Following the Master's Hands: Capturing Piano Performances for Mixed Reality Piano Learning Applications

Citation: Labrou, Katerina, Zaman, Cagri Hakan, Turkyasar, Arda and Davis, Randall. 2023. "Following the Master's Hands: Capturing Piano Performances for Mixed Reality Piano Learning Applications."

As Published: https://doi.org/10.1145/3544549.3585838

Publisher: ACM|Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems

Persistent URL: https://hdl.handle.net/1721.1/150654

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Following the Master's Hands: Capturing Piano Performances for Mixed Reality Piano Learning Applications

Katerina Labrou*
klabrou@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Cagri Hakan Zaman*
zaman@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Arda Turkyasar
Mediate Labs
Somerville, USA
arda@mediate.tech

Randall Davis
davis@csail.mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Figure 1: Tracking a pianist's hand movements using optical motion capture.

ABSTRACT

Piano learning applications in Mixed Reality (MR) are a promising substitute for physical instruction when a piano teacher is absent. Existing piano learning applications that use visual indicators to highlight the notes to be played on the keyboard or employ video projections of a pianist provide minimal guidance on how the learner should execute hand movements to develop their technique in performance and prevent injuries. To address this gap, we developed an immersive first-person piano learning experience that uses a library of targeted visualizations of the teacher's hands and 3D traces of hand movements in MR. Seeing the piano teacher's hands while hearing the music is central to developing the novice's musical intuition. We introduced an end-to-end workflow to accurately capture the pianist's technical gestures and align them with the musical score. We recorded pianists playing technical exercises and music pieces. We developed a multimodal performance dataset (MPD) comprising virtual hand models, keyboard (MIDI) recordings and the corresponding music scores, and different visualizations of hand traces capturing movement. Finally, we developed Pianoverse, an MR application to assist piano learning, and performed exploratory user testing with novice piano players to understand the impact of multimodal representations of movement on skill learning. Our initial observations suggest that apprehending the movement traces of a recorded performance over a physical keyboard increases the learner's ability to position their body and hands correctly and to replicate hand gestures while playing from written music. Further research will focus on automating performance data collection and a comprehensive evaluation of the use of leading movement traces in piano learning.

*Both authors contributed equally to this research.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585838

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; • **Applied computing → Sound and music computing**; **Interactive learning environments**.

KEYWORDS

Mixed Reality, Piano training, Motion Capture, Embodied Computing

ACM Reference Format:

Katerina Labrou, Cagri Hakan Zaman, Arda Turkyasar, and Randall Davis. 2023. Following the Master's Hands: Capturing Piano Performances for Mixed Reality Piano Learning Applications. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3544549.3585838

1 INTRODUCTION

Learning to play the piano is a complex task that requires years of practice to master. The piano technique is typically taken up by imitating and receiving guidance from an experienced teacher. Many people do not have access to piano teachers or the ability to financially support yearlong studies. Self-learning methods such as video tutorials and virtual applications have become popular learning alternatives. However, piano experts, stressing the importance of developing a correct technique at an early learning stage, have expressed concern about the effectiveness of these applications. In this study, we propose a novel approach to immersive piano learning that incorporates principles of embodied cognition, learning from others, and learning with the body. We introduce a novel methodology for controlled recordings of pianists and a method to synchronize the recorded performance data. The method involves capturing the hand movements of expert piano players, aligning them with audio from the digital keyboard and the musical score, and reenacting the recorded performances in a virtual environment. Significant technical contributions of this work include the protocol for recording the pianists' performance and the computational workflows developed to bring the data to a consistent format. The resulting multimodal piano dataset (MPD) contains clean data from 6 performances by three pianists, including motion and MIDI data, such as which note is played, how hard the key is pressed, and which finger is hitting the key, aligned with the measures of the music score. We share our performance dataset openly to assist further research. We also present Pianoverse, an application for selective visualization and playback of piano performances in Mixed Reality (MR). Pianoverse is intended to assist novices and intermediate piano players in learning how to move their hands as they play from a written score. Learners can develop their technique in a multimodal learning environment that fosters imitation through close observation of hand and body gestures. The performer's hand movements are projected in three dimensions over the physical piano keyboard while listening to the recorded sound.

2 BACKGROUND

Over the past years, many digital applications have been introduced to motivate and assist piano learning [10, 15]. Interest in MR applications that provide immersive musical experiences and can support music learning, rehearsing, improvisation, and composition is increasing [16, 17]. Current trends in these applications prioritize visual stimuli by lighting up the keys or using a roll visualization of the score [13]. The user follows the visual cues that indicate the sequence of keys to strike on the keyboard for a selected piece of music. In another approach, researchers suggest first-person, full-scale displays of a pianist's virtual hands projected over a physical piano to assist learning. Xiao and Ishii, stressing the importance of physical movement, use projections of pianists' upper body over a grand piano [19, 20] to motivate piano students. Xiao & Ishii's [19, 20] poetic projections of video captures of pianist's performances over a physical piano focus primarily on transmitting expressive aspects of each expert pianist's performance. Gerry et al. introduce a real-time piano training application in MR that projects a video of the teacher's hands over the learner's piano [3]. Visualizing a performance in 3D demonstrating different ways to correctly execute technical movements and providing spatial guidance beyond the pianist's virtual hands has yet to be tested.

Capturing and synchronizing audio and motion from a piano performance is a complex task. It is particularly challenging to capture the hand movements of a piano player due to the fast and intricate finger movements involved in playing the instrument. In addition, the hands are often close to the piano, which causes occlusions and reflections that obstruct the continuous tracking of the fingers. Several methods have been used for capturing piano players' hand movements. One approach focuses on using depth cameras, such as Microsoft Kinect and Intel RealSense, and video-based tracking of hands [8, 9, 11] However, this technique does not provide adequate spatial and temporal resolution for the task. Other disadvantages are the occlusion of hands and fingers and variation in hand sizes and textures [2, 12].

Optical motion tracking with markers has also been used for studying the hand movements of pianists due to its advantages over the methods mentioned above [1, 2, 4, 6, 12]. Marker tracking involves attaching small reflective markers to the hand and capturing their movements using high-speed infrared cameras. The main advantage of marker-based hand tracking over depth cameras and image-based algorithms is that it provides highly accurate hand tracking because the markers are directly placed over the hand. The higher accuracy of motion capture allows studying anatomical features of the pianist's hands. For example, Tits et al. use motion capture and manually align it with MIDI data to analyze pianist's gestures [18]. Marker-based systems are more robust to occlusions and hand shape and texture variations compared to depth cameras and image-based algorithms[5]. To maintain an uninterrupted line-of-sight for optical hand motion capture, researchers exclusively use digital keyboards because upright and grand pianos obstruct the hands.

In the study, we employ optical motion tracking to capture piano hand movements. We augment the data recording workflow with information obtained from video, audio, and MIDI. To our knowledge, MPD is the first dataset that combines motion capture, MIDI, and digital music scores and is suitable for reenacting a real piano performance in a digital environment.

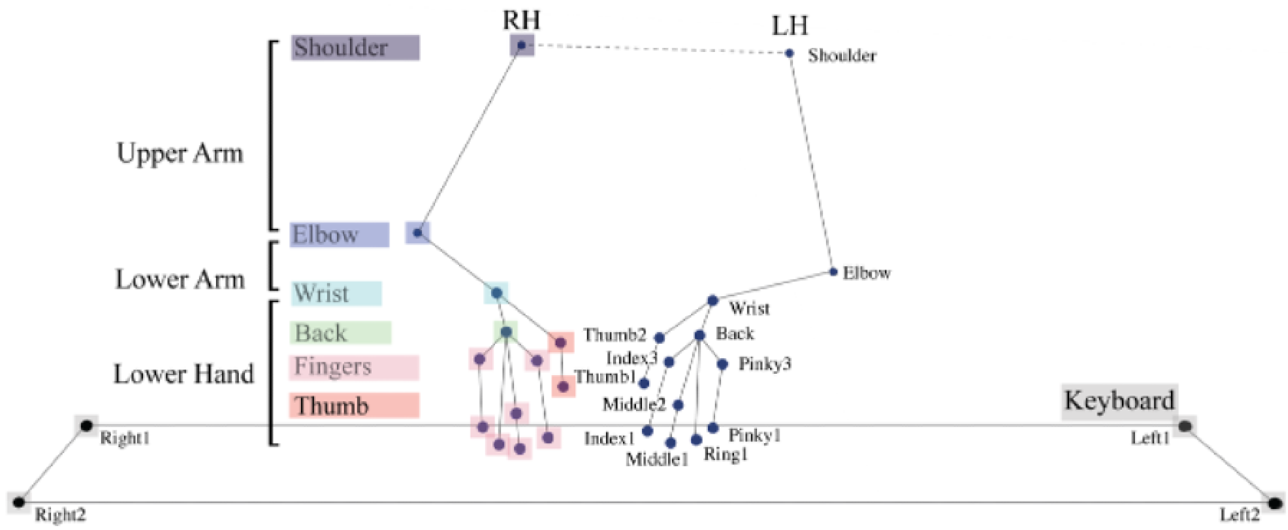

Figure 2: Thirty tracked markers on the keyboard (4) and the pianist's body (13/hand). Hand parts are shown on the right hand, and marker names on the left hand and the keyboard.

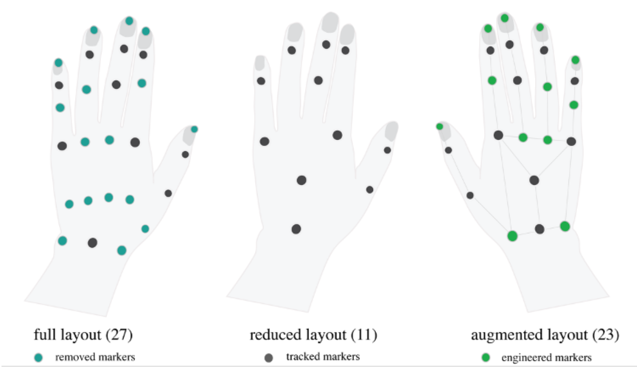

Figure 3: (a) Full-marker layout is frequently used in optical hand motion tracking. (b) Reduced marker layout used in the study. (c) Augmented Marker Layout including generated markers

3 METHODOLOGY

3.1 Experimental Setup

We followed a multimodal approach to record performances using motion capture and video and audio using cameras and environment microphones. The pianists' interactions with the touch-sensitive Yamaha P-45 digital keyboard were recorded through MIDI (Musical Instrument Digital Interface). For motion capture, we used the OptiTrack system. This 27-camera system consisted of 26 infrared Prime 13 cameras and a Prime Color camera for synchronized video capture.

We carefully selected the data to be recorded and determined the specifications and setup for each sensing device. The digital piano keyboard was placed in the center of the cubic recording field of the cameras to maximize tracking. The cameras for the video recordings were positioned to not obstruct the views of the wall-mount cameras for hand tracking. The additional cameras allowed us to visually record the pianists' hand movements in great detail, providing a comprehensive representation of the performance.

3.1.1 Motion Capture. Reflective markers of 7.9 mm were used to track 13 anatomical landmarks of interest, 11 on the pianist's lower hand and two on the upper hand, elbow, and shoulder. Four markers of 9.5 mm were used for tracking the rectangular keyboard boundary. The motion capture system was not sufficient to track a full-marker layout (see Figure 3(a)) due to the considerable distance of wall-mounted cameras in the space. To obtain cleaner data, we used a reduced 11-marker layout shown in Figure 3(b), which we developed to deal with inconsistencies in tracking caused by the markers at close distances to one another. We experimented with other methods, including MoCap gloves and a lightweight 3-camera Optitrack array. Despite the drawbacks related to the scale incompatibility of the wall-mounted cameras, the 26-camera setup provided the best spatial and temporal resolution of the tracked motion. Figure 2 shows the final layout of markers and corresponding labels, including four stationary markers for tracking the keyboard.

3.1.2 Video Recording. The video was recorded with two cameras at different angles. A medium-distance front view of the hands on the keys and the body of the pianists were captured with a digital camera on a tripod in front of the keyboard. A top view of the hands and keyboard was captured with an RGB Prime 13 high-frequency camera synchronized frame by frame with the OptiTrack system.

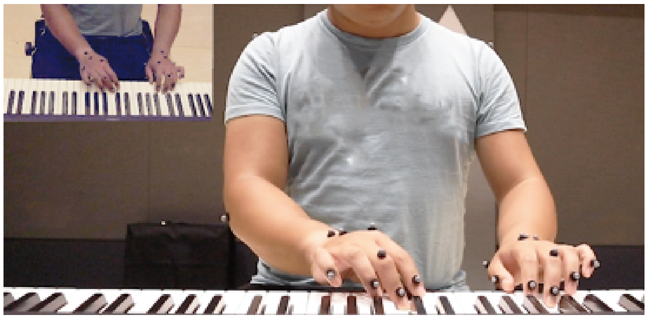

Figure 4: Video recording of a performance. Videos with sound were used to discern the finger used per keystroke.

We used videos to associate played notes with individual fingers. The numbers corresponding to each finger, also called fingerings in music, were noted on the digital score and used to disambiguate finger motion during data processing, as shown in Figure 4.

3.1.3 Audio Recording. We recorded the environment sound with a microphone placed next to an external loudspeaker. The audio recording was in sync with the motion data, permitting an accurate synchronization of the motion and MIDI data at a later stage of data preprocessing by synching the music start time in both audio and MIDI. MIDI data was acquired from the keyboard's USB-to-HOST interface using Garage Band, a music creation software. The resulting MIDI data contain information on keystrokes - specifically, when a key is pressed, how, and for how long. A velocity value corresponding to each key press captures the intensity of a critical press in values ranging from 0-127.

3.2 Data Collection

We recruited six expert pianists (4 male / two female, median age 24) to participate in an hour-long recording session. All of the participants had at least ten years of experience. Before the study, the participants were informed about the performance tasks and requested to practice the music repertoire. At the beginning of each session, we placed the motion capture markers over the pianist's hands and photographed their hand in resting position over a millimetric measuring grid. This measurement allowed us to validate the marker positions in the dataset and provided a reference point for determining the scale and positions of fingers in the video recordings.

The music repertoire comprised a set of technical exercises and two select music pieces. The repertoire is formulated so that most players would be familiar with it before the study, making it easier to prepare within a week or two and record it during a one-hour session. The selected technical exercises were C major and C minor scales and variations and Prelude in C Minor, BWV. 999, J. S. Bach. The pianists performed the technical exercises consecutively, with 5-10 seconds breaks between each part. The Prelude was captured in two separate recordings. Each pianist played the piece two times, one treating it like a technical exercise and the other playing expressively.

3.3 The Multimodal Performance Dataset

Data sourced separately should be temporally and spatially aligned to provide consistent performance representations. The motion data should be annotated according to the hand landmarks and localized relative to the piano keys. To align the performance with the music score, finding a correspondence between the written and the played notes is required. Data Processing steps are illustrated in Figure 5

Labeling is a routine process in motion capture, where markers are annotated through visual observation before the motion data are exported from the acquisition software. In this process, markers may be misreported, reported in the wrong positions, or missing. The problem of misreported markers was partially solved during manual labeling. It was assessed visually when a marker was in the correct position before a label was assigned. To remove noise from the data, we developed a process for removing outliers from the reported data points. The issue of missing markers is also solved within the same process by interpolating the data.

3.3.1 MIDI-Motion Synchronization. We used the audio recorded in the motion acquisition software to synchronize the motion and MIDI data because motion and audio recordings were already in sync. By analyzing the audio signal, we defined the start time of MIDI to be the same as the start time for music.MIDI contains note events corresponding to a pitch with an associated start time and duration. Using this information, we created a structured representation of the MIDI data with a temporal resolution of 240 frames/sec, the same as the motion data. We adjusted the MIDI start time to match with motion. We incorporated the fingering information from the video into the digital music score, MusicXML [7], which already had appropriate data representations for fingerings.

3.3.2 Music Score and Performance Alignment. To temporally align the fingering information with the MIDI data, a correspondence between the notes in the music scores and the notes in the MIDI should be determined. It is an essential step because the actual performance of the pianist almost certainly deviates from the written score. We used an alignment algorithm developed by Nakamura et al. [14], based on merged-output Hidden Markov Models, which can deal with reordered notes due to voice discordance. The alignment allowed us to integrate data on fingerings with the played notes. Extending the structured representation of the midi data, we annotated each note with the hand and finger that plays the note and the measure number.

3.3.3 Motion Localization. To accurately locate hand movements on the piano keyboard, we used keyboard markers as a constant reference across all performances. We also created a spatial model of the keyboard using measurements from the 3D model of the Yamaha P-45 found online. This allowed us to analytically define the boundaries of the white and black keys and the heights of the keys in the digital space, enabling us to visualize the motion data in a 3D virtual environment. To identify individual keystrokes in finger motion, we combined touch-MIDI data, which provides a timestamp for keystrokes in time, and fingering data to search motion locally to identify the motion boundaries. This approach, which required the alignment of motion, touch, and MIDI data, allowed us to accurately identify the keystrokes in the finger motion.

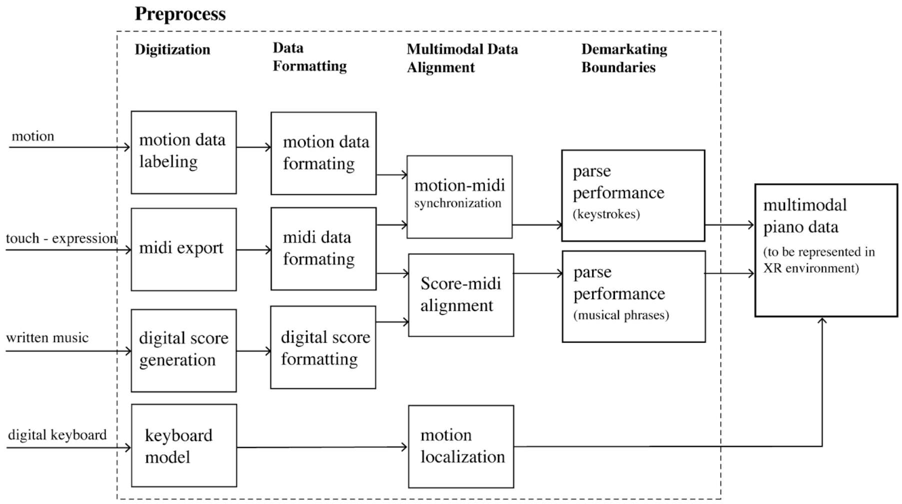

Figure 5: Data Processing System Overview. We sourced information from four different channels: Motion Capture, MIDI, Digital Music Score (MusicXML), and a 3D model of the keyboard used in the experiments. The final Multimodal Performance Dataset is created through formatting, synchronizing, and aligning these channels.

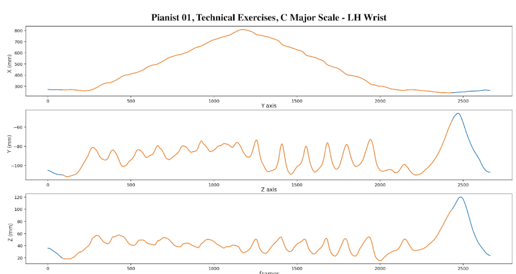

Figure 6: Motion Boundary Detection. Wrist motion close-up of for 01 Technical, C Major Scale. Note duration frames 88-2202. Motion buffer from 0-88 (motion start) and 2434-2667 (first data point with local z-minimum), indicated with blue

3.3.4 Motion Boundary Extraction. To complete the MPD, we identified two types of motion events: individual keystrokes and whole parts (scales, arpeggios, or octaves in the Technical Exercises, and or the entire piece in Bach) in order to separate them from moments of inactivity, such as the pianists' preparation time. We determined keystroke boundaries by searching for local maxima in motion on the finger trajectory around the temporal note event boundaries as depicted in Figure 6. We also used wrist motion to determine the motion boundaries of each part of the Technical Exercises played by the pianists by looking for local minima in the z-axis of the wrist's trajectories. This approach is effective because the movement boundaries correspond to parts that start from a resting position of the hands on the keyboard.

The completed dataset consists of four time synchronized parts per musical score: MIDI, Motion Data, MusicXML, and Segments, which include fingering, measure, and motion boundary information for every note in the score. Figure 7 shows how these data synchronously visualized in a performance. See supplemental materials for a sample data from MPD.

4 PIANOVERSE: A MIXED REALITY PIANO LEARNING EXPERIENCE

We developed a mixed-reality application called Pianoverse. This application projects the recorded data over a real piano and presents the recorded performances in space, allowing the user to observe intricate details of piano playing in realistic conditions. Additionally, users can see their hand movements while a performance is played, allowing them to imitate the pianist's movements. We developed the Pianoverse application using the Unity game engine and deployed it to an Oculus Quest Pro XR device. The application consists of three parts: (1) Alignment: the user is instructed to position their fingers to pre-determined locations of reference markers to align the digital recording with the physical keyboard correctly. (2) Tutorial: Once the alignment is done, the user is presented with prompts explaining the visualization components. (3) Experiments: the user is presented with several performances that show the pianist's hands, the written score, and visual indicators of the keystrokes and hand movements.

4.1 System Setup and Procedures

4.1.1 Aligning the Keyboard. The alignment process consists of two stages. In the first stage, the user locates the leftmost key on the physical keyboard and places their left index finger right next to it. While their index finger is in place, the user makes a pinching gesture with their right hand, fixing the position of the first reference point. The user follows the same step for the right-hand side to fix the position of the second reference point. When

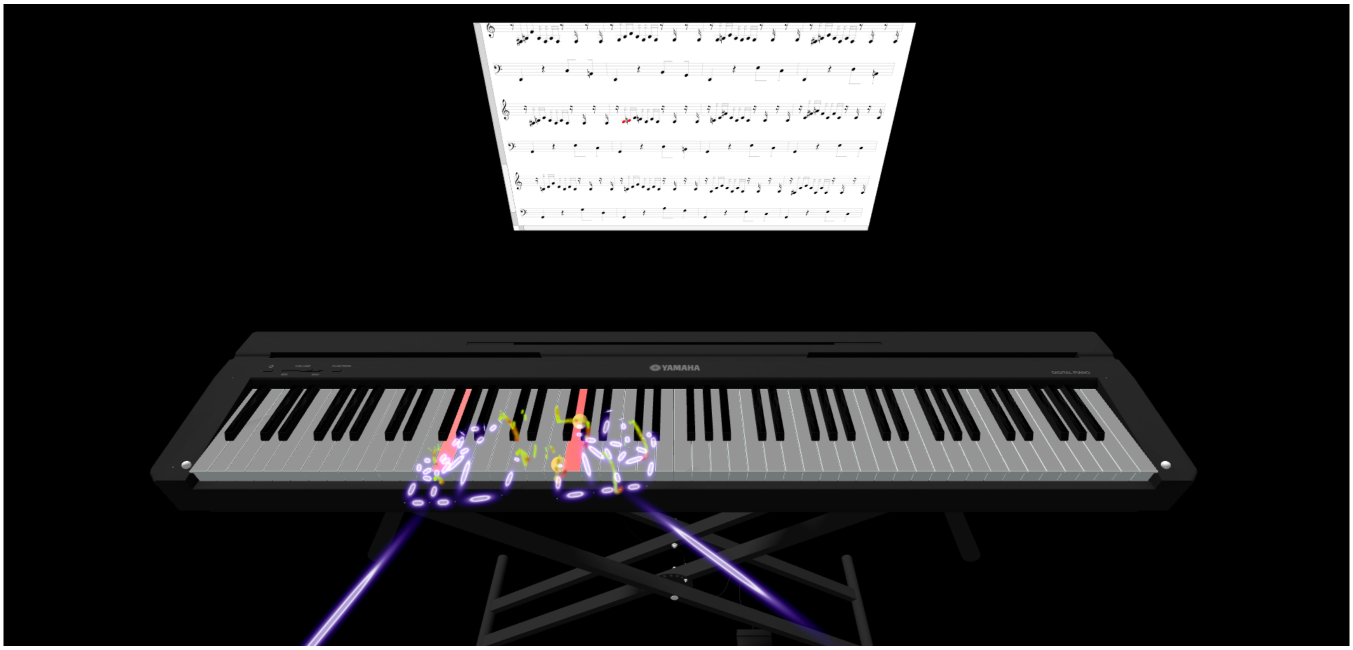

Figure 7: A still from a performance. Notes that are being played are highlighted in red both on the musical score and the keyboard. The fingers pressing the keys have circle indicators on them. Several joints have trailing indicators that visualizes the movement.

the two reference points are set, the digital coordinate system is transformed to align two predetermined markers with the markers located by the user.

4.1.2 3D Hand Representation. We represented hands as semi-transparent joints between augmented marker layouts, which included shoulders and elbows. We decided to give the user an unobstructed view of the keyboard and all fingers. Because the hands were aligned with the digital piano, they more or less overlapped with the user's body.

4.1.3 Movement Traces. The movement traces visualize the hand movement over several seconds so the user can see the gestures, especially the planned hand movements while watching the performance. Planned hand movements are not annotated in the musical score but play an essential role in the performance. Because they show the user the next planned hand movement in real-time, these traces are also helpful for practicing a score concurrently with the recorded performance. To test this aspect of movement traces, we also implemented leading movement traces that precede the performance by one second.

4.1.4 Keystroke Indicators. While the fingers approach the keyboard's surface, it is hard to visually identify if there is a keystroke or a simple touch that does not produce sound. In order to show actual keystrokes to the user, we developed two distinct hit indicators. The first indicator highlights a key for the duration of a keystroke. This indicator simply represents the musical score over the keyboard. However, combined with the 3D hand representation, it also allows the user to understand distinct finger movements in the performance. The second kind of indicator highlights fingertips for the duration of a keystroke. This indicator is also helpful in identifying the correct finger for a keystroke, especially when hands make large movements while engaging the keys.

4.1.5 The Musical Score. The MR application shows the written score as the performance is being played. We highlight the notes as they are being played.

4.2 Preliminary Tests and Observations

Two main aspects of our study require further evaluation. Our data collection and alignment procedure produced empirically robust results.However, the initial dataset is limited to a small repertoire and is not inclusive of all possible hand movements involved in piano playing. In the next steps of study, the authors will expand the dataset with scores that might introduce additional challenges for data collection.

We will further develop Pianoverse for a comprehensive evaluation to understand the impact of the MR experience on piano learning. As a first step, we performed a small pilot study to test the usability of our system and to receive feedback from prospective users. One novice and one experienced player were invited to participate in the pilot study. Before each session, the users were informed about the purpose of the application and were asked to follow the prompts for aligning the keyboard and watching the recorded pieces. The tutorial informed users about different aspects of the interface before the music pieces were displayed.

The alignment procedure was mostly successful despite it's dependence on correct placement of fingers on the respective sides

Figure 8: A still from Pianoverse Mixed Reality application. The subject is attempting to replicate the movements projected on the keyboard. See the supplemental video for an overview of the system.

the keyboard. One user initially failed to align the keyboard because the device localization was reset due to an external error causing digital keyboard model to be misplaced. Upon aligning the keyboard and following the tutorials, both users were comfortably engaged with the recorded performance. Although they were not instructed to do so, both attempted to imitate the movements: They noticed the posture indicated by shoulder and elbow joints and attempted to correct their posture and placed their hands over digital representations.

The speed of the Prelude was a challenge for both subjects. The novice player could not follow the keystrokes in that piece, but they could replicate the parts of C minor and C major scales after watching the performance. The experienced player was better at replicating the parts of the Bach score, although they requested a rewinding feature to repeatedly practice specific parts of the score. Both players remarked that seeing the movement trails allowed them to understand the musical score better. The experienced player said that the onset of the trail was so effective that they automatically followed them to move their hands. While these observations require a comprehensive user study, they indicate an increased engagement and embodied understanding provided by the mixed reality experience.

5 DISCUSSION AND FUTURE STEPS

The contributions of our research include (1) the multimodal performance dataset (MPD) and (2) the technical workflows used to capture and segment musical performance according to the music, as well as (3) a novel approach to immersive instrument learning based on seminal leading traces of bodily movement implemented in (4) Pianoverse.

First, the multimodal performance dataset (MPD) provides a complete visualization of a piano performance consisting of synchronized movements, keystrokes, audio, and written scores. This dataset enables deeper investigations into musical performance and novel applications facilitating embodied, hands-on learning. Second, we contribute a method for segmenting the performances for playback that takes advantage of the aligned music score and keyboard data as well as the recorded hand motion trajectories. The increased control over the recorded motion and MIDI data enables customizing the students' learning experience.

Third, the virtual hands of the pianist and the guiding traces on the physical keyboard aid the learning process by capturing the piano technique and directing the learner's hand posture and movement. The 3D traces enhance the pupil's perception of motion in a traditional piano lesson. The depicted trajectories of bodily movement, suggest a whole new direction in using 3D projections of the master's hands in an immersive setup for skill learning. Pianoverse is the first step in this direction. We employed motion traces as visual cues for piano learning, explored different ways for visualizing motion information and carried out user testing to evaluate their effectiveness in piano practice.

Overall, our work addresses the difficulty of representing important aspects of manual technique, a task that is challenging both in physical and virtual skill learning. Our goal is to further investigate the role of embodiment and spatial perception in conveying the intricacies of skilled movement. We intend to extend the usability of Pianoverse by customizing the experiences for the piano learner by incorporating a larger performance library catering to different hand sizes. We also want to provide feedback on the learner's performance and a way to compare their hand movement to that of the pianist.

We will perform more user studies and comprehensively evaluate the system components. To tackle the limitations of our current approach, we will further automate the process of data collection. Some interesting directions include training machine learning algorithms to label markers and extract fingering information without video recordings.

ACKNOWLEDGMENTS

The authors would like to thank MIT Professor Terry W Knight for her advice and support for this research. The authors also thank the MIT Immersion Lab and the pianists who participated in the studies for their valuable contribution. This research was partly funded by George & Marie Vergottis Foundation.

REFERENCES

[1] Simone Dalla Bella and Caroline Palmer. 2011. Rate Effects on Timing, Key Velocity, and Finger Kinematics in Piano Performance. *PLOS ONE* 6, 6 (06 2011), 1–10. https://doi.org/10.1371/journal.pone.0020518

[2] Shinichi Furuya and Eckart Altenmüller. 2013. Flexibility of movement organization in piano performance. *Frontiers in Human Neuroscience* 7 (2013). https://doi.org/10.3389/fnhum.2013.00173

[3] Lynda Gerry, Sofia Dahl, and Stefania Serafin. 2019. ADEPT: Exploring the Design, Pedagogy, and Analysis of a Mixed Reality Application for Piano Training. In *Proceedings of the 16th Sound and Music Computing Conference SMC 2019 (Proceedings of the Sound and Music Computing Conference, Vol. 1)*. Sound and Music Computing Network, 241–249. https://doi.org/10.5281/zenodo.3249333 16th Sound and music computing conference, SMC2019 ; Conference date: 28-05-2019 Through 31-05-2019.

[4] Werner Goebl. 2017. Movement and touch in piano performance. *Handbook of human motion* (2017), 1–18.

[5] Werner Goebl, Simon Dixon, and Emery Schubert. 2014. Quantitative methods: Motion analysis, audio analysis, and continuous response techniques. *Expressiveness in music performance: Empirical approaches across styles and cultures* (2014), 221–239.

[6] Werner Goebl and Caroline Palmer. 2008. Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research* 186, 3 (2008), 471–479.

[7] Michael Good et al. 2001. MusicXML: An internet-friendly format for sheet music. In *Xml conference and expo*. Citeseer, 03–04.

[8] D.O. Gorodnichy and A. Yogeswaran. 2006. Detection and tracking of pianist hands and fingers. In *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*. 63–63. https://doi.org/10.1109/CRV.2006.26

[9] Aristotelis Hadjakos. 2012. Pianist motion capture with the Kinect depth camera. In *Proceedings of the Sound and Music Computing Conference*. 303–310.

[10] hellosimply.com. 2001. *Simply Piano*. Retrieved January 18, 2023 from https://www.hellosimply.com/

[11] Hui Liang, Jin Wang, Qian Sun, Yong-Jin Liu, Junsong Yuan, Jun Luo, and Ying He. 2016. Barehanded Music: Real-Time Hand Interaction for Virtual Piano. In *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (Redmond, Washington) *(I3D '16)*. Association for Computing Machinery, New York, NY, USA, 87–94. https://doi.org/10.1145/2856400.2856411

[12] Cheryl D. Metcalf, Thomas A. Irvine, Jennifer L. Sims, Yu L. Wang, Alvin W. Y. Su, and David O. Norris. 2014. Complex hand dexterity: a review of biomechanical methods for measuring musical performance. *Frontiers in Psychology* 5 (2014). https://doi.org/10.3389/fpsyg.2014.00414

[13] Diego Molero, Santiago Schez-Sobrino, David Vallejo, Carlos Glez-Morcillo, and Javier Albusac. 2021. A novel approach to learning music and piano based on mixed reality and gamification. *Multimedia Tools and Applications* 80, 1 (2021), 165–186.

[14] Eita Nakamura, Kazuyoshi Yoshii, and Haruhiro Katayose. 2017. Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment.. In *ISMIR*. 347–353.

[15] S. C. Ng, Andrew K. F. Lui, and Alvin C. H. Kwok. 2015. Easy-to-Learn Piano: A Mobile Application for Learning Basic Music Theory and Piano Skill. In *Technology in Education. Technology-Mediated Proactive Learning*, Jeanne Lam, Kwan Keung Ng, Simon K.S. Cheung, Tak Lam Wong, Kam Cheong Li, and Fu Lee Wang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 103–112.

[16] Graham Percival, Ye Wang, and George Tzanetakis. 2007. Effective Use of Multimedia for Computer-Assisted Musical Instrument Tutoring. In *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education* (Augsburg, Bavaria, Germany) *(Emme '07)*. Association for Computing Machinery, New York, NY, USA, 67–76. https://doi.org/10.1145/1290144.1290156

[17] Sophia Ppali, Vali Lalioti, Boyd Branch, Chee Siang Ang, Andrew J. Thomas, Bea S. Wohl, and Alexandra Covaci. 2022. Keep the VRhythm Going: A Musician-Centred Study Investigating How Virtual Reality Can Support Creative Musical Practice. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 220, 19 pages. https://doi.org/10.1145/3491102.3501922

[18] Mickaël Tits, Joëlle Tilmanne, Nicolas D'alessandro, and Marcelo M Wanderley. 2015. Feature extraction and expertise analysis of pianists' Motion-Captured Finger Gestures. In *ICMC*.

[19] Xiao Xiao and Hiroshi Ishii. 2010. MirrorFugue: Communicating Hand Gesture in Remote Piano Collaboration. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction* (Funchal, Portugal) *(TEI '11)*. Association for Computing Machinery, New York, NY, USA, 13–20. https://doi.org/10.1145/1935701.1935705

[20] Xiao Xiao and Hiroshi Ishii. 2016. Inspect, Embody, Invent: A Design Framework for Music Learning and Beyond. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5397–5408. https://doi.org/10.1145/2858036.2858577