Audio Engineering Society

Conference Paper

Presented at the AES International Symposium on
AI and the Musician
2024 June 6–8, Boston, MA, USA

When XR Meets AI: Integrating Interactive Machine Learning with an XR Musical Instrument

Max Graf[1] and Mathieu Barthet[1]

[1]*Queen Mary University of London*

Correspondence should be addressed to Max Graf (`max.graf@qmul.ac.uk`)

ABSTRACT

This paper explores the integration of artificial intelligence (AI) with extended reality (XR) through the development of Netz, an XR musical instrument (XRMI) designed to enhance expressive control using deep learning techniques. Netz implements algorithms to map physical gestures to musical controls, offering customisable control schemes that enhance gesture interpretation accuracy and elevate the overall musical experience. The instrument was developed through a participatory design process involving a professional keyboard player and music producer. The process spanned three phases with corresponding design sessions: exploration, making, and performance & refinement. Initial challenges with traditional computational approaches to hand-pose classification were overcome by incorporating an interactive machine learning (IML) system, enabling personalised gesture control. A set of musical performance tasks encompassing melodies and chord progressions were used to assess the instrument's playability and expressivity in collaboration with our musician partner. Thematic analysis of reflective interviews revealed that the IML system enhanced musical interaction, suggesting AI's potential to improve XR musical performance. Future work will involve a wider range of musicians to assess the generalisability of our findings.

1 Introduction

In the evolving landscape of digital music, the merging of extended reality (XR), encompassing augmented, mixed, and virtual reality, and AI holds the potential to unlock new avenues for musical expression and instrument interaction. This paper describes the development of Netz, an XR musical instrument (XRMI), which exemplifies this synergy by utilising deep learning to personalise gesture control for expressive musical interaction. Netz is a hand-controlled XR instrument that uses machine learning (ML) to refine the translation of musicians' gestures into musical controls. The use of ML reduces the impact of technological sensing errors and empowers expressive control in XR music creation. Ontologically, our AI approach focuses on augmenting musicians' control over the instrument, providing them with new creative possibilities. This differs from AI aimed at automating tasks traditionally done by musicians, such as generative modelling for composition or performance. XR head-mounted displays (HMDs) blend physical and virtual spaces through head-tracking, stereoscopy and real-world scene understanding, enabling new forms of instrument design and interaction. Zellerbach and Roberts broadly defined an XRMI as an "embodied system for expressive musical performance, characterized by the relationships between the performer, the virtual, and the physical environment" [1]. Out of the many current issues in interaction design for XR, we want to highlight two in particular: sensing-based technological faults [2] and accurate and timely control of musical interface elements.

XR HMDs use head, hand, and body pose detectors to capture motion data, forming embodied control interfaces. While hand tracking in XR often employs machine learning for 3D pose generation from camera images, these systems do not always achieve perfect accuracy, particularly in high-resolution finger tracking. This can lead to jitter, tracking loss, and glitches, which negatively impact the performance of hand-controlled XRMIs by affecting expressive control gestures crucial for musicians, as shown in our previous work [3]. These issues are especially important for instrument makers, as expressive control gestures are paramount in musical instruments [4].

This paper explores (i) how constraints for effective instrument interaction with current XR technology were revealed during participatory design, and (ii) how we addressed these constraints through interactive machine learning (IML), leading to improved hand gesture recognition for XR musical interaction. Our main contributions include (i) design insights for XRMIs emerging from participatory design and (ii) an interactive machine learning pipeline for XRMIs. This pipeline allows instrument designers to train deep neural networks on external machines and deploy them on XR devices using a client/server architecture. We also provide (iii) an analysis of the utility of interactive machine learning for control gesture recognition within XRMIs.

2 Background

Existing works in the XRMI domain focus on the design and implementation of instruments [5, 6, 7], user experience [8, 9], general techniques for interaction [10, 11], or collaborative music making [12, 13]. Several approaches have been proposed to circumvent sensing-related limitations of XR devices. One strategy is to create instruments with larger interface control elements, as seen in [14, 8, 9]. Another approach is to introduce auxiliary tracking devices, such as XR device controllers, webcams, or external hand tracking devices to augment the sensors of the HMDs [5, 15, 7]. Our research focuses on a gap in XR musical interfaces: compact instruments that enable accurate control through hand and finger gestures using HMDs, eliminating the need for external tracking devices.

Sensor data representing embodied musical gestures can be processed and stored to enable real-time action-to-sound mapping [16]. Machine learning models can be trained to learn the commonalities and differences between the gestures of different performers, as shown in studies using the Wekinator interactive machine learning system [17]. The key feature of Wekinator is its approach to training. A model in Wekinator learns from user interactions, eliminating the need for a pre-collected training dataset. Françoise et al. built upon this concept by developing probabilistic models mapping movement to sound for real-time performance [18]. The authors employed a *mapping-by-demonstration* technique to establish the relationships between movement and sound. Similar to the Wekinator workflow, models are trained by learning examples of gestures directly from performers. Deep neural networks have shown promise in enhancing digital music instrument control, user enjoyment, and expressive capabilities [19]. The *Fluid Corpus Manipulation* project resulted in techniques for decomposing individual sounds using ML, enabling real-time manipulation, combination, and synthesis for artistic expression [20].

3 Methods

Developed through a user-centered approach, Netz is the result of twelve participatory design (PD) sessions with a professional keyboard player. This collaborative and longitudinal process ensured the instrument caters to the needs of real-world performers. The study received ethical approval at our university (#22.162). The participant was compensated for his time.

Our design objectives fall into two categories: overarching design objectives (ODOs) stemming from our research focus on AI & Music and past work, and functional design objectives (FDOs) identified through participatory design sessions. Our overarching design objectives were as follows:

ODO$_1$: Develop an XRMI that leverages performers' skills and AI for real-time analysis and augmentation of natural hand and finger gestures for music creation.

This type of control addresses a key limitation identified in our previous work [3].
ODO_2: Maintain a compact interface design within the XR environment to ensure it remains visible within the user's field of view (FOV). One major issue with existing XRMs is their physical size, which, when coupled with the limited FOV of HMDs, can lead to performer fatigue due to the high number of head movements required, especially in longer performances [14, 9].
ODO_3: Enable the integration of the virtual interface with a tabletop surface to enable haptic feedback through physical interaction with the surface. This also addresses a previous limitation [3].

3.1 Longitudinal Study with an Expert

We employed a longitudinal participatory design approach with a single participant. This enabled us to iteratively define the instrument's functional design objectives while simultaneously exploring the evolving relationship between the participant and the instrument being developed. Our participant was a 28-year-old male of Indian origin, with professional experience as an audio developer and music producer. He had over 20 years of experience playing piano and keyboard, particularly in electronic music and jazz genres. His musical sophistication was assessed using the Goldsmiths Musical Sophistication Index [21], scoring 103, significantly above the UK's average of 81.58.

Over 12 weeks, the participatory design engaged the participant in three phases: exploration & definition, making, and performance & refinement, drawing inspiration from [22]. Sessions ranged from one to two hours and were all recorded and transcribed.
Through **exploration & definition** sessions (involving mockup workshops and discussions of existing instruments), we identified the participant's needs and brainstormed potential interfaces, ultimately selecting feasible features within time and hardware constraints. In this phase, we formulated functional design objectives for the XRMI. The **making** phase focused on iterative development and testing of an XRMI prototype, gathering participant feedback on instrument features through short performances and interviews.
In the **performance & refinement** phase, the participant received an XR device preloaded with the prototype. Over a three-week period, he used the XRMI to compose music. We investigated the instrument's functionality and any usability or creative limitations encountered through interviews with the participant. The study concluded with a **user test designed to evaluate the IML model**'s effectiveness in addressing instrumental control issues identified during the PD process. To assess the usability, learning curve, and effectiveness of the IML integration within the XRMI prototype, we then conducted a semi-structured **exit interview** with the participant.

3.2 Thematic Analysis Method

We transcribed the audio-recorded interviews verbatim, preserving the participant's responses in full context. We employed an inductive thematic analysis [23] to systematically identify, analyse, and report patterns (codes and themes) within the transcribed data collected for the conducted interviews. Initial coding was carried out by one coder, reading through the transcripts and identifying recurrent ideas. We operationalised the emerging codes into distinct themes. These themes were then compared across the entire dataset to ensure alignment and avoid redundancy. We then linked the themes back to the broader research questions and objectives to understand their significance. This included a detailed analysis of each theme, interpreting its significance and how it connected to the overall instrument design.

4 Design

In this section, we describe the results of the PD sessions, which informed the design of the XRMI. We cover the three session types chronologically.

4.1 Exploration and Definition Phase

In the first two exploration sessions, we conceptualised the XRMI by reviewing both physical and virtual musical instruments with a focus on keyboard-like interfaces. Our functional design objectives were informed by discussions about the design principles for XRMIs from [24] (the importance of feedback, mappings and low latency; making use of existing performer skills; consider HMD ergonomics, especially the weight of the devices). We defined the following FDOs:

FDO_1: Design an intuitive interface inspired by traditional keyboard instruments, but offering extended functionality. This includes continuous pitch control, overcoming the limitation of discrete keys found on

standard keyboards.

FDO$_2$: Create an interface that facilitates improvisation through an isomorphic layout that simplifies transposition of melodies and chord progressions.

FDO$_3$: Implement expressive controls for individual notes, allowing for manipulation of velocity, pitch and timbre to enable nuanced musical expressions tailored to the performer's intent.

Given the musical background of the participant, we focused on developing a keyboard-like XRMI (*FDO*$_1$). Traditional keyboard interfaces lack the ability for continuous pitch control, limiting the expression of glissandos, bends, and other playing techniques. One particular problem with traditional keyboards identified by the participant was the non-invariance for transposition of melodies and chord progressions. Keyboard players need to learn different fingerings to play the same melody in different keys, which can be a barrier for improvisation. To address this, we aimed to structure our instrument along an isomorphic layout in two dimensions *FDO*$_2$. Based on his experience with MIDI polyphonic expression (MPE), the participant proposed incorporating multidimensional control of note attributes for the XRMI, allowing for per-note pitch bending and greater expressivity (*FDO*$_3$).

In session three, we explored a novel approach to support chord creation. Drawing inspiration from the Tonnetz, a visual representation of harmony developed by Euler (1739), we sought to integrate this framework into the interface. This innovative approach would allow users to visually understand pitch relationships, enabling intuitive chord construction and individual note selection within a single interface (*FDO*$_1$ and *FDO*$_2$). While the Tonnetz has influenced the design of isomorphic digital instruments like the Linnstrument, C-Thru Axis-49, and Exquis, its application within an XR environment is, to our knowledge, a novel approach.

4.2 Design Solution

To respond to the overarching and functional design objectives, we set out to design a compact (*ODO*$_2$) keyboard-like (*FDO*$_1$) XRMI suitable for use on a tabletop surface (*ODO*$_3$). This XRMI utilises the *3,7,4-Tonnetz* layout (*FDO*$_2$) for intuitive note and chord production. Figure 1 illustrates this layout through a subsection of the interface. The layout makes the keyboard isomorphic with the property of transpositional invariance, i.e. any given sequence and/or combination of musical intervals has the same shape when transposed to another key. Each node on the Tonnetz grid corresponds to a musical note. We designed the nodes to be sensitive to pitch, dynamics, and timbre changes via control gestures. Chords are obtained through combination of several nodes, or by "playing" the triangles between nodes through custom control gestures described below. Using two hands, a large variety of chord qualities, extensions and voicings can be achieved. We further incorporated portamento effects through sliding movements along the lines connecting nodes, enhancing expressive capabilities (*FDO*$_3$) and enabling microtonal performance. The control of the instrument and its expressive attributes is facilitated by AI (*ODO*$_1$). To address the challenge of playing traditional scales on the 3,7,4-Tonnetz layout due to the large note intervals between nodes, the participant advocated for adding "bridge nodes" to the layout. These intermediate nodes would simplify melody execution across scales and facilitate easy transpositions (Figure 1).

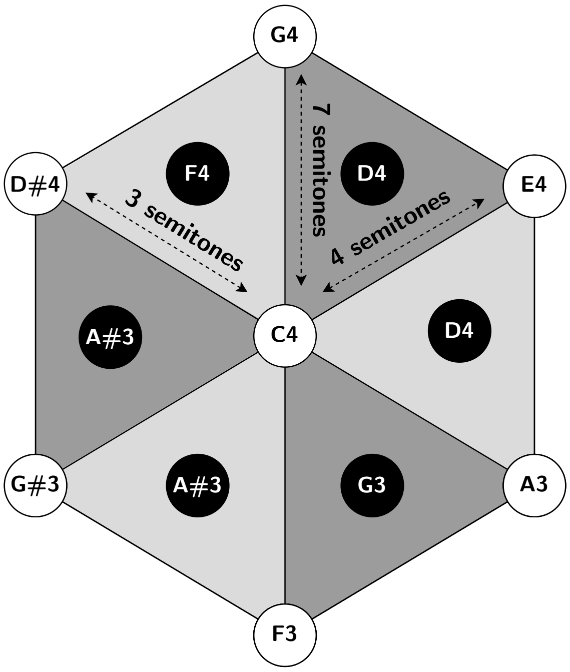


Fig. 1: Schematic rendering of a 3,7,4-Tonnetz section. Triangles form minor (light grey) and major (dark grey) triads. Bridge notes are black nodes.

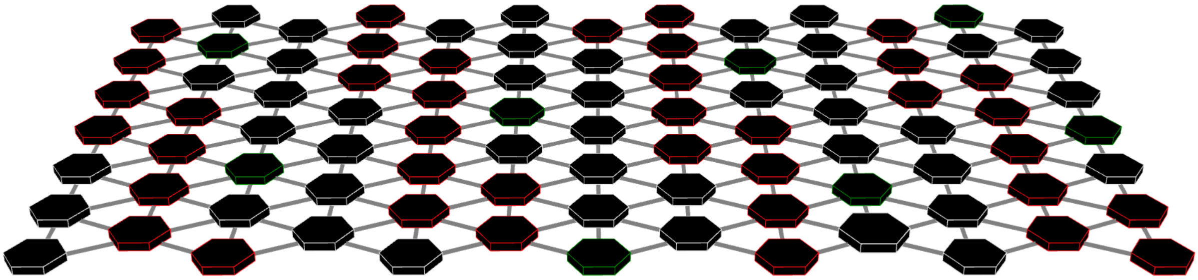


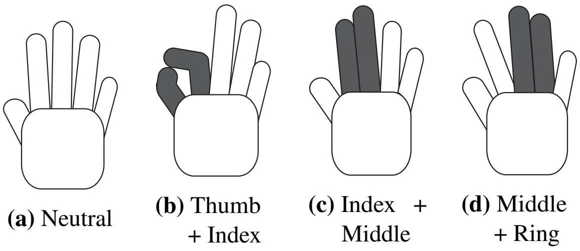
Fig. 2: 3D model of the Netz instrument in XR.

4.3 Making Phase

After the exploration & definition phase we started with the technical implementation of the Netz prototype[1]. We used Unity3D targeted at the Oculus Quest 2 XR device, featuring video passthrough. The main interface, shown in Figure 2, consists of hexagonal nodes arranged in the previously discussed 3,7,4-Tonnetz structure and covers a pitch range of four octaves from A2 to C7. By default, the nodes are set 7 cm apart, maintaining the interface within the user's field of view at a viewing distance of 70 cm (*ODO_2*).

Collision detection mechanisms (computational problem of detecting an intersection of two or more spatial objects) between the player's tracked hand and the interface elements—nodes, lines, and triangles—drive the musical interactions. The system specifically detects fingertip collisions to minimise accidental activations of musical elements through other parts of the hand.

During initial testing, it became evident that the XRMI required a method of differentiating between playing bridge notes (nodes at the centre of triangles) and chords (triangles per-se). We developed a set of specific hand pose configurations to address this, as illustrated in Figure 3. Depending on the pose, either bridge nodes or chords would be triggered upon contact. The poses include open hands (neutral) for playing bridge notes and finger pairings for playing triads in various voicings. Paired index and middle fingers generate the first inversion, and middle and ring fingers yield the second inversion. We opted for these poses as they allow players to create chords and use their remaining fingers to add more notes (extending the basic voicings) by activating additional nodes on the XRMI interface.



(a) Neutral **(b)** Thumb + Index **(c)** Index + Middle **(d)** Middle + Ring

Fig. 3: Hand pose configurations in the XRMI. The relevant paired fingers are shaded in grey.

Initial testing with a prototype rule-based algorithm, designed to classify hand poses based on fingertip distances, resulted in frequent unintended activations of either bridge nodes or chords. This indicated the inadequacy of simple rule-based approaches. Discrepancies in hand sizes among different users (developer and participant) further complicated the interaction, often requiring adjustments to the algorithm.

4.4 Interactive Machine Learning Model

To address this issue, we explored the use of deep neural network (DNN) models to improve gesture recognition accuracy (*ODO_2 & FDO_3*). We developed a custom IML pipeline integrated with the XRMI, shown in Figure 4. This pipeline allows performers to capture their own hand gesture data, train the model using this data, and run the trained model within the immersive application. The data collection involves playing 100 notes or chords for each hand pose configuration, generating around 3000 samples per pose in 1-2 minutes. These samples are then sent to an external server for training, and the trained model is returned to the XR device for real-time execution. This separation of concerns (design principle for separating computer

---

[1] While the participant was actively involved in the design of the instrument, he did not participate in the technical implementation of the XR application.

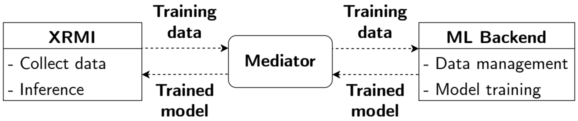


Fig. 4: Conceptual model of the IML pipeline

applications into distinct sections) allows us to dynamically adapt frontends (XR instruments) and backends (compute machines with GPUs) as necessary.
The neural network, designed in PyTorch, comprises approximately 50,000 parameters and uses eight inter-finger distance measurements per hand as inputs. This setup helps the model remain invariant to hand rotations, provided the hand tracking system accurately detects the hands. The hand joint data is processed at 50 frames per second to classify hand poses, which then determine the interaction configuration with Netz.

The effectiveness of the IML model was evaluated through a user test comparing task performance with and without the model activated. The test focused on the accuracy of playing note sequences and chord progressions, specifically measuring the number of incorrect notes or chords played. For a quantitative analysis of the user test results, please refer to [25]. Here, we focus on the subjective evaluation of the model and its effect on musical performance with the instrument through an analysis of the interviews conducted during the peformance & refinement sessions of the study, including the final exit interview.

4.5 Performance & Refinement Phase

After the making sessions, we provided the participant with an XR device preloaded with the instrument prototype[2]. He kept the device for three weeks to practice and compose music with the XRMI. Ten days into the performance & refinement sessions, we gave the IML model to the participant. For the following ten days, he focused on playing the instrument with the integrated IML model, actively comparing this experience to his initial ten days without the model's assistance.

4.6 Thematic Analysis Results

Five key themes emerged from our analysis of the interviews conducted during the final three sessions and the exit interview. Code occurrences are reported in brackets.

1. Technological errors and IML model impact (26)
The discussions frequently highlighted frustrations with hand tracking inaccuracies and unintended note activations during performances. These issues significantly hindered user experience, especially when the IML model was not active. This created a disconnect between the imagined music and the instrument capabilities (*"I was never able to properly recreate what's going on in my head on the interface... It would mis-trigger so badly."*). This led the participant to adapt his playing style (*"I played whatever did not mis-trigger at all."*). After introducing the IML model, it proved effective in mitigating accidental note/chord activations (*"I could tell if the model was on, it would trigger fewer random [notes]", "I think it's great how much I can control the instrument with the model running."*). While this had a positive effect on the participant's confidence with the instrument, he also stated that he *"sometimes was expecting too much of [the model]"*.

2. Exploration and learning experience (15)
The participant highlighted the exploratory nature of learning Netz, comparing it to the process of mastering a completely new instrument. This theme encompasses the challenges and experiences of adapting to a new interface, specifically the isomorphic note layout (*"Surprisingly it wasn't too hard to remember the positions of the [nodes]. It wasn't confusing to find notes here and there, it was very explorative."*).

3. Physical and interface design concerns (12)
Comments highlighted challenges with the size of the instrument and difficulty reaching certain areas (*"The interface was too large to play with so I could never get to the corners."*). The participant suggested that an ideal interface would be smaller, allowing for effortless access to all interface elements, eliminating the need for excessive stretching of the arms.

4. Control and interaction (9)
The participant appreciated how Netz's control features allowed them to shape the sound dynamically, specifically the instrument's responsiveness to their hand gestures (*"The timbre controls that were mapped to the wrist movements... That added so much to the whole experience, for me it was very easy to control."*). Beyond appreciating the control features, the participant actively discussed the value of the tactile feedback when using the instrument on a tabletop. He highlighted how

---

[2] Please refer to `https://netz-website-xi.vercel.app/` for a demonstration of the final version of the Netz instrument.

effectively he could control Netz through adjusting the placement of the interface on the table.

5. Flow (6)
Despite the technical challenges, the participant described moments of connection with the instrument (*"There were multiple moments where things clicked throughout the weeks."*). However, these moments were *"[...] short lived [...]. I wish I could have kept going on with an idea for 10 or 15 minutes, but sometimes [accidental note activations] would occur and it would take me out."*

5 Discussion

The results from our participatory design sessions indicated that the integration of an IML model into the XRMI led to a perceived improvement in gesture recognition accuracy. This was evident in the performance & refinement sessions, where the participant experienced fewer unintended note/chord activations and reported a more enjoyable, less interrupted interaction with the instrument. These improvements highlight the importance of adaptive and user-specific AI systems in complex interaction environments like XR, where standard gesture recognition algorithms may fall short.
The use of a deep neural network (DNN) to process hand pose data directly on XR devices introduces a scalable solution that can be adapted for other applications requiring real-time gesture recognition. This approach could be particularly beneficial in other domains such as XR-based gaming, virtual training simulations, and rehabilitative technologies.

Despite these positive outcomes, several problems remain. The primary issue were accidental note activations due to technological errors pertaining to hand tracking. These errors still occurred after the integration of the IML model, albeit less frequently. Our participant perceived a limitation in the expressive potential of the Netz instrument. This can be attributed to the inherent tension between the instrument's physical size and the accuracy of hand tracking technology (themes 1, 3 and 5): larger interfaces reduce accidental note activations but require more space and consequently, more body movement. Conversely, smaller interfaces minimise physical strain but current hand-tracking limitations lead to frequent accidental activations. These errors underline the open challenges in achieving robust interaction within XR environments and suggest the need for further optimisation of the hand tracking algorithms, as well as hardware improvements.

Additionally, our study's limitation to a single participant reduces the generalisability of the results. While the in-depth, longitudinal participatory design approach provided valuable insights into the design and functionality of Netz, future studies should include a larger and more diverse group of participants to further validate and refine the findings.

6 Summary

The Netz XR musical instrument provides a compelling case study for AI integration within interactive music XR systems. We showed that our interactive machine learning approach can effectively improve the musical interaction experience in XR by reducing interruptions during practice and performance. As XR technology continues to evolve, the insights gained from this study can inform future developments in the field, pushing the boundaries of what is possible in XR interaction design and musical expression. Future advancements in technology and further empirical research are essential to overcome existing limitations and fully realise the potential of AI-enhanced musical instruments in XR environments.

7 Acknowledgements

This work was supported by the UKRI EPSRC AI & Music Centre for Doctoral Training (EP/S022694/1). We would like to thank our participant, Clifford Moses Manasseh, for all the insightful conversations and his contribution to this study.

References

[1] Zellerbach, K. C. and Roberts, C., "A Framework for the Design and Analysis of Mixed Reality Musical Instruments," in *NIME 2022*, 2022.

[2] Benford, S., Schnädelbach, H., Koleva, B., Anastasi, R., Greenhalgh, C., Rodden, T., Green, J., Ghali, A., Pridmore, T., Gaver, B., Boucher, A., Walker, B., Pennington, S., Schmidt, A., Gellersen, H., and Steed, A., "Expected, sensed, and desired: A framework for designing sensing-based interaction," *ACM Transactions on Computer-Human Interaction*, 12(1), pp. 3–30, 2005, doi:10.1145/1057237.1057239.

[3] Graf, M. and Barthet, M., "Mixed Reality Musical Interface: Exploring Ergonomics and Adaptive Hand Pose Recognition for Gestural Control," in *International Conference on New Interfaces for Musical Expression*, 2022, doi:10.21428/92fbeb44.56ba9b93.

[4] Rasamimanana, N., “Towards a Conceptual Framework for Exploring and Modelling Expressive Musical Gestures,” *Journal of New Music Research*, 41(1), pp. 3–12, 2012, doi:10.1080/09298215.2011.642391.

[5] Mäki-Patola, T., Laitinen, J., Kanerva, A., and Takala, T., “Experiments with virtual reality instruments,” in *Proceedings of the 2005 conference on New interfaces for musical expression*, NIME ’05, pp. 11–16, National University of Singapore, SGP, 2005.

[6] Zielasko, D., Rausch, D., Law, Y. C., Knott, T. C., Pick, S., Porsche, S., Herber, J., Hummel, J., and Kuhlen, T. W., “Cirque des bouteilles: The art of blowing on bottles,” in *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 209–210, 2015, doi:10.1109/3DUI.2015.7131774.

[7] Hamilton, R., “Coretet: A Dynamic Virtual Musical Instrument for the Twenty-First Century,” in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1395–1395, 2019, doi:10.1109/VR.2019.8797680, iSSN: 2642-5254.

[8] Deacon, T., Stockman, T., and Barthet, M., “User Experience in an Interactive Music Virtual Reality System: An Exploratory Study,” in M. Aramaki, R. Kronland-Martinet, and S. Ystad, editors, *Bridging People and Sound*, volume 10525, pp. 192–216, Springer International Publishing, Cham, 2017, doi:10.1007/978-3-319-67738-5_12, series Title: Lecture Notes in Computer Science.

[9] Çamcı, A., Vilaplana, M., and Wang, R., “Exploring the Affordances of VR for Musical Interaction Design with VIMEs,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 121–126, Birmingham, UK, 2020.

[10] Berthaut, F., “3D interaction techniques for musical expression,” *Journal of New Music Research*, 49(1), pp. 60–72, 2020, doi:10.1080/09298215.2019.1706584.

[11] Deacon, T. and Barthet, M., “Spatial Design Considerations for Interactive Audio in Virtual Reality,” in *Sonic Interactions in Virtual Environments*, Human-Computer Interaction Series, pp. 181–217, Springer, 2022.

[12] Hamilton, R. and Platz, C., “Gesture-based Collaborative Virtual Reality Performance in Carillon,” in *Proceedings of the 2016 International Computer Music Conference*, p. 5, Utrecht, Netherlands, 2016.

[13] Men, L. and Bryan-Kinns, N., “LeMo: Supporting Collaborative Music Making in Virtual Reality,” in *IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, 2018, doi:10.13140/RG.2.2.27033.93289.

[14] Berthaut, F., Desainte-Catherine, M., and Hachet, M., “Interacting with 3D Reactive Widgets for Musical Performance,” *Journal of New Music Research*, 40(3), pp. 253–263, 2011, doi:10.1080/09298215.2011.602693, publisher: Routledge _eprint: https://doi.org/10.1080/09298215.2011.602693.

[15] Moore, A. G., Howell, M. J., Stiles, A. W., Herrera, N. S., and McMahan, R. P., “Wedge: A musical interface for building and playing composition-appropriate immersive environments,” in *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 205–206, IEEE, Arles, France, 2015, doi:10.1109/3DUI.2015.7131772.

[16] Wanderley, M. M. and Depalle, P., “Gestural control of sound synthesis,” *Proceedings of the IEEE*, 92(4), pp. 632–644, 2004, doi:10.1109/JPROC.2004.825882, conference Name: Proceedings of the IEEE.

[17] Fiebrink, R. and Cook, P. R., “The Wekinator: A System for Real-Time, Interactive Machine Learning in Music,” *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010.

[18] Françoise, J., Schnell, N., Borghesi, R., and Bevilacqua, F., “Probabilistic Models for Designing Motion and Sound Relationships,” *Proceedings of the 2014 international conference on new interfaces for musical expression*, p. 6, 2014.

[19] Smith, J., “Human-AI Partnerships in Generative Music,” in *International Conference on New Interfaces for Musical Expression*, 2022, doi:10.21428/92fbeb44.1aaaf7ab.

[20] Tremblay, P. A., Roma, G., and Green, O., “Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit,” *Computer Music Journal*, 45(2), pp. 9–23, 2022, doi:10.1162/comj_a_00600.

[21] Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L., “The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population,” *PLOS ONE*, 9(2), p. e89642, 2014, doi:10.1371/journal.pone.0089642, publisher: Public Library of Science.

[22] Zayas, L. and McPherson, A., “Dialogic Design of an Accessible Digital Musical Instrument: Investigating Performer Experience,” 2022.

[23] Braun, V. and Clarke, V., “Thematic analysis,” in *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*, APA handbooks in psychology®, pp. 57–71, American Psychological Association, Washington, DC, US, 2012, doi:10.1037/13620-004.

[24] Serafin, S., Erkut, C., Kojs, J., Nilsson, N. C., and Nordahl, R., “Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions,” *Computer Music Journal*, 40(3), pp. 22–40, 2016, doi:10.1162/COMJ_a_00372.

[25] Graf, M. and Barthet, M., “Reducing Sensing Errors in a Mixed Reality Musical Instrument,” in *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST ’23, pp. 1–2, Association for Computing Machinery, New York, NY, USA, 2023, doi:10.1145/3611659.3617210.