Melody and Motion: Integrating Guitar Gestures with Musical Patterns for Extended Control in Live Performance and Metaverse Applications

Nishal Silva
University of Trento
Trento, Italy
nishal.silva@uinitn.it

Marcelo M. Wanderley
McGill University
Montreal, Canada.
marcelo.wanderley@mcgill.ca

Luca Turchet
University of Trento
Trento, Italy
luca.turchet@unitn.it

***Abstract*—This paper introduces a novel system for detecting gestural movements of guitarists during live performances, integrated with a pre-existing musical pattern detection framework. The system enables guitarists to trigger peripheral devices through combined musical and physical gestures, extending instrumental control beyond traditional interfaces. We conducted a case study evaluation with five professional guitarists to explore the continuum between natural and theatrical instrument movements, examining how these gestures can be leveraged for expressive control. Technical evaluation of the system using a 500-event test corpus demonstrated a precision of 0.79, recall of 0.76, and F1 score of 0.78 for combined gesture and pattern detection. Through a set of interviews we further investigated practical applications within Musical Metaverse environments, highlighting opportunities for immersive performance experiences. Our findings reveal how combining musical content with physical expression creates intentional and novel performance controls, opening novel possibilities for interactive music that bridges traditional performances with virtual environments.**

***Index Terms*—Musical Pattern Detection, Gesture Detection, Guitar Performance, Musical Metaverse.**

I. Introduction

> "Basically it's a gesture which happens on the spur of the moment. I think, with guitar smashing, just like the performance itself; it's a performance, , it's an act..."

Pete Townshned, the guitarist for the popular rock band The Who, made the above statement during a 1968 interview for Rolling Stone Magazine about smashing guitars onstage[1]. This quote shows that guitarists frequently employ significant physical gestures as a core aspect of their live performances, using movement not only for sound production but also for theatrical effect and audience engagement. Research in music performance studies has shown that gestures serve multiple roles, from marking musical structure to communicating expressive intent and enhancing the visual impact of the performance [1], [2].

Notable examples from the literature and performance history illustrate the use of theatrical gestures in guitar playing:

- Jimi Hendrix was known for his flamboyant stage presence, often playing the guitar behind his back, between his legs, and with his teeth [3];
- Chuck Berry is credited for the popularization of the duckwalk, a stage move that became one of his signature moves and was replicated by many musicians [4];
- Jimmy Page of Led Zeppelin is known for incorporating a violin bow into his guitar solos [5], and performing with the instrument held high over his head. The Metropolitan Museum of Art describes Page's use of the violin bow to explore new sonic possibilities in psychedelic rock [5], [6];
- Angus Young of AC/DC is famous for his high-energy stage antics, which include running, jumping, performing guitar solos while lying on his back, and being carried through the audience;
- Iron Maiden are known for their theatrical performances. The scholarly discussion presented by [7] notes the inclusion of dynamic stage movement and theatrical gestures, particularly by the guitarists.

The use of movements and gestures, specifically the relationship between human movement and music research is explored in [8]. The authors define gestures as body movements that evoke meanings. The authors present methodological approaches that can be used in gesture research. The study discusses functional and communicative aspects of musical gestures, categorizing them into performer-performer and performer-perceive classes. The study also discusses using gestures as controls to extract the expressiveness of the human body.

While these expressive gestures and theatrics are central to the visual experience of a live performance, they are intertwined with the underlying music. Just as physical movements can dramatize moments within a performance, recurring melodic and rhythmic motifs provide the structural framework that guides both performer and audience through the musical narrative. Recurring melodic and rhythmic motifs often serve as cues for these dramatic movements, which suggests a link between what is played and how it is performed.

Studies have shown that humans typically associate struc-

[1] https://www.rollingstone.com/music/music-news/pete-townshend-talks-mods-recording-and-smashing-guitars-79369/

ture in music through the concept of repetition [9], [10]. While research on pattern detection in recorded music is common, its real-time counterpart has been relatively unexplored. In our previous work, we presented an algorithm to detect predefined patterns from the output of a musical instrument in real-time. We developed methods for symbolic monophonic pattern detection, and polyphonic audio pattern detection [11], [12].

We also proposed several prototypes of Smart Musical Instruments (SMIs), which were equipped with embedded computing devices running the Real-time Pattern Detection (RTPD) system [13]. The musician could define a set of musical patterns that would be used in a performance, which are used to train the Recurrent Neural Network (RNN) model. Once trained, the instrument is able to detect when the musician plays one of the patterns. This detection can be used as a trigger to control any number of peripheral devices or external stage equipment wirelessly. We conducted a multisensory concert experience, where three SMI controlled several stage lights, a smoke machine, 3D graphics on Mixed Reality (MR) headsets worn by the audience members, as well as haptic vibrations on audience members smartphones. The concert was conducted to evaluate the experience of musicians, as well as audience members while using the system. A video of the concert and of the ecosystem is available online[2].

How musical patterns function within guitar music, and how their integration with gestures can extend creative expression in live applications warrant a novel an interesting research. This paper highlights our work in developing a SMI with embedded musical pattern detection as well as gesture detection. The musician can take advantage of the combined use of patterns and gestures during their performances. As an example, the musician could define a set of musical patterns, as well as a set of gestures that can be done with an electric guitar. Upon training the model successfully, the musician can use the combination of gesture and pattern as a trigger for external peripherals or stage devices. We envision the use of such systems within the Musical Metaverse, where the gesture + pattern can be translated into the avatar of the user, and also as a trigger for various digital effects. The gesture detection is fulfilled by an Inertial Measurement Unit (IMU) sensor attached to the electric guitar. The same embedded computing device used for the RTPD is used in the gesture detection. The operation of the SMI is identical to our previous prototypes in every other way, where its self-powered, and capable of transmitting control messages wirelessly via Wi-Fi.

In this paper, we also present a case study conducted to identify a continuum between naturally occurring gestures, and gestures done purposely to induce a theatrical effect into the performance. We interviewed 5 professional guitarists who are experienced with live performances to identify the gestures they do, both involuntarily and purposely. We use the case study as motivation behind developing the gesture detection system, and we also use the musicians input to define possible use cases within virtual environments.

[2]https://youtu.be/axEHkdnxFB8

The rest of this paper is structured as follows. Section II discusses the current state of the art in the domains of pattern detection and gesture detection. Section III presents our case study conducted with guitarists. Section IV and Section V presents the methodology, and Section VI presents the results of the technical evaluation conducted. In Section VII, we discuss possible use cases within the Metaverse. Section VIII presents the conclusion and future work.

II. Related work

The combination of gesture detection, RTPD, and SMIs is a novel and interdisciplinary field bridging computer science, music technology, and performance studies [14]. This literature review summarizes the current research across multiple domains that are essential to understand the technical foundation and creative applications our work.

To date, numerous studies have explored the integration of sensors within musical instruments, or creating novel instruments with built-in sensing capabilities. A comprehensive review of sensors in musical expression devices, reported in [15], shows that 24 out of 75 projects using accelerometers also incorporate gyroscopes, while 11 projects implement complete MARG (Magnetic, Angular Rate, and Gravity) sensor configurations combining accelerometers, gyroscopes, and magnetometers.

The authors of [16] introduce a system composed on an inertial sensor and a hexaphonic nylon guitar for capturing strumming gestures. The study revealed a 0.77 normalized covariance coefficients of the displacement magnitudes, which show an acceptable trade-off between flexibility, portability and low cost when compared to high-end motion capture systems.

The TRAVIS II augmented violin proposed in [17] explores a flexible linear-potentiometer and FSR-based solutions for position, pressure, and force sensing integrated into violin and guitar fingerboards. Such augmented violin incorporates four conductive-PLA strips for touch tracking, along with four body-mounted FSRs with on-board Wi-Fi, to enable wireless operation while maintaining classical playability.

The Hyperbow introduces a commercial carbon fiber violin bow outfitted with a sensing system to measure changes in position, acceleration, and the downward and lateral strains of the bow stick [18]. The study uses electromagnetic field sensing, commercial MEMS accelerometers, and foil strain gauges. The authors show that the device could be easily controlled by a player using traditional right hand bowing technique.

The commercially available DigitAize Smart Violin incorporates a sensor equipped fingerboard for accurate finger position detection, along with 3D motion tracking to allow the violin body gestures to modulate sounds [19]. The K-Bow is also a commercially available bow equipped with sensors that measure bow position, acceleration, pressure and grip to control and translate such information into MIDI data for control. The bow is available for violin, viola, cello, and double bass [20].

In a different vein, musical pattern detection has been a popular field of research in Music Information Retrieval (MIR). A comprehensive survey of offline pattern recognition methods for polyphonic music transcription is presented in [21]. The authors examine signal processing techniques, machine learning models, and hybrid approaches, to establish a foundation for subsequent research. Similarly, [22] offers an in-depth overview of pattern discovery in music, reviewing current methodologies, identifying key challenges, and outlining potential directions for future research.

The paper reported in [23] introduced novel algorithms for matching patterns in symbolically encoded polyphonic music, allowing for pattern matching under transposition invariance. Their approach demonstrated improved accuracy and efficiency compared to previous methods, particularly in handling complex polyphonic structures. The authors of the study described in [24] introduced a method for extracting motivic patterns in polyphonic music based on adaptive redundancy filtering. Their approach could identify significant musical motifs in complex polyphonic textures, providing valuable tools for music analysis and composition.

In our previous work, we introduced several SMIs with RTPD capabilities. The method relied on a RNN-based model, which was trained with a synthetically generated dataset with possible expressive variations done to a ground truth. The training dataset was created using a rule-based model, inspired by a comprehensive dataset which include musical patterns along with repetitions containing artistic variations. The SMIs were evaluated with user studies involving both musicians and audience members.

In summary, despite musical pattern detection and gesture detection has been investigated independently, to the best of our knowledge research has not been conducted to combine the two approaches to allow extended expressive control by a musician. The work presented by this paper aims to bridge this research gap, and to possibly inspire future works in such multimodal expressive controls.

III. Case study: The continuum between natural movements and theatrical gestures

Our study is primarily focused on the electric guitar, as guitarists often move their instrument for theatrical effects. The work presented in [1] highlights how pronounced fretboard gestures and body movements by guitarists like Alex Lifeson of Rush and Sister Rosetta Tharpe, not only reflect musical phrasing but also serve communicative and theatrical functions. Similarly, motion-capture studies reveal that guitarists may exaggerate movements at key points –such as transitions between song sections– to emphasize narrative or emotional content. Further research reported in [2] found that guitarists adjust their posture and movement style in response to musical timing and expressive intent, supporting the view that such gestures are deliberate and meaningful rather than incidental.

The comprehensive analysis presented by [5] examines the kinetic qualities of Led Zeppelin's music, focusing on how the band members used their bodies in performance. The author presents a detailed analysis of physical gestures employed by Jimmy Page - the guitarist of Led Zeppelin. The study identifies movements that represent musical content, and also gestures to mark rhythmic emphasis, and gestures to represent abstract musical ideas. The study presents photographs and diagrams of the guitarist performing some of his most iconic gestures live.

Significant gestural movements by guitarists are well documented, and an essential feature of a live performance, serving both musical and theatrical purposes. With this in mind, we conducted a case study with 5 professional guitarists (all male, mean age 32.2, experienced in live performances with bands) to identify a continuum between functional movements—those done to have better technique or ease of playing, and theatrical movements—those done deliberately for visual impact. The participants were shown videos and images of guitarists using theatrical gestures during live performances, after which they were asked questions about the movements they do while playing. The questions regarding the movements were split into two sections as shown below:

Natural Movements

- Q1 What unconscious movements do you make when concentrating complex passages?
- Q2 How much do you tilt the guitar upwards when playing on higher frets? Do you do it for ease of playing, or for theatrics?
- Q3 Do you move your instrument in other ways for ease of playing or to have better technique during certain sections? What sections are they, and how do you move the instrument?
- Q4 When you prepare for a challenging passage or solo, do you find yourself making preparatory movements?
- Q5 Can you recall any repetitive or habitual movements that are essential for your playing but not intended for the audience?
- Q6 Have you ever noticed a functional movement becoming exaggerated over time, perhaps evolving toward something more theatrical? Can you give an example?

Theatrical Movements

- Q7 What are the most deliberate, visually impactful gestures you use on stage?
- Q8 For each theatrical gesture, when in the music do you typically perform it (e.g., during a solo, chorus, or a particular lyric)?
- Q9 How do you decide when to incorporate a theatrical gesture versus keeping your movements more restrained? Does the size or energy of the audience influence this?

The participants were then shown recordings of concerts conducted within the Metaverse. This step was done to ensure that they have a sufficient understanding of the visual effects, avatars, and creative possibilities within the Metaverse. Then they were asked the following questions about how they would translate their movements and musical patterns within the virtual world.

Q10 In a virtual concert, which of your movements would translate well to an avatar?

Q11 How could virtual effects amplify your gestures—trails, particle effects, environmental responses? What would enhance rather than distract from your performance?

Q12 Describe scenarios where your guitar movements could trigger virtual effects—lighting changes, environmental shifts, audience interaction elements.

A. Functional-to-Theatrical Continuum

All participants exhibited various movements during technically demanding passages. Some participants expressed adaptive, comfort seeking behaviors, stating “I shift my weight between the two legs. I adjust myself to feel as much comfortable as possible”. Another participant expressed shifting the body for better visibility: “I tend to lean in towards the guitar to have a better line of sight at the fretboard”. A majority of the participants (4 of 5) acknowledged that they tilt the guitar upwards for easy access to higher frets: “I tilt the guitar quite often, its primarily for ease of access when playing in higher frets”.

Movement patterns were tied to specific techniques and dynamics, and also key phrases in music. One participant described context-specific adjustments: “I move the instrument slightly if it helps me play better. I tend to tilt the guitar when playing arpeggios to keep my thumb in the middle of the neck. Its usually the changes in dynamics”. Another participant emphasized preparatory movements: “I brace myself when I am playing fast sections. Also I tend to brace myself during sections where only I am playing to keep better concentration”. Participants also expressed the movements after specific phrases: “Usually, if a guitar solo ends with bend or a note high on the freeboard, I lift my guitar for theatrical effect. It’s like a cut to the audience that the solo is over”; “Sometimes, when I have to hold a chord, I point the guitar at someone in the audience who looks like they are enjoying the show”.

All participants expressed various subconscious movements that are essential to them, but not specifically intended to the audience. Sudden repositioning of the guitar, checking the pickup selector switch, looking at others in the band, standing on one leg, and noodling around the fretboard when not playing were some of the habitual movements highlighted.

Moreover, all participants acknowledged the transition from functional to theatrical movements. Once participant stated to be conscious of some of the functional movements: “I don’t think they have exaggerated much, but they probably have become more refined and elegant through the years - so that it’ll look a better on stage”. Another participant Stated that they deliberately emphasize some movements for theatrics: “I do tilt the guitar upwards when playing at higher frets. But if the section of the song is energetic and pumped, I deliberately raise it more, sometimes I even walk to front of stage to do it”. Another participant discussed how they use theatrical movements for function: ”I point the guitar at the crowd. It looks very cool, but it also gives me better visibility of the fretboard”.

Furthermore, all participants linked theatrical movements to musical dynamics and technical difficulty: “It’s not really during a specific section, but more about changes in dynamics. If I am playing a particularly aggressive section, I tend to be a bit more theatrical. But when I play difficult and challenging passages that require a lot of concentration, I tend to stay more still”. The decision to incorporate theatrical gestures was influenced by multiple factors. One participant emphasized energy reciprocity: “It usually depends on the energy I have, and also on the energy that the audience projects”. Another participant provided a more comprehensive framework: ”It depends on how much focus is required of me. If the part is easy, I am more inclined to do big gestures. If the audience is engaged, my movement will also increase for sure”.

B. Metaverse Integration Insights

All participants expressed eagerness to translate their movement to virtual avatars. One participant, with a high experience in Virtual Reality (VR) environments stated: “If you can have full body and instant tracking, you can translate almost any gesture. However I believe that big gestures where my whole body and the instrument is used will be better to translate to my avatar”. Another participant suggested the potential for movement amplification: “The type of movements will be the same. It might be fun to experiment with even more movements. Somehow seeing an avatar, or seeing other people as avatars, I’ve found that it will exaggerate my movements”.

The participants had different views on visual effects that could be triggered using the system. Notably, some participants expressed the need for novel visual effects that will work exclusively on the virtual world. Environment changes, transporting the users to a different virtual location, changing the size of the avatars, virtual light trails and particles were some of the visual effects suggested by the participants.

C. Case Study Insights

The case study revealed a complex movement taxonomy that extends beyond simple functional-theatrical categorization. The findings suggest that guitarist’s movements exist on a continuum with multiple overlapping functions, including comfort optimization, technique enhancement, sound production, and audience engagement. The participants demonstrated that movements can serve multiple purposes simultaneously, with functional movements often evolving into more refined theatrical expressions over time.

Based on the responses of the participants, we selected two distinct, yet common movements done by guitarists to evaluate the gesture detection system. Tilting the neck upwards was one of the most common gestures done by guitarists, and the rotation of the guitar neck horizontally towards the audience were selected due to their popularity among guitarists, and also their clear distinction in a 3-dimensional axis which has the potential for an accurate initial investigation. Figure 1 shows

Fig. 1. The first author raising the fretboard during a solo. Piazza Duomo, Trento, 2025. Photo credits: Axsel Sociati @*photo_sociati*.

Fig. 2. A bass guitarist (Steve Harris, Iron Maiden) pointing the fretboard forward at the audience. Iron Maiden the Future Past Tour, 2025. Image extracted from the official webpage[4].

a guitarist tilting the guitar up, and Figure 2 shows a guitarist moving the neck horizontally towards the crowd.

The responses of the participants also revealed that it is common to move the instrument after certain segments or sections. This is further supported by the works presented by [25], where the authors discovered a correlation between a musicians' movements across different performances of the same composition which suggests that even subtle gestures can be consistent and potentially meaningful in musical performance. This fact is a fundamental motivation for our research on combining musical patterns and gestures to allow extended controls. Our motivation is also supported by [5], which discusses movements that represent musical content (such as raising the guitar for high notes). Moreover, the work presented by [26] also supports the notion that body movements served to convey the performers' musical interpretations of the score.

IV. Real-Time Musical Pattern Detection

In our previous work, we introduced the polyphonic audio pattern detection system to enable the creation of SMIs [12]. The system consists of a Raspberry Pi 4 and a HiFiBerry DAC+ADC device that can be integrated with conventional electric instruments. The instrument's audio output is split, with one path directed to the detection system and the other to standard amplification or PA systems. Wireless connectivity on the Raspberry Pi 4 is used to transmit Open Sound Control (OSC) messages upon successful pattern detection, allowing for flexible control of networked devices. Figure 3 shows a block diagram of all the components of the RTPD enabled SMI.

The RTPD is done using a RNN trained to classify incoming audio segments into a predefined set of musical patterns. The detection operates in real-time: audio is sampled in 30 ms windows, a duration chosen to match the temporal resolution of human auditory perception. Each window is processed to extract Mel Frequency Cepstral Coefficients (MFCCs), which provide a compact, perceptually relevant representation of both spectral and temporal characteristics crucial for music pattern analysis.

MFCC parameters are optimized for the embedded platform: FFT size of 2048, hop size of 512, and 10 Mel bins. These settings provide a balance between computational efficiency and classification accuracy. The extracted MFCCs are input to the RNN, which outputs the predicted pattern class. Upon detecting a pattern, the system transmits an OSC message corresponding to the detected pattern, enabling real-time interaction with external devices.

The RNN model consists of a stacked architecture comprising of:

- A Masking layer to handle variable-length input sequences,
- A Long Short-Term Memory (LSTM) layer with 256 units,
- A Gated Recurrent Unit (GRU) layer with 128 units,
- A Simple RNN layer with 256 units,
- A final dense layer with a tanh activation, outputting one value per pattern class.

Zero-padding is applied during training to standardize input dimensions. The model is designed to be lightweight for real-time inference on embedded hardware, while maintaining high classification accuracy.

A robust training dataset is essential for generalizable real-time detection. Following our previous work [11], a rule-based approach is used to generate synthetic variations of each ground-truth musical pattern. Rules are applied to the MFCC

[4] https://www.ironmaiden.com/media/the-future-past-tour/

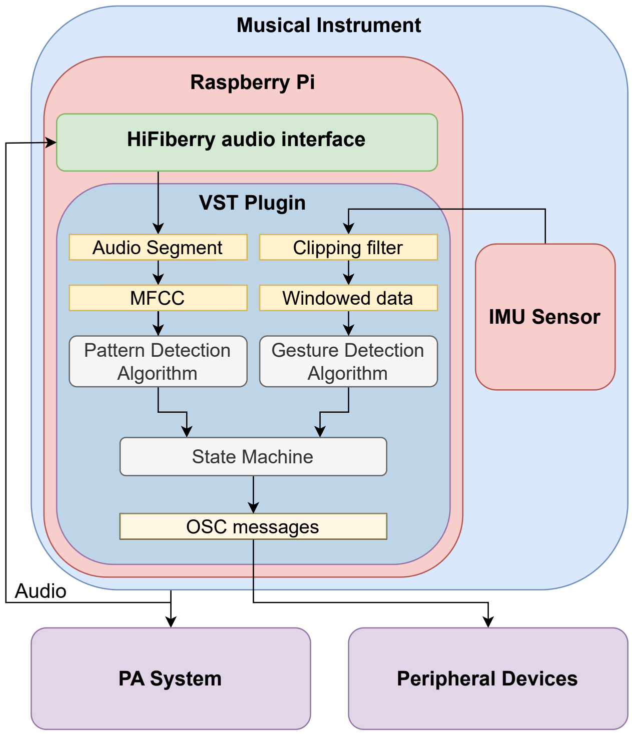

Fig. 3. Block diagram of the RTPD and gesture detection done within the SMI.

spectrograms to simulate expressive performance variations, some of them being:

- Up to 24% added notes,
- Up to 3% removed notes,
- 9% added pauses,
- 1% removed pauses.

Each ground-truth pattern is expanded to approximately 10,000 variations, capturing a wide range of possible musician interpretations. This approach ensures the model is resilient to expressive deviations and adaptable to diverse musical styles.

V. Proposed Method

The proposed method consists of two independent detection algorithms systems working in parallel. The pattern detection algorithm, as well as the gesture detection algorithm both run within the Virtual Studio Technology (VST) plugin independently, their outputs combined by a state machine to obtain a combined output that is dependent on both the musical pattern and gesture.

The gesture detection is done using a BNO055 IMU sensor. The system is implemented using a sliding window approach, similar to the RTPD system (see Section IV) and a RNN based deep learning inference engine. The BNO055 sensor serves as the primary data acquisition unit, providing 9-axis inertial measurement capabilities including a 14-bit accelerometer, 16-bit gyroscope, and magnetometer. The sensor operates at a sampling frequency of 100 Hz with 10ms intervals. Figure 5 shows the developed SMI prototype, where the IMU sensor is attached to the headstock (Figure 6), as it is possibly the location that experiences the most movement in a electric guitar. The IMU is connected directly to the GPIO port of the Raspberry Pi 4, also mounted to the guitar body (Figure 6).

The BNO055 sensor is configured for IMU mode operation with I2C communication. The sensor provides calibrated acceleration data in three orthogonal axes (X, Y, Z), at a sampling frequency of 10 Hz, with automatic sensor fusion capabilities. Calibration procedures follow standard protocols where the gyroscope requires static positioning, while the accelerometer necessitates 6-axis calibration across different orientations [27].

The algorithm first filters the accelerometer data by means of a hard clipping filter to prevent saturated measurements, noise, and artifacts from introducing cumulative errors. Following the filtering, the X, Y, Z data is obtained using a sliding window. A fixed-size buffer of 100 samples corresponding to a 1 second temporal window is used to allow sufficient temporal context for gesture pattern recognition while maintaining a real-time operation. The sliding window operates as a First In First Out (FIFO) queue, where new samples are continuously added while the oldest samples are removed when the buffer reaches capacity. A feature aggregation is then done by concatenating the X, Y, and Z vectors into a single vector to be fed to the RNN for classification.

We designed the RNN model with a simpler architecture than the one for RTPD, as the task required a significantly smaller latent space. A RNN architecture with a single LSTM layer with 256 units was empirically found to be sufficient to detect the two gestures through acceleration. The RNN model was trained with a custom made dataset consisting of repetitions of the two gestures.

The gesture detection, as well as the RTPD is done in parallel by exploiting the multi-threading capabilities of the embedded system. The combined result of both system is obtained using a Hierarchical Finite State Machine (HFSM), which provides the capability to ensure a detection only when the musical pattern, and then the gesture is detected sequentially. This step was motivated by the participants of the case study who expressed introducing a theatrical element by the use of big gestures immediately after playing an iconic musical phrase.

The state machine architecture has three distinct states and four transition paths. Figure 4 illustrates the complete operational flow from musical pattern recognition through gesture detection. Below are the state descriptors:

- **Idle State** serves as the default operational mode where the system continuously monitors audio input for musical patterns, and the IMU sensor for gestures.
- **Pattern Detected State** is the intermediate state activated when the system identifies a musical pattern that may precede a gesture. In this state, the system will pause the RTPD and focus on gesture detection. The system maintains two exit conditions: successful gesture detection or timeout protection to prevent indefinite processing.

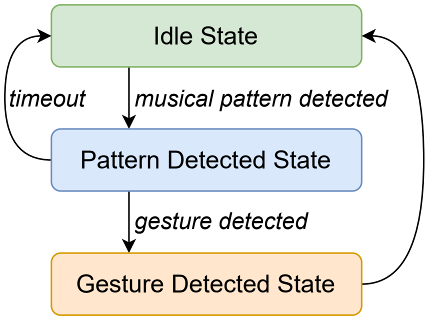

Fig. 4. HFSM diagram for RTPD and gesture detection.

Condition	Description	Trials
P1 $\rightarrow$ G1	Musical pattern 1 immediately followed by gesture 1	100
P2 $\rightarrow$ G2	Musical pattern 2 immediately followed by gesture 2	100
P1 $\rightarrow$ G2 or P2 $\rightarrow$ G1	Correct pattern with wrong gesture	100
Pattern-only	Either pattern without a gesture	100
Gesture-only	Either gesture without a pattern	100

TABLE I
TEST CORPUS CONDITIONS AND TRIALS

- **Gesture Detected State** represents the brief transitional state where the system has successfully identified musical pattern, immediately followed by a gesture. This state triggers the transmission of the corresponding Open Sound Control (OSC) command to external systems and immediately returns to the Idle state for continued monitoring.

VI. EVALUATION AND RESULTS

To asses the accuracy of the system, we conducted a technical evaluation. The evaluation consisted of a setup where the RTPD system, as well as the gesture detection system were trained to detect two patterns each. The evaluation was done by one of the authors, where they performed a 500 event test corpus in a controlled setting. The evaluation was done using the developed prototype SMI, which integrates a Raspberry Pi 4, HiFiBerry audio interface, and a BNO055 IMU sensor mounted on the headstock.

Two distinct musical patterns (P1 and P2) were pre-defined and trained into the RTPD system. Two gross rotational gestures – neck raise (G1) and forward tilt (G2) – were selected based on case study insights. The evaluation was distributed across five experimental conditions to evaluate both true positives and potential failure modes. The test corpus is presented in Table I. The precision, recall and F1 score of the evaluation are presented in Table II.

Though not part of the technical evaluation, the tester noted the *forgiving* nature of the HFSM, as mis-played patterns never triggered gestures, and stray motions during silences were ignored. This qualitative observation shows a benefit in combining musical and physical cues to use as triggers.

Precision	0.79
Recall	0.76
F1 score	0.78

TABLE II
PRECISION, RECALL, AND F1 SCORE OF THE TECHNICAL EVALUATION.

VII. DISCUSSION

In this paper, we introduced a novel system intended to allow extended controls to a guitarist. The system comprises a musical pattern detection system as well as a gesture detection system working in tandem and in real-time to allow the creation of a unified SMI. Both the RTPD system and the gesture detection system are accomplished through the use of RNN-based deep learning. A HFSM is used to combine the two detection systems to allow a musician to use musical a pattern, immediately followed by a gesture as an extended control to their performance. We envision the usage of such systems within the Musical Metaverse [28], where the musical patterns and gestures control digital effects in the virtual world.

The technical evaluations was done to assess the accuracy of the system. We selected two musical patterns and two gestures (raising the guitar neck upwards and turning the guitar neck forward pointing at the audience) to assess the accuracy. Evaluations were done using a 500-event test corpus where $P_1 \rightarrow G_1$ and $P_2 \rightarrow G_2$ represented the correct sequence of patterns and gestures, $P_1 \rightarrow G_2$ and $P_2 \rightarrow G_1$ represented the wrong sequence of patterns and gestures, pattern only, and gesture only events where P_1 is Pattern 1, and G_1 is Gesture 1. The evaluations revealed a precision of 0.79 and recall of 0.76. These results show that integrating musical pattern and gesture detection has promising potential applications. Qualitative feedback from the evaluation suggests improved performer confidence when gestural control is associated with musical structure.

The case study conducted with five professional guitarists revealed an interesting continuum of functional and theatrical movements in live performance. Skill-driven, comfort-oriented gestures (i.e., neck tilting for higher-fret access) were observed to evolve into deliberate visual flourishes (i.e., neck raises following iconic sections). Participants reported that musical dynamics, technical difficulty, and audience energy jointly inform their choice to emphasize gestures. The selection of the two gestures to evaluate the system was motivated by the responses to our case study.

The integration of musical pattern and gesture detection can open a wide array of creative and interactive possibilities within the Musical Metaverse. Insights from the case study questionnaire reveal that all guitarists and envisioned novel applications that extend beyond traditional performance, enabling immersive, participatory, and expressive experiences in virtual environments.

The visual effects triggered by the proposed system can be real-time avatar mirroring, allowing the audience to witness expressive movements; or initiate unique avatar animations such as changing the costume, or size of the avatar. Musi-

Fig. 5. The developed SMI prototype, with the embedded system and IMU sensor.

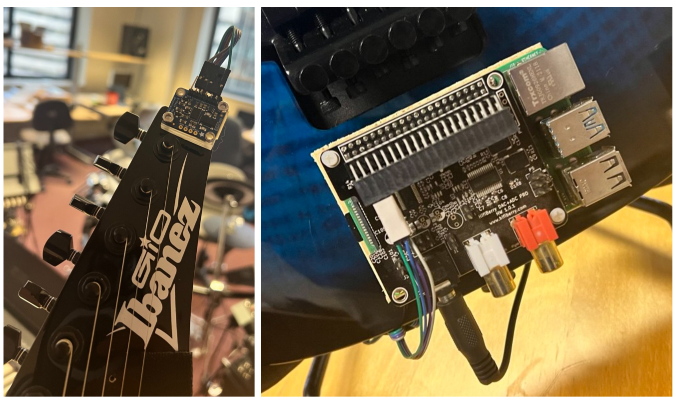

Fig. 6. The embedded computing system is attached to the body of the SMI, and the IMU sensor is attached to the headstock.

cians also expressed interest in translating their gestures into environment changes, 3D graphics, or other environmental and ambient effects. Musicians can also assign combinations of patterns and gestures to control digital effects of external devices, both in the physical and virtual world. Hence, SMIs with embedded sensors can become programmable controllers for the Metaverse, allowing artists to experiment with new forms of expression and interaction.

One of the limitations of our study is the focus on only two musical patterns and two gestures. Expanding to a larger, more diverse vocabulary may call for more optimization to the model. Moreover, the current gesture set is restricted to gross rotational movements; finer-grained controls (e.g., hand-shape changes) could be addressed in future work. Finally, the system's performance under noisy stage environments and network congestion also warrants further study.

VIII. Conclusion and Future work

This paper presented a novel system that integrates real-time musical pattern detection with gesture recognition for the electric guitar, intended for performers to achieve extended control in virtual environments. The pattern detection, and the gesture detection is done using trained Recurrent Neural networks, and a hierarchical finite state machine ensures that only the intended combinations of musical patterns and gestures are captured.

A user-centered case study with 5 professional guitarists revealed a nuanced continuum between functional and theatrical movements, which influenced the selection of gestures and motivating the system's design. Technical evaluation demonstrated a precision of 0.79, recall of 0.76, and F1 score of 0.78 in a 500-event test corpus. Qualitative feedback highlighted the system's reliability and its potential to boost performer confidence by aligning gestural control with musical structure. The findings shows the potential of integrating multimodal inputs for smart musical instruments, which can open new possibilities for creative expression in both physical and virtual contexts.

While the current study focused on two musical patterns and two gross gestures, future work will expand the gesture vocabulary, test the system with a broader range of performers and environments, and further explore applications in immersive and accessible music-making. We also plan to incorporate additional gestures (e.g., strum force, hand shapes) via multi-sensor fusion enhance the system. We hope to conduct further longitudinal studies with diverse performer populations to assess learning curves, creative adoption, and the impact of embodied control on musical expression. We also plan to investigate adapting gesture and pattern detection to recognize a broader range of expressive inputs for users with physical disabilities. We hope our initial investigative work lays a foundation for more immersive, interactive, and expressive musical experiences that bridge physical and digital performance spaces.

Acknowledgment

We acknowledge the support of the MUSMET project funded by the EIC Pathfinder Open scheme of the European Commission (grant agreement n. 101184379). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Innovation Council. Neither the European Union nor the European Innovation Council can be held responsible for them.

References

[1] S. Gardner and N. J. Shea, "Gestural perspectives on popular-music performance," *Music Theory Online*, vol. 28, no. 3, August 2022. [Online]. Available: https://mtosmt.org/issues/mto.22.28.3/mto.22.28.3.gardnershea.pdf

[2] M. R. Haugen, G. S. Câmara, K. Nymoen, and A. Danielsen, "Instructed timing and body posture in guitar and bass playing in groove performance," *Musicae Scientiae*, vol. 28, no. 3, pp. 407–420, 2023, original work published 2024.

[3] S. K. Lavine, ""are you experienced'? the life, music, and legacy of jimi hendrix"," *Student Publications*, 2023. [Online]. Available: https://cupola.gettysburg.edu/student_scholarship/1086

[4] T. Taylor, "The origins of chuck berry's famous duck walk," *Far Out Magazine*, 2021, accessed July 7, 2025. [Online]. Available: https://faroutmagazine.co.uk/the-origins-of-chuck-berrys-famous-duck-walk/

[5] S. Fast, *In the Houses of the Holy: Led Zeppelin and the Power of Rock Music*. Oxford University Press, 09 2001. [Online]. Available: https://doi.org/10.1093/oso/9780195117561.001.0001

[6] J. K. Dobney and C. J. Inciardi, *Play It Loud: Instruments of Rock & Roll*. New York: The Metropolitan Museum of Art, 2019.

[7] L. Meller, "Historical themes in iron maiden songs (part ii): From the inquisition to the second world war," *Revista Brasileira de Estudos da Canção*, vol. 4, pp. 190–203, 2013. [Online]. Available: https://rbec.ect.ufrn.br/data/_uploaded/artigo/N4/RBEC_N4_A14.pdf

[8] A. Refsum Jensenius, M. M. Wanderley, R. I. Godøy, and M. Leman, "Musical gestures : concepts and methods in research," in *Musical gestures : sound, movement, and meaning*, Godøy, Rolf Inge and Leman, Marc, Ed. Routledge, 2009, pp. 12–35. [Online]. Available: http://doi.org/10.4324/9780203863411-8

[9] R. Dannenberg and N. Hu, "Linear time for discovering non-trivial repeating patterns in music databases," in *ISMIR 2002 Conference Proceedings: Third International Conference on Music Information Retrieval*, 2002, pp. 63–70.

[10] I. Y. Ren, H. V. Koops, A. Volk, and W. Swierstra, "In search of the consensus among musical pattern discovery algorithms," in *Proceedings of the 18th International Society for Music Information Retrieval Conference*, 2017, pp. 23,27.

[11] N. Silva and L. Turchet, "Real-time pattern recognition of symbolic monophonic music," in *Proceedings of the 19th International Audio Mostly Conference*, ser. AM '24. New York, NY, USA: Association for Computing Machinery, 2024.

[12] ——, "Real-Time Audio Pattern Detection for Smart Musical Instruments," *Journal of Audio Engineering Society*, 2025 (submitted).

[13] ——, "User-Centered Evaluation of Smart Musical Instruments with Embedded Real-Time Pattern Detection," in *Proceedings of the Immersive and 3D Audio conference*, 2025, (submitted).

[14] E. Volta and N. D. Stefano, "Using wearable sensors to study musical experience: A systematic review," *Sensors*, vol. 24, no. 17, p. 5783, 2024. [Online]. Available: https://doi.org/10.3390/s24175783

[15] C. B. Medeiros and M. M. Wanderley, "A comprehensive review of sensors and instrumentation methods in devices for musical expression," *Sensors (Basel)*, vol. 14, no. 8, pp. 13 556–13 591, July 25 2014.

[16] S. Freire, G. Santos, A. Armondes, E. Meneses, and M. Wanderley, "Evaluation of inertial sensor data by a comparison with optical motion capture data of guitar strumming gestures," *Sensors*, vol. 20, no. 19, p. 5722, 2020. [Online]. Available: https://doi.org/10.3390/s20195722

[17] T. Grosshauser and G. Tr´oster, "Finger position and pressure sensing techniques for string and keyboard instruments," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Daejeon, Republic of Korea: Graduate School of Culture Technology, KAIST, May 2013, pp. 479–484. [Online]. Available: http://www.nime.org/proceedings/2013/nime2013_286.pdf

[18] D. Young, "The hyperbow controller: real-time dynamics measurement of violin performance," in *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, ser. NIME '02. SGP: National University of Singapore, 2002, p. 1–6.

[19] Bridge Street Violins, "Digitaize smart violin – instrument and midi controller in 1," 2024, accessed July 10, 2025. [Online]. Available: https://bridgestreetviolins.com/violins-for-sale/p/digitaize-smart-violin-instrument-and-midi-controller-in-1

[20] Keith McMillen Instruments, "K-Bow: A MIDI Controller Translating Violin Bow Gestures into MIDI Data," n.d., accessed July 10, 2025. [Online]. Available: https://www.keithmcmillen.com/labs/k-bow/

[21] F. Jamshidi, G. Pike, A. Das, and R. Chapman, "Machine learning techniques in automatic music transcription: A systematic survey," 06 2024.

[22] B. Janssen, W. de Haas, A. Volk, and P. Van Kranenburg, "Discovering repeated patterns in music: state of knowledge, challenges, perspectives," 10 2013.

[23] A. Laaksonen and K. Lemström, "Advanced Polyphonic Music Pattern Matching Algorithms with Transposition Invariance," *Journal of Music Information Retrieval*, vol. 10, no. 2, pp. 45–67, 2024.

[24] O. Lartillot and P. Toiviainen, "Motivic matching strategies for automated pattern extraction," *Musicae Scientiae*, vol. 11, no. 1_suppl, pp. 281–314, 2007.

[25] M. M. Wanderley, "Quantitative analysis of non-obvious performer gestures," in *Gesture and Sign Language in Human-Computer Interaction*, I. Wachsmuth and T. Sowa, Eds. Springer, 2002, pp. 241–253.

[26] M. M. Wanderley, B. W. Vines, N. Middleton, C. McKay, and W. Hatch, "The musical significance of clarinetists' ancillary gestures: An exploration of the field," *Journal of New Music Research*, vol. 34, no. 1, pp. 97–113, 2005.

[27] H. Khan, A. S. A Rashid, and N. Nasir, "Evaluation and calibration of mems-imu sensors for real-time landslide detection and monitoring system," *Engineering Research Express*, vol. 7, no. 2, p. 025104, apr 2025. [Online]. Available: https://dx.doi.org/10.1088/2631-8695/adce56

[28] L. Turchet, "Musical metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1811–1827, 2023.