ADEPT: Exploring the Design, Pedagogy, and Analysis of a Mixed Reality Application for Piano Training

Lynda Joy Gerry, Sofia Dahl, Stefania Serafin
Aalborg University, Copenhagen
Department of Architecture, Design and Media Technology
lyn,sof,sts@create.aau.dk

ABSTRACT

One of the biggest challenges in learning how to play a musical instrument is learning how to move one's body with a nuanced physicality. Technology can expand available forms of physical interactions to help cue specific movements and postures. This cueing can reinforce new sensorimotor couplings to enhance motor learning and performance. Using Mixed Reality (MR), we present a system that allows students to share a first-person audiovisual perspective with a piano teacher. Students place their hands into the virtual gloves of a teacher. Motor learning and audio-motor associations are reinforced through motion feedback and spatialized audio. The Augmented Design to Embody a Piano Teacher (ADEPT) application is an early design prototype of this piano training system.

1. INTRODUCTION

This paper presents the Augmented Design to Embody a Piano Teacher (ADEPT) system and explains the motivation for its design to train piano playing. The ADEPT system is a Mixed Reality (MR) application in which students share a first-person, embodied perspective with a piano teacher to facilitate learning the proper finger, hand, wrist, and torso configurations to produce various sounds on the piano. The ADEPT system virtually overlays a video recording showing the teacher's hands on top of the student's own hands into the students head-mounted headset. The ADEPT system is inspired by embodied music cognition, which emphasizes the role of human bodily movement in music perception and performance, and makes muscle memory the main focus of musical training and analysis [1]. This differentiates the ADEPT system from the prevailing approaches which often analyses skill of playing in terms of key press onset and release [2, 3]. Instead, embodied music cognition views piano playing as a nuanced and specialized bodily knowledge [4]. Previous technology-enhanced piano training applications aim to train playing the correct key(s) versus training optimal sound-producing movements [5]. Rather than memorizing each individual note to be played, trained musicians use muscle memory and fine motor skills. Consequently, the ADEPT system is designed to reinforce muscle memory rather than rote learning of symbolic musical notation, using visual and audio perspective-taking as a tool to guide sensorimotor skills development, combined with motion tracking and feedback to enhance musical action cueing.

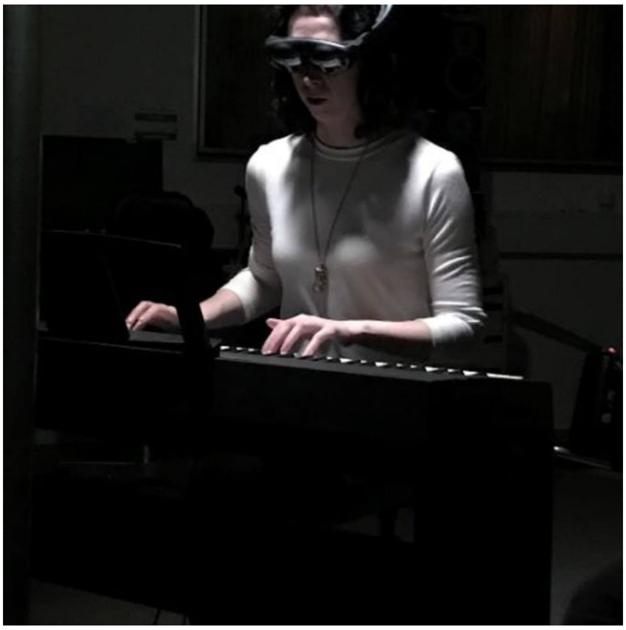

Figure 1. Image showing a user performing piano through the ADEPT system on the Magic Leap headset.

2. RELATED RESEARCH

The pedagogy of the ADEPT system is based on training bodily knowledge and sensorimotor skills. The idea of using virtual overlays to create the illusion of sharing an embodied perspective with the piano teacher is inspired by the instructional technique of having a piano student place their hands on top of the piano teachers hands while playing scales or simple tunes [4]. Moreover, the ADEPT system aims to to train muscle memory for novice piano students, specifically in knowing how to move to produce certain sounds on the piano. ADEPT is geared towards refining the students' experience of their own body and movements towards developing a more nuanced bodily knowledge more akin to that of an expert pianist.

The ADEPT system is inspired by the [**?**, **?**, **?**] framework proposed by Xiao and Ishii [6]. This framework emphasizes musical expression and takes an experiential perspective towards developing musical expertise. The target student users for ADEPT are adult musicians new to the piano. Thus, ADEPT trains piano playing in a very imitation-based way, hoping that this can ultimately help the students to internalize the muscle patterns while playing the piano.

2.1 Music and Piano Trainer Applications

Piano training applications are a subset of technology-enhanced musical education systems [7]. The goal of these systems is to enhance the learning environment for greater individualization, real-time feedback, and multi-sensory cues. Virtual content augmentations gamify the experience of learning, and increase motivation and interest in learning the piano [8]. In general, previous piano training applications have involved three primary components: using visual cues to indicate which key(s) to press, presenting alternative visualizations of musical notation, and increasing sight-reading proficiency.

The primary augmentation for piano training has been to present visual cues on top of piano keys to guide keyboard playing. The training emphasis of these systems is learning to play the correct keys according to the musical score. For example, key press has been indicated through line pointers above the key highlighting the next key to be pressed [8] and a red highlight with incorrect key press [9]. Another variation of training correct keyboard fingering is to show how long a key should be pressed. For instance, HoloKeys presents a green glow on the key at the moment it should be pressed, which disappears as soon as it should be released [10]. Similarly, the P.I.A.N.O. system involves highlighting the current and next key to play, and uses a Guitar Hero approach with dropping lines approaching the keyboard from a far distance to demonstrate which notes to play, and for how long they should be played [9]. In addition to presenting visual cues, some music trainer applications also present auditory cues to piano students using real-time sonification to analyze the student's playing sounds and provide auditory feedback when the student plays the wrong note [11].

Visual cues also help to train students' abilities to imagine some of the sound-movement feedback cycles vital for playing a musical instrument. Specifically, expressive parameters might be difficult for a music teacher to communicate to a student. Hence, for example, the PianoFORTE system helps to train expressive features of piano playing by providing visualizations for dynamics, tempo, articulation, and hand synchronization [12]. Similarly, the Andante system utilizes visual animations of people of various weights walking across the keyboard in sync with the keys as they are played in sequence [13]. This is designed to help the student to visualize another aspect of correct keyboard fingering, which is knowing how hard to press the keys.

It is important to note that the goal of many technology-enhanced piano systems is to compliment and to function as an add-on to traditional piano teaching in-person with a piano teacher. For example, the Piano Tutor [14] is intended to help students practice in between lessons with the teacher, and we intend for ADEPT to function similarly. Additionally, however, technology-enhanced musical education has the possibility to make the learning process easier and more intuitive, as well as to invite new avenues for personal reflection on one's own performance and process. Specifically, one of the major benefits of these systems is that they can help the student to develop online self-analysis skills while playing [14].

2.1.1 Passive Haptic Learning

Another approach of augmentation is to provide haptic information. For example, the MobileMusicTouch is a piano training tecnique haptic feedback with vibration motors inserted at the metacarpophalangeal joints (knuckles) of a glove to help the student to understand which fingers to use to play which keys [15]. One advantage of Passive Haptic Learning (PHL) [16] is that it allows the student to memorize fingering patterns for playing various short music pieces without requiring conscious effort or attention. Indeed, participants were able to retain fingering patterns while wearing MobileMusicTouch even when viewing a film and playing a memory game [15].

Although passively learning finger patterns while attending other stimuli can be convenient, it raises the question as to whether it is the right approach to teach the overall movement control needed for music performance. As noted by Xiao [6], most of the technology-enhanced musical education systems have a focus on the score and the associated errors.

2.2 Mixed Reality Applications

Mixed Reality (MR) environments present virtual overlays and augments that directly interact with the users physical environment, real world objects, and natural movements [17]. Augmented Reality (AR) is a subset of MR, and the two terms are often used interchangeably. The emphasis on MR in this case is to highlight the main interaction with real-world objects (i.e., playing a real piano), which is supported by virtual augments, rather than having the primary interaction be with virtual content within a real environment (i.e., flying a virtual plane that follows the constraints of the real physical environment), as is often the case with popular AR applications. Previous educational benefits have been demonstrated in MR based on its ability to extend embodied actions with high-fidelity multisensory stimuli and real-time feedback [18], specifically by using various types of cueing for different bodily actions.

Previous piano training MR systems that help cultivate higher levels of musicianship focus on facilitating playing a system from memory, and training skills related to musical improvisation. For example, Handel [19] presents a visual overlay of sheet music notation on top of the pianist's fingers while they attempt to play the piece from memory. Similarly, systems like Stanford's Pianolens facilitates learning and rehearsing new music with an interactive sheet music display that imitates a piano roll [20].

More recently, a few MR piano training applications have projected fingers and hands on top of the keyboard or on top of the player's hands to guide piano playing [21]. These systems use various 2D and 3D graphical representation of an experienced pianist's hands and fingers. For example, Teomirn uses the HoloLens display and projects a geometric 3D display of a virtual hand that the student can place their own hand into to follow along while playing. However, the virtual hand is not very realistic, and only roughly helps to guide the student's movements. The ADEPT system instead depicts precise finger, wrist, and upper arm positions, angles, rotations, and movements of the piano teacher for the student to follow. Moreover, these MR designs are mostly focused on helping the student to know which finger to use to play which key.

As noted by Xiao [6] , the primary element that is lacking in all of these systems is an emphasis on the bodily movements of the piano student, and training how to move the body in the correct ways as a focus for the training and technology-enhanced design. MirrorFuge is one exception [21]. The MirrorFugue system presents a projection-mapped video stream of a pianists hands on top of a physical keyboard. Subjects reported that seeing the hands of an instructor was more helpful than screen-based instruction or abstract visual cues (a small dot indicated keys pressed by the expert pianist) [21]. Due to a small sample (5 subjects), the results were not statistically significant but suggest that using the first-person perspective to present the instructors hands from the same egocentric orientation decreased the amount of time that it took for students to learn simple melodies. These results are promising for the development of AR systems focusing on sharing an embodied perspective with an expert pianist who guides the students movements, and this is precisely the target of the ADEPT system. In the next section, we describe virtual embodiment and introduce the concept of augmented embodiment, which is a key design principle behind the ADEPT system.

2.3 Virtual and Augmented Embodiment

Virtual embodiment is a technique used in virtual reality (VR) to create the illusion of becoming a virtual avatar [22]. Virtual embodiment allows users to see and hear from the first-person, embodied perspective of another real person or a virtual character and perform a task together, such as hand-drumming [23, 24] Synchronous stimuli presented to the visual system in VR, combined with the physical body in reality, create various bodily illusions that make the user feel that the avatar body is their own body [25]. This induces strong psychological effects on the user, specifically identification with the avatar body [26]. Taking the perspective of an expert in virtual embodiment studies has been shown to increase confidence and improve performance on related tasks [4, 27–30].

In addition to the strong psychological and learning effects of virtual embodiment, learning a new task from a first-person perspective improves retention of instructional material. For example, sequences of chess movements were more accurately retained when presented in first-person perspective in VR, as compared to exocentric, screen-based perspective [31]. Memory retention is stronger when events are presented from an egocentric versus an exocentric or allocentric point of view [32]. To this end, egocentric VR has been used in memory rehabilitation to increase procedural learning in patients with memory impairments, transferable to real-world environments [33]. Thus, delivering piano instruction in the ADEPT system should support better retention of the finger sequences involved in playing and better performance outcomes in a shorter time.

Another reason why virtual embodiment may increase fine motor skills is that the observation of hand movements elicits motor-evoked potentials in the observer in the specific muscles that would be involved in executing the movements [34]. Research on the mirror neuron system in humans indicates that during the observation of another persons bodily state, the same neural structures are activated in the observer [35]. This effect is even stronger when observing hand movements from a first-person perspective, when one's own hand positions and movements are congruent with those observed [32]. Moreover, visual feedback using video is very common for both piano teachers and students to adjust and adopt better postures for playing, and this has been enhanced with 3D visualization of postural information [36].

We here introduce the concept of Augmented Embodiment, in which the user's point of view is not fully overtaken by that of another virtual avatar or real person, but is instead augmented with a virtual projection of another person's embodied perspective super-imposed on top of one's own. This is the core design feature implemented in the ADEPT system. Augmented embodiment can allow a student to perform and observe an action at the same time, from the same view-point, and in the same way as a teacher with real-time feedback, a phenomenon not possible in physical reality [37].

2.4 Embodied Music Cognition

Adaptive and immersive virtual environments involve new strategies for sensorimotor training and can induce brain reorganization, presently tested therapeutically in clinical populations recovering from stroke [38]. In one such study, a Virtual Piano Trainer system found that adaptive motion feedback providing information about position adjustments in the fingers and hands increased the accuracy and duration of muscle activity, expediting the recovery of these fine motor capabilities. Motion feedback support has previously been shown to increase skills ability and retention while learning technical motor skills [39]. Multisensory feedback can encourage plasticity within the sensory-motor cortex and enhance motor performance [40].

3. SYSTEM OVERVIEW AND DEVELOPMENT

The ADEPT system is programmed in Unity version 2018.1.9f2-MLTP8.1 with C# and uses the Magic Leap Package Manager with Lumin SDK v0.19.0 and Device Driver version 0.94. The setup involves virtually overlaying a video recording from a head-mounted camera worn by the teacher into the student's head-mounted display (HMD), in our case, the Magic Leap 1 (ML-1, see Figure 2). The Unity programming environment used the Media Player example from the Magic Leap Unity SDK and played the video on top of a spherical mesh. The system displays the teacher's hands on top of the student's own hands, and highlights each key when pressed. We used Real-World Reconstruction in Unity, which uses ML-1 depth sensors and spatial computing to track the environment in real-time, which helped for tracking the location of the user's physical piano. AR Registration of the physical piano was accomplished using Vuforia Engine version 8.1 and four fiduciary markers at the four corners of the piano. The cameras on the ML-1 register the four corners of the piano and Vuforia image recognition from piano models registers the approximate positions of the keys.

The system is a a work in progress, and this section reviews the implementation that has been developed so far as well as features that are continuing to undergo further development. For the development we recorded a young piano teacher with 15 years of experience who had taught piano for 3 years.

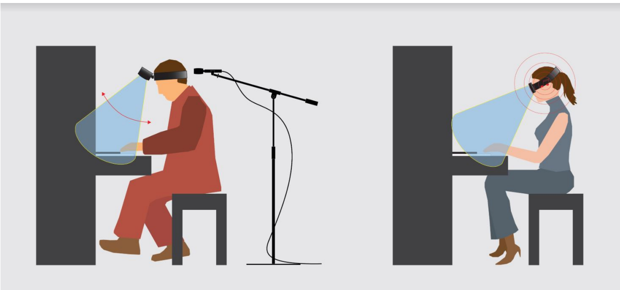

Figure 2. This is the overview of the system. The left side of the figure shows the piano teacher wearing the head-mounted camera with the recording microphone directly behind his head. On the right, we see the user (piano student) wearing the Magic Leap headset with spatialized audio. The teacher and the student have the same audiovisual perspective and orientation on the piano.

3.1 User Experience

Participants were seated at a Yamaha P45 digital piano with controlled lighting for optimal display of the virtual overlays, and ability to still see one's own piano and hands clearly. In the videos, the piano teacher is seated at an upright grand piano. Students were instructed to listen to the teacher's instructions, and then to place their hands on top of hers and play along with her when she instructed them to do so. The video sequences involved first an observation sequence in which the teacher showed the fingering movements. Then the teacher would instruct the students to get ready and place their fingers, and she would count down from three for when they should start playing along with her.

3.2 Visual Environment

The visual environment consists of 360 degree video footage that has been captured using the Garmin VIRB 360 camera with a head-mounted strap (made by Go Pro) to create a head-mounted camera worn by the piano teacher (see left panel in Figure 2). Multiple recordings were done of different short sequences in which the piano teacher explains how the fingers are numbered, how to place the hands on the keyboard, and how to play basic scales.

One concern with using overlays is that the visual environment quickly becomes cluttered. Even with the virtual piano and piano keys spatially aligned with the user's physical piano (using Vuforia ARCamera and TargetMarkers), early prototype testing indicated visual clutter. By creating an alpha channel in Adobe After Effects, we were able to reduce most of the visual noise from the piano by only having the virtual hands and the current key note pressed in the video with a blue shader to highlight it.

The first recordings were done with the top of the grand piano removed with the intention of allowing the student to see the hammer-head moving with the key depressed. When superimposed in the MR headset and environment, the visual environment appeared very busy and cluttered. Thus, inspired by the MirrorFugue project [21], we replaced the top of the grand piano and recorded the 360 video footage such that the top of the piano created a reflection of the piano teacher (see Figure 3). There are three reasons that it is valuable to show the virtual reflection of the piano teacher. The first is to cultivate greater social presence [41], so that the student can see the face of the piano teacher while she speaks and gives verbal instructions. Social presence is the feeling of being there together with another real person in an online, digital, or virtual remote collaboration, and it has shown to have significant effects on user satisfaction [41]. The second reason is modeled from of the design of virtual embodiment studies in VR. The setup for these studies involves a virtual mirror, in which the user can see him or herself reflected as the avatar they embody [22]. Thus, we wanted to include a similar 'virtual mirror' to create a greater sense of psychological identification with the piano teacher that might improve confidence and performance. The final reason for including the virtual mirror is that eventually we would like to visually annotate the virtual reflection of the teacher with real-time motion feedback from the student's movements to help the student notice the difference between their movements and those of the teacher. A related example comes from the i-Maestro musical training application for violin [42], which used a 3D Augmented Mirror showing synchronized video and motion data for bowing trajectories on the violin. This 3D Augmented Mirror enhanced students' understanding of the correct bowing techniques and body postures for playing the violin.

3.3 Audio Recording and Feedback

The teacher was seated at a Yamaha upright piano situated in a large room (volume: around 400 m^3) against a wall. A Sennheiser Ambeo VR Mic was positioned upright slightly above and behind the expert's head in order

Figure 3. This figure is a model representing what the user sees inside the headset in the Mixed Reality environment. The user sees the virtual hands of the teacher on top of the physical keyboard. There is a blue highlight on the currently pressed key. The user also sees the face and upper body of the pianist in the reflection on the upright grand piano.

to get a binauralised sound source. All sound files were encoded to B-format thanks to Ambeo A-B Format Converter then decoded to binaural format through FB360 Converter.

3.3.1 Audio Spatialization

Spatialized audio is used in ADEPT to differentiate the sounds of the teachers playing from the students playing, ultimately so that the student can learn to attenuate to their errors in pitch and timing while playing. Different development versions of the system have explored various ways to spatialize audio. The first design was to spatially misalign the teacher's point of view from the actual physical piano of the student. Here, we offset the sound virtually by using the Facebook 360 Audio Spatializer plugin in Reaper. The sound was made to spawn at locations directly above and to the left of the user's physical piano, above and to the right, matching the user's piano, and also slightly below. This spatial offset design was inspired by the visual offset design used by Xiao Xiao in MirrorFugue [21]. Preliminary data from pilot testing has shown that students find the audio spatialization offsets of slightly above and to the right and left to be the most comfortable and intuitive to follow than audio from below or matched to their piano location.

The second audio spatialization technique we are exploring is using binaural recordings rather than ambiosonics to capture the spatial perspective of the teacher. Based on the previous design results, we will be offsetting the spatialized spawning location of this binaural audio to be slightly above the user's piano. In future user studies, the binaural audio will also be accented by personalized 3D sound that account for the shape of the user's ears, which may enhance the effect.

4. FUTURE DIRECTIONS

Rather than using 360 degree video, we are currently exploring using volumetric video capture of the teacher's hands using the Structure Sensor and the Microsoft Kinect 2 depth-sensing cameras. We will also explore using Leap Motion hands to display the teacher's hands with a custom mesh of the teacher's actual hands from photogrammetry scans. A benefit for both of these designs is the possibility to easily resize and rescale the 3D hand models to more appropriately fit the student. Between these two designs, we will select the one that seems to have the most optimal display quality for the project. The virtual mirror will be present, not as now with the real reflection of the teacher in the piano, but instead as a darkened video screen that we will position in the Unity environment to spawn in the reflection of the piano. The goal for the design moving forward is to present the minimum amount of information necessary to facilitate user performance with the highest degree of quality. Currently, we are also developing a display with volumetric video capture and 3D motion capture data overlaid, which we could project as a 3D augmented Mirror like the i-Maestro project [42]. We hope that this can help the student to understand motion trajectories for the piano, combined with real-time motion feedback visual annotations.

Currently, the visual component of our design prototype has only explored adjusting the opacity of the video overlay of the teacher's visual perspective. In future user testing, we are also exploring having the teacher's body and hands appear next to the student, or above the student's hands on top of the keyboard, to explore if this makes it easier for students to follow the fingering patterns of the teacher. We will also test having the teacher's hands to be present from a first-person perspective aligning with the student's visual perspective in the observation phase, and then to appear above the student's hands during the ”play along” phase. And finally, another visual design prototype we plan to test involves having the teacher seated next to the student, and to allow the student to slide over on their piano bench to "sit into" the body and first-person perspective of the teacher.

We are also implementing a user interaction for the student to trigger hearing more or less from teacher's point of view. This means that the user will be able to select how much she sees or hears from the teachers point of view. That is, in future user testing, the system will be presented as though the teachers embodied sense reality is something that the student can choose to enter into. For instance, the visual overlay will be first set around 50 percent opacity, but the student can decrease or increase the opacity to see more or less from their own or the teacher's point of view. Similarly, the student can lean her head forward into a virtual sphere to hear more strongly from the teachers perspective.

Audio signal processing of the piano playing is being incorporated using pitch Pure Data, so that we can have an auditory analysis using the frequency to MIDI converter, pitch detection, and tempo tracking. We will use information to detect errors in playing the incorrect key or the correct key with the wrong finger. Ultimately, this analysis will allow the system to provide real-time feedback with visual motion annotation, pitch slider effects, and potentially also haptic feedback to the student.

The role of haptic feedback in learing musical movements has becoming increasingly vital to this project, and we are exploring the possibility of using a wrist actuator with five vibrating motors to represent movements for each of the fingers. Ideally, this haptic wearable device should be non-intrusive for the student's playing, which is why we are planning to test a wrist-worn device, as opposed to previous gloves which were not worn while playing the piano.

Lastly, we are collaborating with piano teachers at the Rhythmic Music Conservatory in Copenhagen to collect qualitative data on piano pedagogy towards a participatory design approach to make the technology more specific to enhance and compliment the students' learning experience. Additionally, we continue to conduct user testing with the system to address ongoing challenges during development. One of the goals in doing this is to better understand music and piano pedagogy and to explore the ways that the system can actually compare to and enhance traditional face-to-face piano instruction with a teacher. We hope that the system can eventually deliver piano instructions in a way that acknowledges the deep and complex history of piano pedagogy techniques, whereas the current focus in design and development has been much more focused on previous technology-enhanced systems for training to piano.

4.1 Motion Capture and Feedback

We are beginning to explore the major features of musical movement that distinguish expert pianists (masters) from more novice pianists, and also to explore the movement patterns characteristic of very novice adult students learning the piano for the first time. The purpose of this motion analysis is to create an evaluation metric for performances as an outcome to target as a result of the training.

Motion capture from the teacher will be captured using Leap Motion (mounted above to approximate head position) and photo-electric sensors with an optitrack system. Motion data from the student will be captured using the Leap Motion (mounted above at same coordinate ratio to where it had been mounted for the teacher) and Microsoft Kinect as both infrared and skeletal tracking systems. Motion data will capture finger and joint positions from both hands, as well as temporal sequencing and timing of movements.

Novice pianists focus on the extremities while playing, particularly having the correct fingering patterns, whereas expert pianists feel the music through their entire body [43]. Arms, wrists, and upper torso posture movements are usually introduced and trained at more advanced stages of learning, and are trained in isolation [43]. Learning how to move the body relies heavily on imitation. Observation and imitation of expert performance allows students to experience how music is felt in the body of another player [6]. Moreover, visual feedback about motor performance, accuracy, and adjustments can help improve reflection on one's own body and performance, and is often used in piano performance [36]. A previous training system used electromyography (EMG) to measure muscle activity in the thumb and successfully delivered biofeedback to help students achieve optimal muscle activation. While motion feedback does not deliver information directly about muscular activations, this still indicates a strong potential for motion feedback to imrpove motor performance while playing piano [44]. Thus, the primary goal of the motion feedback is to support successive adaptations in motor learning and performance. In order to monitor and evaluate the performance, the movements of the students will be captured with similar means.

4.2 Potential Challenges

Previous AR piano training applications have used Vuforia object tracking for matching the virtual and real pianos, but the distance at which the virtual piano appears from the user is still not quite correct. Thus, a potential challenge for the next stage of development is to ensure that the 3D model of the piano teacher's hands appears at the current depth and distance from the user. We found that we were able to control this apparent distance with the 360 degree video, but depth perception accuracy in the Magic Leap headset is not as precise as it could be. Specifically, we will want to make sure that the spatial configurations of the hands are easy to see and understand in a three-dimensional way, and that the user can perceive the distance and depth of the fingers accurately. A related challenge is that the teacher's hand blends in with the white keys on the keyboard, so we might need to put an outline around the hand and fingers, shade the hand with a color shader, or have the teacher wear colored gloves to increase the contrast.

Visual and aural latency could be potential challenges that could disrupt the user experience, specifically if the two sensory channels are out of sync. In future developments, the auditory mix between teacher's and student's sound could be a bit problematic if there is latency between visual and auditory information.

4.3 User Studies

Future user studies will be conducted on the binaural audio perspective-taking to see which settings optimize user experience. We will also explore adding user interactions to trigger the intensity of the perspective taking, and measure which decisions users choose to make at which time points to gain a sense of the usefulness of the user interactions, and which user actions should be programmed to trigger those visual and auditory effects.

4.4 Experimentation

We will compare two groups of students learning either with the ADEPT system or with the same content presented on a 2D video screen. The main target of experimentation is to explore how well students can actually follow the teacher. Thus, we will be measuring interpersonal synchrony comparing motion capture data recorded from the teacher with that of the student. Our hypothesis is that interpersonal synchrony during training will predict better performance out of the system. Afterwards, we will ask the students to play the same sequences from memory and measure performance accuracy using motion capture analysis with the Musical Gestures toolbox in Matlab and video analysis using Elan Software for Transcription. We will also have an expert panel of professional piano instructors who will rate the performance of the users who had been trained with the system, as compared to just watching a video.

5. DISCUSSION AND CONCLUSIONS

In this paper we presented ADEPT, a system for facilitating learning to playing the piano. The ADEPT system aims at teaching musical movements on the piano in an embodied way so that the student learns to move like a professional pianists. We use audiovisual perspective taking with a piano teacher to help students orient to the correct hand, finger, wrist and upper torso positions for sound-producing movements on the piano. By introducing the notion of Augmented Embodiment, the student can see and hear a blend of his or her own body and that of the teacher from a first-person perspective. Increased user interactions to control the intensity of the audiovisual perspective taking are currently being implemented and tested. Initial prototype testing indicates that positioning the sounds of the teacher' piano playing slightly above the student optimizes comfort, and further testing will determine if this also optimizes performance. In conclusion, the ADEPT system offers a new design technique using modern Augmented Reality technology with audio-visual perspective taking and feedback to help teach piano to novice students.

Acknowledgements and Author contributions

We would like to thank Christie Anne Jeannine Laurent for her patience and excellent piano playing and instruction to create the audiovisual instructional content. We would also like to thank Daniel Ludwig for creating the figures that demonstrate the overall design schema for the project and the anonymous reviewers for their helpful comments in improving the paper.

Author Gerry was responsible for designing and developing the application, as well as collecting initial prototype data. Author Dahl designed data collection and analysis tools for motion capture and analysis. Author Serafin designed data collection and analysis tools for audio spatialization. All authors participated in writing and revising the manuscript.

This work is supported by NordForsk's Nordic University Hub Nordic Sound and Music Computing Network NordicSMC, project number 86892.

6. REFERENCES

[1] M. Leman, *Embodied music cognition and mediation technology.* Mit Press, 2008.

[2] S. Shirmohammadi, A. Khanafar, and G. Comeau, "MIDIATOR: A tool for analyzing students piano performance," *Revue de recherche en ducation musicale*, vol. 24, pp. 35–48, 2006.

[3] S.-H. Lee, "Using the personal computer to analyze piano performance," *Psychomusicology: A Journal of Research in Music Cognition*, vol. 8, no. 2, p. 143, 1989.

[4] R. Lindgren, "Generating a learning stance through perspective-taking in a virtual environment," *Computers in Human Behavior*, vol. 28, no. 4, pp. 1130–1139, 2012-07. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0747563212000234

[5] X. Xiao, "Reflecting music through movement: a body-syntonic approach to playing [with] the piano." 2016. [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/107576/974646581-MIT.pdf?sequence=1

[6] X. Xiao and H. Ishii, "Inspect, embody, invent: A design framework for music learning and beyond," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, 2016, pp. 5397–5408. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2858036.2858577

[7] G. Percival, Y. Wang, and G. Tzanetakis, "Effective use of multimedia for computer-assisted musical instrument tutoring," in *Proceedings of the international workshop on Educational multimedia and multimedia education - Emme '07*. ACM Press, 2007, p. 67. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1290144.1290156

[8] J. Chow, H. Feng, R. Amor, and B. C. Wunsche, "Music education using augmented reality with a head mounted display," *User Interfaces*, vol. 139, p. 8, 2013.

[9] M. Weing, A. Rhlig, K. Rogers, J. Gugenheimer, F. Schaub, B. Knings, E. Rukzio, and M. Weber, "P.i.a.n.o.: enhancing instrument learning via interactive projected augmentation," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication - UbiComp '13 Adjunct*. ACM Press, 2013, pp. 75–78. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2494091.2494113

[10] D. Hackl and C. Anthes, "HoloKeys - an augmented reality application for learning the piano," in *Proceedings of the 10th Forum Media Technology and 3rd All Around Audio Symposium*, 2017-11, pp. 140–144.

[11] S. Ferguson, "Learning musical instrument skills through interactive sonification," in *Proceedings of the 2006 conference on New interfaces for musical expression*. IRCAMCentre Pompidou, 2006, pp. 384–389.

[12] S. W. Smoliar, J. A. Waterworth, and P. R. Kellock, "pianoforte: A system for piano education beyond notation literacy," in *ACM Multimedia*, vol. 95, 1995.

[13] X. Xiao, B. Tome, and H. Ishii, "Andante: Walking figures on the piano keyboard to visualize musical motion," in *New Interfaces for Musical Expression (NIME)*, 2014, pp. 629–632.

[14] R. B. Dannenberg, M. Sanchez, A. Joseph, P. Capell, R. Joseph, and R. Saul, "A computer-based multimedia tutor for beginning piano students," *Journal of New Music Research*, vol. 19, no. 2-3, pp. 155–173, 1990.

[15] D. Kohlsdorf and T. Starner, "Mobile music touch: The effect of primary tasks on passively learning piano sequences," in *International Symposium on Wearable Computers (ISWC) 2010*. IEEE, 2010, pp. 1–8.

[16] K. Huang, E. Y.-L. Do, and T. Starner, "Pianotouch: A wearable haptic piano instruction system for passive learning of piano skills," in *2008 12th IEEE International Symposium on Wearable Computers*. IEEE, 2008, pp. 41–44.

[17] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: a class of displays on the reality-virtuality continuum," H. Das, Ed., 1995-12-21, pp. 282–292. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=981543

[18] D. L. Dieker, M. Hynes, C. Stapleton, and D. C. Hughes, "Virtual classrooms: STAR simulator," *New Learning Technology SALT*, vol. 4, pp. 1–22, 2007.

[19] "Personal contextual awareness through visual focus," vol. 16.

[20] E. Strasnick, A. Chambers, L. Jiang, and T. Xiaonan, "Pianolens: An augmented reality interface for piano instruction," 2017. [Online]. Available: https://cs.stanford.edu/people/estrasni/otherprojects/pianolens.html

[21] X. Xiao and H. Ishii, "Mirrorfugue: communicating hand gesture in remote piano collaboration," in *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. ACM, 2011, pp. 13–20.

[22] B. Spanlang, J.-M. Normand, D. Borland, K. Kilteni, E. Giannopoulos, A. Pomés, M. González-Franco, D. Perez-Marcos, J. Arroyo-Palacios, X. N. Muncunill *et al.*, "How to build an embodiment lab: achieving body representation illusions in virtual reality," *Frontiers in Robotics and AI*, vol. 1, p. 9, 2014.

[23] K. Kilteni, I. Bergstrom, and M. Slater, "Drumming in immersive virtual reality: The body shapes the way we play," p. 9, 2013.

[24] D. Banakou, P. D. Hanumanthu, and M. Slater, "Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias," *Frontiers in Human Neuroscience*, vol. 10, 2016-11-29. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00601/full

[25] L. Maister, M. Slater, M. V. Sanchez-Vives, and M. Tsakiris, "Changing bodies changes minds: owning another body affects social cognition," *Trends in cognitive sciences*, vol. 19, no. 1, pp. 6–12, 2015.

[26] M. Slater and M. V. Sanchez-Vives, "Transcending the self in immersive virtual reality," *Computer*, vol. 47, no. 7, pp. 24–30, 2014.

[27] M. J. Traxler and M. A. Gernsbacher, "Improving written communication through perspective-taking," *Language and Cognitive Processes*, vol. 8, no. 3, pp. 311–334, 1993.

[28] J. Guegan, S. Buisine, F. Mantelet, N. Maranzana, and F. Segonds, "Avatar-mediated creativity: When embodying inventors makes engineers more creative," *Computers in Human Behavior*, vol. 61, pp. 165–175, 2016.

[29] L. J. Gerry, "Paint with me: Stimulating creativity and empathy while painting with a painter in virtual reality," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 4, pp. 1418–1426, 2017.

[30] D. Banakou, S. Kishore, and M. Slater, "Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias," *Frontiers in Psychology*, vol. 9, p. 917, 2018.

[31] M. Slater, V. Linakis, M. Usoh, and R. Kooper, "Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess," in *ACM virtual reality software and technology (VRST)*, vol. 163. ACM Press, 1996-07, p. 72.

[32] R. Watanabe, T. Higuchi, Y. Kikuchi, and M. Taira, "Visuomotor effects of body part movements presented in the first-person perspective on imitative behavior: Movement and first-person perspective," *Human Brain Mapping*, vol. 38, no. 12, pp. 6218–6229, 2017-12. [Online]. Available: http://doi.wiley.com/10.1002/hbm.23823

[33] B. M. Brooks and F. D. Rose, "The use of virtual reality in memory rehabilitation: Current ndings and future directions," *NeuroRehabilitation*, vol. 18, no. 2, pp. 147–157.

[34] L. Fadiga, L. Fogassi, G. Pavesi, and G. Rizzolatti, "Motor facilitation during action observation: a magnetic stimulation study," *Journal of Neurophysiology*, vol. 73, no. 6, pp. 2608–2611, 1995-06. [Online]. Available: http://www.physiology.org/doi/10.1152/jn.1995.73.6.2608

[35] S. D. Muthukumaraswamy, B. W. Johnson, and N. A. McNair, "Mu rhythm modulation during observation of an object-directed grasp," *Cognitive Brain Research*, vol. 19, no. 2, pp. 195–201, 2004-04. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0926641003002994

[36] J. Mora, W.-s. Lee, G. Comeau, S. Shirmohammadi, and A. Saddik, "Assisted piano pedagogy through 3d visualization of piano playing," in *2006 IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE 2006)*. IEEE, 2006, pp. 157–160. [Online]. Available: http://ieeexplore.ieee.org/document/4062532/

[37] K. Alaerts, E. Heremans, S. P. Swinnen, and N. Wenderoth, "How are observed actions mapped to the observers motor system? influence of posture and perspective," *Neuropsychologia*, vol. 47, no. 2, pp. 415–422, 2009-01. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0028393208003801

[38] S. V. Adamovich, G. G. Fluet, E. Tunik, and A. S. Merians, "Sensorimotor training in virtual reality: A review," *Neurorehabilitation*, vol. 25, no. 1, pp. 29–44, 2009. [Online]. Available: http://www.medra.org/servlet/aliasResolver?alias=iospress&genre=article&issn=1053-8135&volume=25&issue=1&spage=29&doi=10.3233/NRE-2009-0497

[39] S. P. Swinnen, "Information feedback for motor skill learning: A review," *Advances in motor learning and control*, pp. 37–66, 1996.

[40] H. Sveistrup, "Motor rehabilitation using virtual reality," *Journal of NeuroEngineering and Rehabilitation*, p. 8, 2004.

[41] S. T. Bulu, "Place presence, social presence, co-presence, and satisfaction in virtual worlds," *Computers & Education*, vol. 58, no. 1, pp. 154–161, 2012.

[42] K. Ng and P. Nesi, "i-maestro: Technology-enhanced learning and teaching for music." in *NIME*, 2008, pp. 225–228.

[43] S. Furuya and H. Kinoshita, "Organization of the upper limb movement for piano key-depression differs between expert pianists and novice players," *Experimental Brain Research*, vol. 185, no. 4, pp. 581–593, 2008-03. [Online]. Available: http://link.springer.com/10.1007/s00221-007-1184-9

[44] R. Montes, M. Bedmar, and M. S. Martin, "Emg biofeedback of the abductor pollicis brevis in piano performance," *Biofeedback and self-regulation*, vol. 18, no. 2, pp. 67–77, 1993.