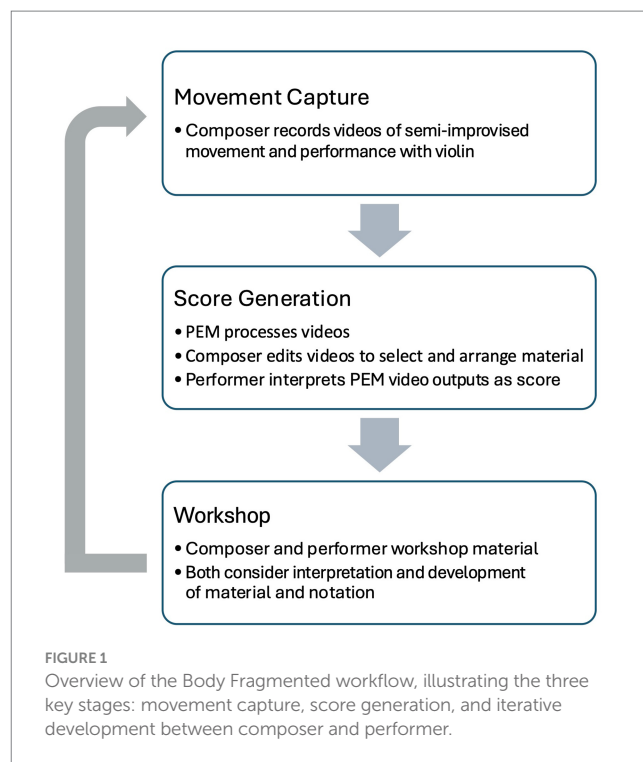
This process established a framework for collaboration and interpretation between the composer and the performer that could be repeated throughout the development stage.

3.1.1 Version 1 – initial experiments and PEM video outputs

The composer wrote text instructions for 5 sections as material to improvise with for the recordings. These instructions included: “tapping body,” “small figures, fragmented, tiny movements growing larger,” “plucking holding at waist,” “disengage and re-engage in a new way,” and “as slow as you can bow.”

The composer recorded 3 videos of themselves responding to these instructions by moving, interacting with and playing the violin. Audio was also captured. The movement ranged from slow, subtle movements with the bow on the violin, to full body movements including changing positions of the violin and body from standing to kneeling. The movements produced included:

- 1 Movements intrinsic to the sound-making actions



- 2 Movements intended to reinforce musical gesture
- 3 Non-sound-producing gestures which aim express musical ideas and influence “emotional expression, timing, musical structure and audience perception” (Mainsbridge and Beilharz, 2014).

The videos were edited with sections selected compositionally arranged to be fed into the PEM. The selection was guided by the question ‘what does that movement express and how could that sound?’, where the movement was deemed to be expressive and suggest many sonic possibilities. The audio was used to create an audio-reactive geometric shape which accompanied the rendered video (Figure 2; Supplementary material version 1 example).

After the material was shared with the performer, and they had time to reflect on it, the composer and performer met online to discuss and workshop the material. Through a dialogic reflexivity the composer and performer explored individual interpretations considering what the movement expresses, what sonic material it evokes, how it felt for the composer to record the movement, how the performer experiences responding to it, and the implications for the composition and performance outcome.

Initial responses from the performer indicated that they could discern musical content from the rendering. The discussion centred on interpreting the PEM videos, exploring values approaches to engaging with it, such as treating it as a duet, or imitation, and considering what additional musical context or notation might support the videos. The audio-reactive geometric shape received little discussion, suggesting it contributed minimally to score information or interpretive value.

3.1.2 Version 2 – refinement and interpretation

Following the same process, five new videos were rendered, this time excluding the audio-reactive geometric shape. This reflects the

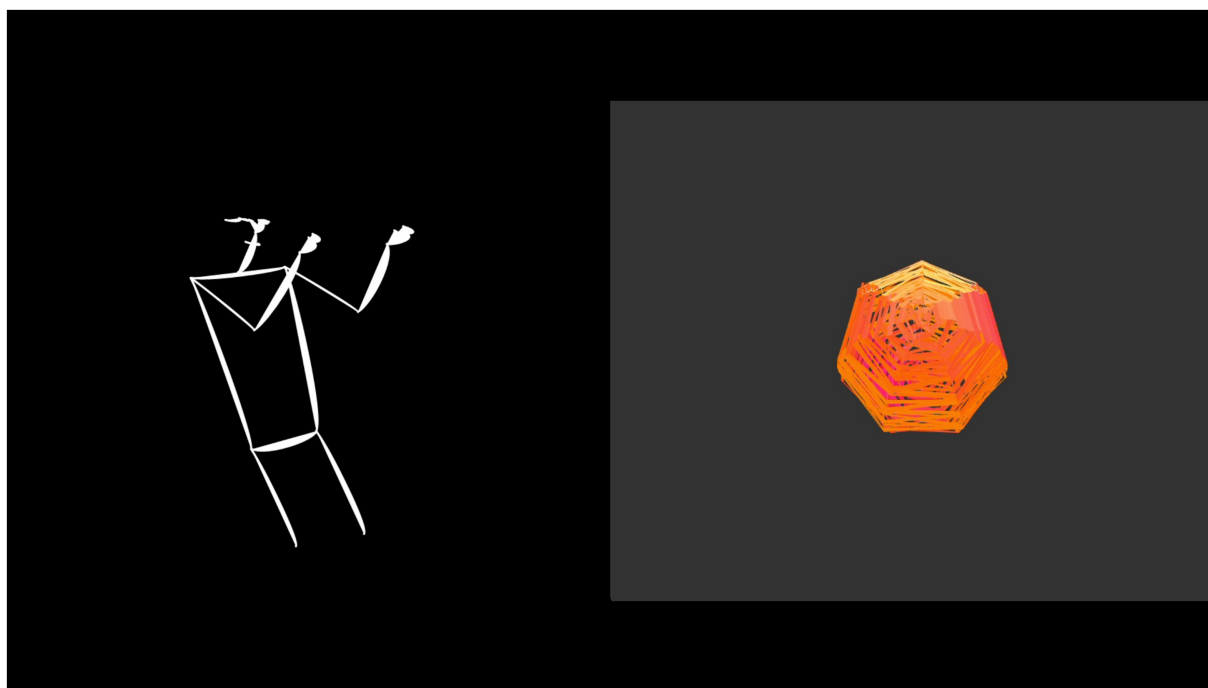


FIGURE 2
PEM figure and audio-reactive geometric shape still from video.

non-linear working method, which allows for discarding of elements that no longer serve the current enquiry, with the understanding that they can be revisited if they become relevant again.

These new renderings exhibited visual artefacts, such as jitter and inaccuracies in movement representation. For example, an estimation error presented as a sudden jerking motion at the left elbow appeared due to occlusion, even though no such movement occurred in the original video (see [Supplementary material version 2](#) example). We discussed how to interpret these anomalies, whether to treat them as errors and disregard or consider as integral features of the rendering. We decided that any information present in the score, including artefacts of the PEM process, can be interpreted and potentially realised in the performance.

The long form videos require the performer to follow the material in real-time without the ability to preview upcoming content. This constraint contrasts with standard score-reading practises, where performers read beyond where they are playing ([Perra et al., 2021](#)). The inability to do so in video playback may limit the performers interpretive flexibility and agency.

3.1.3 Version 3 – performance implementation

To enhance performer agency and interpretive flexibility, the videos were divided into 17 shorter clips, allowing the performer to progress the piece at their own pace (see [Supplementary material version 3](#) examples). These clips were embedded within a Max patch, with each segment linked to corresponding audio processing preset parameters. A footswitch was used to advance the clips and simultaneously update the patch settings. In addition, a preview of the upcoming video was displayed

in a smaller window, allowing the performer to look ahead, restoring a key aspect of traditional score-reading practises.

IMUs are not used to capture movement for the notation. However, an IMU is used in the performance of the work. The [MUGIC Motion \(2025\)](#) sensor is worn by the performer in performance to track movement of the right (bowing) arm and wrist in real-time. This data is integrated with the sound processing shaping the manipulation and processing of the live violin. In some instances, the data is mapped directly, in other cases the movement is analysed to determine its characteristics and therefore the data is more indirectly mapped to the process. This indirect mapping can allow for greater expression and autonomy for the performer by not restricting their movement. For example, where a period of minimal movement or stillness is identified by the IMU, this increases the reverb size, which allows the performer to build large sounds through very subtle movements.

3.2 Pose estimation methodology

Various PEMs can be used for the video analysis. These models use Convolutional Neural Networks (CNNs) to extract features and classify the image being provided ([Roggio et al., 2024](#)). The frame-by-frame output provides the user with positional information for anatomical landmarks (keypoints) which can be used to gain insight about the movement in real-time. Initially, PoseNet was used but the results exhibited too much jitter to be deemed usable. BlazePose from MediaPipe, which is trained on sports movements and produces 33 body keypoints ([Bazarevsky et al., 2020](#)), provided significantly better accuracy for capturing musicians' movements in home and studio

settings, and so was used for this project. It required less filtering and smoothing, thereby reduced the risk of eliminating expressive qualities.

Our abilities to sense quite accurately both the actual movements and their expressive and emotive features become even more remarkable when we try to replicate these abilities with machines. What's easy for us may be very difficult or even impossible for machine-based systems of vision (Godøy and Leman, 2010).

In this project the expression in movement is just as important as the expression in sound, and this project is premised on the hypothesis that the musical expression of the composition exists within the movement required to perform it. Therefore, the ability to reflect musical expression is prioritised above accuracy of capture. Godøy and Leman (2010) highlight that “gestures are often intended to express, rather than to denote.” In this discussion the use of the term gesture blends to both refer to a movement that expresses musical meaning and denotes musical ideas (not deconstructed musical parameters).

3.2.1 Addressing occlusion and smoothing

Although the model did introduce some movement errors, these were mostly dealt with through refining the filtering process and, in some cases, were retained as part of the model's agency and contribution to the collaboration.

A common issue encountered was with occlusion, particularly when the instrument obscured the performer. Several adjustments were made to improve tracking quality: ensuring that no parts of the body go off-camera, adjusting the performer's angle to the camera to reduce self-occlusion, and using contrasting colours between the instrument and clothing.

Several rule-based methods were applied to the PEM output for temporal smoothing and outlier rejection, based on frame-by-frame comparisons.

- 1 Confidence-based filter: BlazePose provides a confidence score for each keypoint. Only keypoints with a confidence score above 0.4 (40%) are retained. When a keypoint's confidence fell below this threshold, the previous frame's value was used instead. This helps remove unreliable keypoint detections whilst maintaining continuity during temporary occlusions.
- 2 Frame-by-frame averaging: a two-point moving average is used for temporal smoothing between frames. Although this introduces a slight temporal lag, it is not problematic in this context as the original video is not displayed.
- 3 Outlier rejection: a distance thresholding is used to filter out implausible keypoint movements, those deemed physically impossible for the body or keypoint to move between those two points within consecutive frames. The threshold was determined through trial and error, calibrated to the composer's movement and interaction with the violin.

For flexibility and accuracy, these thresholds would need to be customised to each instrument as the speed and range of movement would differ significantly. It may even be necessary for the thresholds to be set for each individual to respond to their movement characteristics.

4 Results and discussion

The composition, commissioned and performed by violinist Larissa O'Grady, was presented in a live performance with a Max patch with embedded videos of the PE figure depicting the movement as score for each section. The performer could see a small video within the patch or fullscreen rendered and used a footswitch to progress the videos and other electronics processing. A video of the performance is available from [Hugh Lane Gallery \(2023\)](#).

Each clip looped continuously until the performer advanced to the next via the footswitch. The video display was also visible to the audience, allowing them to observe the relationship between the figure and the performer. At times, the movement produced very quiet sounds, in some instances the gestures were interpreted directly, but may differ in performance qualities, such as being smoother, more dramatic, or more expressive. Some interpretations appeared to extend or evolve from the movement. Movements such as jerking did not always generate sounds directly but instead conveyed expressive qualities such as tension ([Figure 3](#)).

One particular clip depicting both arms moving slowly and in circles followed by increased movement through the body of the figure resulted in circular bowing for string crossings. The section is fluid and softer. The video depicted the left arm slowly moving towards the head and the right arm drawing downwards from the body. Although no violin is visible in the video renderings, the relational movement of the arms strongly implies its presence.

The looping of the short duration clips led to repetition in the material, and a non-linear structure. Whilst the piece unfolds as a sequence of musical events led by the video clips, the structure is non-linear as the substructures resist directionality and continuity ([Vickery, 2011](#)), instead presenting as fragmented moments from which thematic material may emerge.

A subsequent performance of the work was presented without the PE videos. In this iteration, a graphic score was used in conjunction with the Max patch. A video of this version is available from [CMC Ireland \(2024\)](#). There was significant reflection and composer-performer collaboration and development time in between the two versions.

4.1 Interpreting movement and reflections on the guiding question

The guiding question, ‘what does that movement express and how could that sound?’, was not intended to elicit a fixed sonic response, but instead to serve as a conceptual anchor throughout the project. It invited an open exploration of movement, and its potential to express musical ideas. This framing, alongside the PEM renderings, encouraged both composer and performer to carefully consider the nuances of gesture, posture, and motion, and to consider how these might suggest or evoke sonic material, rather than prescribe it.

In practise, this question shaped the selection of material from the PEM renderings. For example, in version 1, non-sound-producing movements were retained for their expressive intent and non-obvious sonic possibility, such as shifts in body weight, and atypical postures. The question also encouraged an openness to interpreting movement which had more obvious interpretations available, such as circular movements in the right arm. The performer's responses varied: some

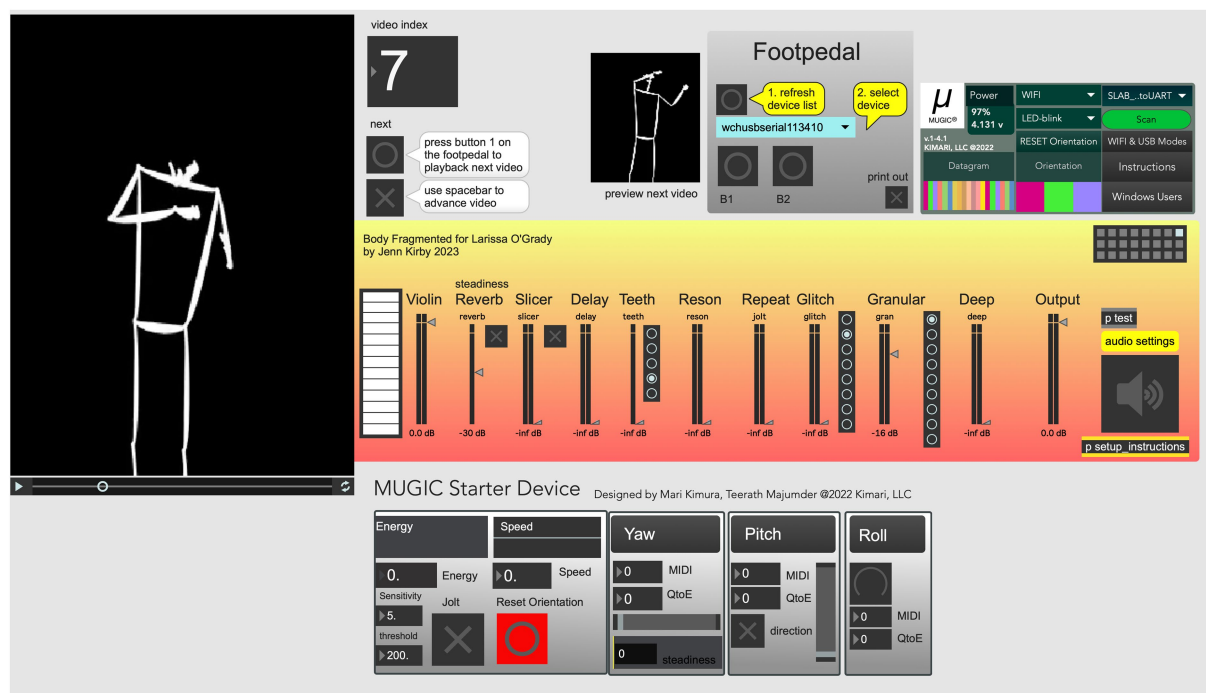


FIGURE 3
Max patch with embedded videos.

gestures were mirrored in bowing direction or articulation, whilst others inspired contrasting textures, and movement responses, revealing an interpretive relationship which can respond to a variety of aspects within the movement to elicit musical intention. This interpretive freedom shows the use of pose estimation in this project not as a prescriptive tool, but as a medium for embodied musical dialogue.

The following discussion outlines the collaborative process, the challenges and opportunities which arose during the process, and the role and impact of PE in the composition and performance outcomes.

4.2 Performer reflections

The performer's reflection is based on an analysis of interview responses to questions provided by the composer, with direct quotations drawn from these responses.

The performer expressed that the animated format initially provided an immediate connection but raised questions regarding how the performer should or could respond to the movements presented. The performer considered three potential approaches:

- 1 Exact replication: Should they mimic the movements precisely?
- 2 Personalised variation: Should they echo the movements whilst incorporating their own style in a follow/lead pattern?
- 3 Improvised duet: Should they allow the movements to inform their response, akin to interacting with a human in an improvisational setting?

The performer opted for the third approach, engaging with the video score as if in an improvisational duet with another performer, which she explained "felt more natural."

The PE prompted the performer to adopt an experimental approach to movements, that might otherwise be considered "poor posture, bad technique or untraditional." Embracing this experimental approach to movements created "space for the inclusion of new to me sounds" generated from her violin and bow movements. The performer described the experience as one that necessitated a thorough interrogation of various elements, such as "purpose/style/movement range/allowances of freedom." She noted that working with the moving figure, the score felt more deeply connected to the composer than if the movement had used abstract shapes. This could be due to the figure containing substantial form information and joint trajectories (Giese et al., 2008) as motion displays closer to human form exhibit more expressiveness (Moura et al., 2023). This connection to the figure led the performer to want to seek the composer's permission regarding the figure's movements.

This signifies a distinct difference in a relationship with the figure in the videos compared to traditional notated or graphic scores. Feeling a more immediate connection to the figure, facilitated a "call and response" interaction. However, the relationship the performer formed with the figure led to a sense of "duty and loyalty" that made her feel less free to "change tack radically during the performance." Instead, she realised that she was subconsciously tied to repeating what was practised during rehearsals. The fixed videos format restricted some performance possibilities, because the interaction was one way, and the figure could not respond to the variation and improvisation from the performer.

The performer's reflections reveal how this embodied composition through PE can alter artistic practises, creating more freedom in some directions and creating restrictions in others. Navigating the complexities of compositional communication, between composer,

performer and score, led to exploring new dimensions of movement and sound.

Through this process, performer and composer engaged in meaningful discussions about the role of movement in the video score. In situations where collaborative dialogue is not feasible, the performer noted that there is potential to enhance the animated character through narrative or direction within the score, before attempting to connect with the figure, thus providing alternative dialogues and interpretations.

4.3 Composer reflections

The composer's reflection is presented as a reflective analysis, drawing on notes, sketches, and video material.

The PE provided a useful and inspiring means of considering body movement in instrumental composition. Importantly, it addressed a key aim of the project, to provide a means of movement-led instrumental composition.

Outlined above is how focusing on body input into electronic composition leads to considering what sound could emerge from movement, what movement expresses, what musical meaning a gesture can hold, and how the output derives from this intuition-led embodied methodology. Working with PE has provided a means of exploring movement-led instrumental composition, shared between composer and performer. The interpretations and understandings of movement can differ greatly, much like the deviations between sonic and physical gesture. These deviations in meaning and interpretation between performer and composer can produce added value to the process. These variations can be viewed as a compositional dialogue and can reveal additional parts of our artistic expression. As is common in a collaborative process, the score incorporates these discussions.

The PE method has been valuable as a means of capturing compositional ideas and movement improvisations. The pose figure representations remove a human physical form and allow for a magnification of macro and micro movement. [Moura et al. \(2023\)](#) in their comparative study found that the perceptual effect through subjective evaluation of body movement in various simplified forms (point-light, stick-figure, body-mass or skeleton) was dependent on the task associated with the movement. They highlight significant differences in simplified and human-like displays, noting that the lack of information for volume and depth in stick-figures “may obscure.

the attribution of expressive properties to the illustrated motion, a task with higher complexity.

levels than mere motion discrimination.” The stick figure from this study and the output from the PEM are not directly relatable however, as the 33 anatomical keypoints from BlazePose enables some depiction of rotational movement. [Nusseck and Wanderley \(2009\)](#) found that larger body movement amplitudes contributed more to the viewer's perception over individual limbs and anatomical keypoints, and musical attributes appeared to be located as holistic characteristics.

The PE was initially intended to provide a means to embody the composition and fully represent the music within movement for reading and interpretation by the performer. However, it was not sufficient to achieve this fully but through discussions in workshops we found what musical information and intention it could express. Discussions about the material between composer and performer are

essential to ensure that meaning is “co-constructed” ([Boutard, 2016](#)), and that a shared understanding and interpretation emerges from this collaboration.

4.4 Reactions, limitations and the usefulness of pose estimation as notation

Upon reviewing the videos outputted by the PEM, the performer expressed that she could discern musical material within them, and could hear music, noting that the figure initially offered a more direct connection. Whilst this initial positive reaction was significant, as we continued with the process we encountered limitations with the output material. As the performer began to interpret the composition, the detail in the videos proved insufficient for deeper extrapolation. Although it was possible to respond to these questions with further improvisation, we worked to determine the degree of improvisation and what intentions are attempted to be conveyed in the video material.

The level of detail required for improvisation is inherently subjective and individually defined among composers, performers, and in collaboration. We sought to understand how effectively the material communicates the composer's intentions and how much freedom the performer has in their interpretation. If compositional intentions are not conveyed clearly, improvisation or interpretation may overshadow them. In some cases, a composition may possess minimal intentions or allow for completely free interpretation by the performer. However, even this openness constitutes an intention that must be communicated or assumed.

In this composition, it became clear through discussions that more intentions existed than what the provided material conveyed. Therefore, additional details and materials were necessary to adequately express these intentions. This realisation prompted discussions about what form these [Supplementary material](#) should take. To avoid deconstructing the material, we opted for some use of Western staff notation with graphic scores and descriptive elements for individual sections, as well as providing recommended pitch material. Interestingly, this approach rendered the video material of PE movement scores unnecessary to follow for the second performance. This observation raises questions about the effectiveness of the movement scores particularly when they can be substituted with graphic and text-based notation. Upon reflection, the primary functions of the video material were to provide the performer with information on the musical expression and intentions of the piece, and to provide the composer with a means of capturing and articulating those intentions.

The graphic score produced derived in part from the video material and in part from the collaborative discussions through engaging with that material. The PE technology facilitated the capture and exploration of compositional ideas and embodied processes. However, the video material of pose movement was less effective as a standalone notation method to be read in live performance for what we wanted to achieve. It partially encapsulated the composer's intentions and facilitated interesting collaborative discussions to allow the composition to fully emerge from the process.

Composers often utilise video materials to demonstrate specific techniques within a notated score. Whilst these materials serve as supplements to the score, they may have been created prior to any notation; thus, it is possible to consider that the notation enhances

the video content rather than vice versa. The sequence in which materials are created plays a crucial role in the compositional process. Similarly, how performers engage with these materials can significantly impact their interpretative process. Further enquiry may be warranted regarding whether the video material output from the PEM provides more detail about the notated content or if vice versa holds true.

With the performer improvising with fixed video material, a notable discrepancy arises. The performer highlighted feeling not being free to make radical changes from performance to performance, due to improvising with a fixed performer who is not responding to these changes. Given the availability of movement data to which the figure is drawn from, there is potential to enhance the systems interactivity, and thereby enable more compositional and performance opportunities. The figure could be programmed to adapt dynamically to the performer's movement, by utilising the IMU sensor to provide real-time data. This could enable the figure to respond interactively to performer movements. This approach would create a more symbiotic relationship between the human improviser and the digital entity and could allow for the deeper exploration through improvisation which the fixed method had restricted.

5 Conclusion

For PEMs to effectively support a movement-led methodology, their integration early in the compositional process is essential. Nonetheless, certain parameters, such as conceptual framing or textual descriptions, can be established in advance to provide interpretive context and constraints. These [Supplementary material](#) can support dialogue and help shape the direction of the work. Importantly, a movement-led approach does not require that all compositional material originate from PEM videos.

Movement tracking via PE can be integrated at various stages of the composition process, offering versatile functions that are valuable to composers and performers interested in movement-led methods. In this project, PEM videos were most effective for generating and capturing ideas and facilitating collaborative discussions, though they were insufficient on their own to fully notate or communicate all aspects of the composition. Ongoing reflexive practise played a crucial role in identifying where additional interventions were needed.

The tools used in creative practise inevitably shape the creative output. In this case, characteristics of the PEM rendering were integrated into the realisation of the work. This methodology encouraged an exploration of the affordances and limitations of PEMs, considering how their unique characteristics may contribute to new forms of movement-led composition. The approach taken here integrated PE with existing notation methods, including graphic scores, text-based scores, and Western staff notation.

The Body Fragmented project demonstrates the potential of PE technology to meaningfully contribute to instrumental composition practises by centring the body as a primary source of musical expression. The collaboration with violinist Larissa O'Grady revealed both the creative opportunities and limitations of using PE as a compositional and notational tool. Whilst the technology facilitated a movement-led approach and provided a novel means of capturing and communicating musical ideas, it also underscored the need for

further refinement to better capture the nuances of musical performance. This aligns with findings by [Jin et al. \(2024\)](#) who noted that current PEMs lack the nuance to depict domain-specific information.

The guiding question posed at the outset, 'what does that movement express and how could that sound?', served as a useful prompt throughout the creative process. It helped frame the project as an open enquiry into the expressive potential of the movement, inviting both the composer and performer to explore how gesture might suggest, rather than dictate sonic material. This question supported a methodology grounded in embodied experimentation, collaborative interpretation, and iterative development.

Initial discussions around expression and interpretation revealed a wide range of possibilities. However, as the process evolved, the videos themselves offered diminishing interpretive value, prompting the creation of additional notation materials. This showed that the PEM videos were more effective as catalysts for dialogue and creative response, rather than prescriptive score. The resulting composition emerged not from a direct mapping of movement to sound, but from a shared exploration of how movement might be felt, understood, and interpreted musically.

This highlights the importance of collaborative dialogue between composer and performer in co-constructing meaning and interpretation. The range of possible interpretations and responses to the PEM videos was broader than anticipated, making shared discussion essential for narrowing interpretive possibilities and arriving at a shared response to the question of "how could that sound." Future work could explore ways to enhance the interactivity of PE systems, enabling more dynamic and responsive performance environments.

This method of working offers promising new pathways for composers whose practises may not be fully supported by existing notation methods. By centring embodied practise, considering the physicality of both composer and performer, and embracing the ambiguity of gesture, this methodology expands the possibilities for movement-led instrumental composition. Much like the innovations introduced through the use of wearable sensors, PEMs enable new modes of expression and collaboration, supporting the development of composition practises emerging from movement and embodied interaction.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

JK: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The Body Fragmented

composition was funded by the Arts Council of Ireland. Open access publication costs were covered by University of Liverpool.

Acknowledgments

The author thanks Larissa O'Grady, who commissioned the composition, for the collaborative project, and her contribution to the article through her interview responses on the methods and collaboration. The author also thanks the Contemporary Music Centre, Ireland, for supporting this project through their Contemporary Artists Network.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. The author acknowledges the use of Perplexity and Copilot, for aiding in text editing and rephrasing during the writing of the paper.

References

- Aslan, J., and Lloyd, E. (2016). Breaking boundaries of role and hierarchy in collaborative music-making. *Contemp. Music. Rev.* 35, 630–647. doi: 10.1080/07494467.2016.1282610
- Baalman, M. A. J. (2017). "Interplay Between Composition, Instrument Design and Performance" in Musical Instruments in the 21st Century. eds. T. Bovermann, A. de Campo, H. Egermann, S. I. Hardjowirogo and S. Weinzierl (Singapore: Springer), 225–241. doi: 10.1007/978-981-10-2951-6_15
- Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T. L., Zhang, F., and Grundmann, M. (2020). BlazePose: on-device real-time body pose tracking. *arXiv*. doi: 10.48550/arXiv.2006.10204
- Bazoge, N., Gagne, R., Nouviale, F., Gouranton, V., and Bossis, B. (2019). Expressive potentials of motion capture in musical performance. In Proceedings of the International Conference on New Interfaces for Musical Expression International Conference on New Interfaces for Musical Expression, Porto Alegre, Brazil, Zenodo. 266–271. doi: 10.5281/zenodo.3672954
- Bevilacqua, F., Naugle, L., and Dobrian, C. (2001). Music control from 3D motion capture of dance. University of California, Irvine. Available online at: https://music.arts.uci.edu/dobrian//motioncapture/SoundControl_MotionCapture.pdf
- Boutard, G. (2016). Solo works of mixed music with live electronics: a qualitative enquiry in timbre and gesture from the performer's perspective. *Musicae Sci.* 20, 361–391. doi: 10.1177/1029864916634421
- Brown, C., Clemmons, M., Greenberg, I., and Brimhall, B. (2021). "Skin hunger" in Proceedings of the International Conference on New Interfaces for Musical Expression, Shanghai, China, 1–6. doi: 10.21428/92fbeb44.2407697b
- Caramiaux, B., and Tanaka, A. (2013). "Machine Learning of Musical Gestures" in Proceedings of the International Conference on New Interfaces for Musical Expression. Seoul, South Korea, Zenodo. 513–518. doi: 10.5281/zenodo.1178490
- Chen, X., Liu, W., Bao, Q., Liu, X., Yang, Q., Dai, R., et al. (2024). Motion capture from inertial and vision sensors. *arXiv*. doi: 10.48550/arXiv.2407.16341
- CMC Ireland. (2024). Larissa O'Grady - 'Body Fragmented' [video]. Youtube. Available online at: <https://www.youtube.com/watch?v=0HKoOAItthw> (Accessed July 5, 2025).
- Collins, N., Kiefer, C., Patoli, Z., and White, M. (2010). Musical exoskeletons: experiments with a motion capture suit. In Proceedings of the International Conference on New Interfaces for Musical Expression. International Conference on New Interfaces for Musical Expression, Sydney, Australia, Zenodo. 455–458. doi: 10.5281/zenodo.1177749
- Dahl, S., and Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music. Percept.* 24, 433–454. doi: 10.1525/mp.2007.24.5.433
- Das, K., de Paula Oliveira, T., and Newell, J. (2023). Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework. *Sci. Rep.* 13:22880. doi: 10.1038/s41598-023-49360-2
- Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: two distinctive case studies. *Psychol. Music* 40, 595–633. doi: 10.1177/0305735612449896
- Emmerson, S. (2011). "Combining the acoustic and the digital: music for instruments and computers or prerecorded sound" in The Oxford handbook of computer music. ed. R. T. Dean (Oxford Academic). doi: 10.1093/oxfordhb/9780199792030.013.0009
- Erdem, Ç., Schia, K. H., and Jensenius, A. R. (2020). Vrengt: a shared body-machine instrument for music-dance performance, in Proceedings of the International Conference on New Interfaces for Musical Expression. International Conference on New Interfaces for Musical Expression, Zenodo. 186–191. doi: 10.5281/zenodo.3672918
- Frid, E. (2019). Accessible digital musical instruments—a review of musical interfaces in inclusive music practice. *Multimodal Technol. Interact.* 3:57. doi: 10.3390/mti3030057
- Giese, M. A., Thornton, I. M., and Edelman, S. (2008). Metrics of the perception of body movement. *J. Vis.* 8:13. doi: 10.1167/8.9.13
- Godøy, R. I., and Leman, M. (Eds.) (2010). *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge. doi: 10.4324/9780203863411
- Goebel, W., Dixon, S., and Schubert, E. (2014). "Quantitative methods: motion analysis, audio analysis, and continuous response techniques" in Expressiveness in music performance: Empirical approaches across styles and cultures. eds. D. Fabian, R. Timmers and E. Schubert (Oxford: Oxford Academic).
- Google AI. (2025). MediaPipe solutions guide. Available online at: <https://ai.google.dev/edge/mediapipe/solutions/guide> (Accessed February 1, 2025).
- Green, O. (2014). "NIME, Musicality and Practice-led Methods" in Proceedings of the International Conference on New Interfaces for Musical Expression (London, United Kingdom, Zenodo: International Conference on New Interfaces for Musical Expression), 1–6. doi: 10.5281/zenodo.1178776
- Harraway, D. (1991). "A cyborg manifesto: science, technology, and socialist-feminism in the late twentieth century" in Simians, cyborgs, and women: the reinvention of nature (New York: Routledge). 149–181.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomp.2025.1570296/full#supplementary-material>

- Hope, C. (2017). Electronic scores for music: the possibilities of animated notation. *Comput. Music. J.* 41, 21–35. doi: 10.1162/comj_a_00427
- Hope, C. (2020). The future is graphic: animated notation for contemporary practice. *Organised Sound* 25, 187–197. doi: 10.1017/S1355771820000096
- Hsu, A., and Kemper, S. (2019). Enacting sonic-cyborg performance through the hybrid body in Teka-Mori and why should our bodies end at the skin? *Leonardo Music J.* 29, 83–87. doi: 10.1162/lmj_a_01069
- Hugh Lane Gallery. (2023). Sundays at noon concert catch-up: Larissa O'Grady, violin [video]. Youtube. Available online at: <https://youtu.be/jNseAwXIOe?si=3wdaHROQoy62kX7Q&t=13> (Accessed July 5, 2025).
- Jensenius, A. R. (2018). "Methods for studying music-related body motion" in *Springer handbook of systematic musicology*. ed. R. Bader (Berlin, Heidelberg: Springer).
- Jensenius, A. R. (2022). *Sound Actions: Conceptualizing Musical Instruments*. Cambridge, MA: MIT Press.
- Jin, Y., Qiu, Z., Shi, Y., Sun, S., Wang, C., Pan, D., et al. (2024). Audio matters too! Enhancing markerless motion capture with audio signals for string performance capture. *ACM Trans. Graph.* 43, 1–10. doi: 10.1145/3658235
- Jourdan, T., and Caramiaux, B. (2023). "Machine Learning for Musical Expression: A Systematic Literature Review" in *Proceedings of the International Conference on New Interfaces for Musical Expression (Mexico City, Mexico, Zenodo: International Conference on New Interfaces for Musical Expression (NIME))*, 319–331. doi: 10.5281/zenodo.11189198
- Jürgens, S., Correia, N. N., and Masu, R. (2020). Designing glitch procedures and visualisation workflows for markerless live motion capture of contemporary dance [paper presentation]. 7th International Conference on Movement and Computing, Jersey City/Virtual, NJ, United States: Association for Computing Machinery. doi: 10.1145/3401956.3404225
- Kanga, Z. (2016). Gesture-technology interactions in contemporary music. *Contemp. Music. Rev.* 35, 375–378. doi: 10.1080/07494467.2016.1258104
- Kim-Boyle, D. (2010). Real-time score generation for extensible open forms. *Contemp. Music. Rev.* 29, 3–15. doi: 10.1080/07494467.2010.509588
- Kimura, M. (2012). "Extracting Human Expression For Interactive Composition with the Augmented Violin" in *Proceedings of the International Conference on New Interfaces for Musical Expression (Michigan, USA, Zenodo: International Conference on New Interfaces for Musical Expression)*. doi: 10.5281/zenodo.1178305
- Kojs, J. (2011). Notating action-based music. *Leonardo Music J.* 21, 65–72. doi: 10.1162/LMJ_a_00063
- Kramer, J., Parker, M., Castro, D., Burrus, N., and Echtler, F. (2012). *Hacking the Kinect*. Berkeley, CA: Apress.
- Lim, M., Kotsani, N., and Hartono, P. (2022). "Handmate: An Accessible Browser-based Controller for Web Audio and Midi using AI Hand-Tracking" in *International Conference on New Interfaces for Musical Expression (Auckland, New Zealand)*. doi: 10.21428/92fbeb44.40c13869
- Mainsbridge, M., and Beilharz, K. (2014). "Body As Instrument: Performing with Gestural Interfaces" in *Proceedings of the International Conference on New Interfaces for Musical Expression (London, United Kingdom, Zenodo: International Conference on New Interfaces for Musical Expression)*, 110–113. doi: 10.5281/zenodo.1178859
- Martinez Avila, Juan P., Tsaknaki, Vasiliki, Karpashevich, Pavel, Windlin, Charles, Valenti, Niklas, Höök, Kristina, et al. (2020). Soma design for NIME. *Proceedings of the International Conference on New Interfaces for Musical Expression*. International Conference on New Interfaces for Musical Expression, Birmingham, United Kingdom, Zenodo, 489–494. doi: 10.5281/zenodo.4813491
- Medeiros, C. B., and Wanderley, M. M. (2014). A comprehensive review of sensors and instrumentation methods in devices for musical expression. *Sensors* 14, 13556–13591. doi: 10.3390/s140813556
- Morales-Manzanares, R., Morales, E. F., Dannenberg, R., and Berger, J. (2001). Sicib: an interactive music composition system using body movements. *Comput. Music. J.* 25, 25–36. doi: 10.1162/014892601750302561
- Moura, N., Fonseca, P., Goethel, M., Oliveira-Silva, P., Vilas-Boas, J. P., and Serra, S. (2023). The impact of visual display of human motion on observers' perception of music performance. *PLoS One* 18:e0281755. doi: 10.1371/journal.pone.0281755
- Moussas, T., Kotsani, N., and Georgaki, A. (2024). "Performing voices in immersive theater: kinesthesia 2.0 for embodied performance" in *Proceedings of the 21st Sound and Music Computing Conference (Portugal: Porto)*, 40–43.
- Movella, Xsens MVN. (2024). Xsens MVN animate. Available online at: <https://www.movella.com/products/motion.capture/xsens-mvn-animate> (Accessed December 4, 2025).
- Movella, Xsens Sirius. (2024). Xsens Sirius Series. Available online at: <https://www.movella.com/products/sensor-modules/xsens-sirius-series> (Accessed December 4, 2025).
- MUGIC Motion (2025). MUGIC, Revolutionize your creativity with advanced motion sensor Tec. Available online at: <http://www.mugicmotion.com> (Accessed July 5, 2025).
- Nichols, C. (2025). "Motion Capture for Musical Expression" in *Collaboration, Engagement, and Tradition in Contemporary and Electronic Music*. eds. D. Payling, M. Estibeiro and D. Cotter. 1st ed. (Routledge), 337–366. doi: 10.4324/9781003430360-19
- Nogueira, M. R., Menezes, P., and Maças de Carvalho, J. (2024). Exploring the impact of machine learning on dance performance: a systematic review. *Int. J. Perform. Arts Digit. Media* 20, 60–109. doi: 10.1080/14794713.2024.2338927
- Nusseck, M., and Wanderley, M. M. (2009). Music and motion—how music-related ancillary body movements contribute to the experience of music. *Music. Percept.* 26, 335–353. doi: 10.1525/mp.2009.26.4.335
- Paredes, V., Françoise, J., and Bevilacqua, F. (2022). Entangling practice with artistic and educational aims: Interviews on technology-based movement-sound interactions: NIME. doi: 10.21428/92fbeb44.5b9ac5ba
- Perra, J., Poulin-Charronnat, B., Baccino, T., and Draï-Zerbib, V. (2021). Review on eye-hand span in sight-reading of music. *J. Eye Mov. Res.* 14. doi: 10.16910/jemr.14.4.4
- Ray, L. S. S., Zhou, B., Suh, S., and Lukowicz, P. (2024). A comprehensive evaluation of marker-based, markerless methods for loose garment scenarios in varying camera configurations. *Front. Comput. Sci.* 6:9925. doi: 10.3389/fcomp.2024.1379925
- Redhead, L. (2017). Entoptic landscape and ijereja: music as an iterative process. *New Sound Int. J. Music.* Belgrade: Department of Musicology, Faculty of Music Art, University of Belgrade. 49, 97–113. Available online at: <http://www.newsound.org.rs/pdf/en/ns49/08.L.Redhead.pdf>
- Redhead, L. (2022). Iterative processes: notating, composing, and performing as journey forms. *Contemp. Music. Rev.* 41, 266–280. doi: 10.1080/07494467.2022.2080459
- Richards, J. (2006). "32kg: Performance Systems for a Post-Digital Age" in *Proceedings of the International Conference on New Interfaces for Musical Expression (Paris, France, Zenodo: International Conference on New Interfaces for Musical Expression)*, 283–287. doi: 10.5281/zenodo.1176995
- Roggio, F., Trovato, B., Sortino, M., and Musumeci, G. (2024). A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: a narrative review. *Heliyon* 10:e39977. doi: 10.1016/j.heliyon.2024.e39977
- Santos, G., Wang, J., Brum, C., Wanderley, M. M., Tavares, T., and Rocha, A. (2021). Comparative Latency Analysis of Optical and Inertial Motion Capture Systems for Gestural Analysis and Musical Performance. *NIME 2021*. doi: 10.21428/92fbeb44.51b1c3a1
- Sauer, T. (2009). *Notations 21*. 1st Edn. New York: Mark Batty.
- Silang Maranan, D., Fdili Alaoui, S., Schiphorst, T., Pasquier, P., Subyen, P., and Bartram, L. (2014). "Designing for movement: Evaluating computational models using LMA effort qualities" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Association for Computing Machinery)*, 991–1000. doi: 10.1145/2556288.2557251
- Sivarathinabala, M., and Abirami, S. (2014). "Motion tracking of humans under occlusion using blobs" in *Advanced computing, networking and informatics- volume 1. Smart innovation, systems and technologies*. eds. M. Kumar Kundu, D. Mohapatra, A. Konar and A. Chakraborty, vol. 27 (Cham: Springer).
- Smith, J. (2022). Human-AI Partnerships in Generative Music. *International Conference on New Interfaces for Musical Expression, Auckland, New Zealand*. doi: 10.21428/92fbeb44.1aaaf7ab
- Stone, K. (1980). *Music notation in the twentieth century: a practical guidebook*. 1st Edn. New York: W. W. Norton.
- Tanaka, Atsu. Sensor-based musical instruments and interactive music, In Roger T. Dean (Roger T. Dean (ed.), *The Oxford handbook of computer music* (2011) Oxford: Oxford Academic
- Tanaka, A., and Donnarumma, M. (2019). "The body as musical instrument" in *The Oxford handbook of music and the body*. eds. Y. Kim and S. L. Gilman (Oxford: Oxford Academic).
- Taylor, A. (2016). "Collaboration" in contemporary music: a theoretical view. *Contemp. Music. Rev.* 35, 562–578. doi: 10.1080/07494467.2016.1288316
- Tobita, K., and Mima, K. (2024). Enhanced detection of musical performance timings using MediaPipe and multilayer perceptron classifier. *AI Comput. Sci. Robot. Technol.* 3:2. doi: 10.5772/acrt.20240002
- Traver, V. J., Latorre-Carmona, P., Salvador-Balaguer, E., Pla, F., and Javidi, B. (2017). Three-dimensional integral imaging for gesture recognition under occlusions. *IEEE Signal Process. Lett.* 24, 171–175. doi: 10.1109/LSP.2016.2643691
- Vear, C. (2019). *The digital score: Musicianship, creativity and innovation*. 1st Edn. New York: Routledge.
- Vickery, L. (2011). The evaluation of nonlinear musical structures. *Proceedings of the 2009 Totally Huge New Music Conference*, Vol. 3, Perth, Western Australia. Edith Cowan University. Available online at: <https://ro.ecu.edu.au/soundscripts/vol3/iss1/18/>.
- von Laban, R., and Lawrence, F. C. (1974). *Effort: economy of human movement*. 2nd Edn. London: Macdonald & Evans.

Wanderley, M. M. (2022). "Motion capture of music performances" in *The Oxford handbook of music performance*. ed. G. E. McPherson, vol. 2 (Oxford: Oxford Academic).

Waters, S. (2021). The entanglements which make instruments musical: rediscovering sociality. *J. New Music Res.* 50, 133–146. doi: 10.1080/09298215.2021.1899247

Xambó, A., Broadhurst, S., and Price, S. (2017). "Embodied music interaction: creative design synergies between music performance and HCI" in S. Broadhurst and

S. Price (Eds.), *Digital bodies: Creativity and technology in the arts and humanities* (London, United Kingdom: Palgrave Macmillan), 207–220. doi: 10.1057/978-1-349-95241-0_14

Yang, Q., and Essl, G. (2012). "Augmented Piano Performance using a Depth Camera" in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Michigan, USA, Zenodo: International Conference on New Interfaces for Musical Expression), 203–206. doi: 10.5281/zenodo.1178455