Banquiero-Garófalo, MM.; Valdeolivas, G.; Trincado, S.; García, N.; Juan, M. (2023). Passthrough Mixed Reality With Oculus Quest 2: A Case Study on Learning Piano. IEEE Multimedia. 30(2):60-69. https://doi.org/10.1109/MMUL.2022.3232892

Passthrough Mixed Reality with Oculus Quest 2: A case study on learning piano

Mariano Banquiero, M.-Carmen Juan, Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, 46022 Valencia, Spain

Gracia Valdeolivas, Departamento de Pedagogía, Universitat Jaume I, 1206 Castellón, Spain

Sergio Trincado and Natasha García, Music School U.M. Santa Cecilia of Benicàssim, 12560 Benicàssim, Spain

Abstract—Mixed reality (MR) in standalone headsets has many advantages over other types of devices. With the recent appearance of the Passthrough of Oculus Quest 2, new possibilities open up. This work details the features of the current Passthrough and how its potential was harnessed and its drawbacks minimized for developing a satisfying MR experience. It has been applied to learning to play the piano as a use case. A total of 33 piano students participated in a study to compare participants' interpretation outcomes and subjective experience when using a MR application for learning piano with two visualization modes (border lines on all the keys (Wireframe) vs. solid color hiding the real keys (Solid)). The two visualization modes provided a satisfying experience. Even though there were no significant differences in the analyzed variables, the students preferred the Solid mode, indicating that short-distance Passthrough limitations should be minimized in application development.

Extended Reality (XR) is a term used to refer to technology-mediated experiences that combine real and virtual environments and objects, where the "X" is a placeholder for (A)R (Augmented Reality), (M)R (Mixed Reality), or (V)R (Virtual Reality). XR is experiencing its peak, largely thanks to the appearance of hardware (e.g., Oculus Quest 2, HoloLens 2) and software (e.g., Vuforia, ARCore, ARKit, SparkAR) with more and more features. VR headsets have been used for AR experiences by adding cameras and placing them on the front of the VR headsets. The headsets (non-standalone) have to be connected to a computer. Therefore, adding a new cable (e.g., for the USB camera) that is connected to the computer is not a big problem. With the appearance of standalone headsets (e.g., Oculus Quest), greater freedom of movement has been achieved which allows headsets of this type to be used to develop applications that previously had not been possible. Moreover, Oculus has recently released an experimental API called Passthrough that allows the users to visualize their real-world environment in 3D thanks to the device's front-facing cameras. This new functionality opens the possibility of having MR in a standalone VR headset such as the Oculus/Meta Quest 2 without the need to add any cameras. A limitation of this new functionality is that the image of the real world is of low quality, has some instability, and is currently displayed in grayscale. A second limitation is that the running application does not have direct access to the information of the real-world image. It behaves as if it were a fixed background that is later replaced by the real image captured by the sensors in a totally transparent way to the app.

Even though the potential of this new functionality is great, these limitations affect the specific use that can be given to it. In this work, different techniques are explored to overcome these limitations and to develop a MR application for learning to play the piano. The developed application uses Passthrough to show the user the real image of the piano on which a previously loaded piece of music is to be played. The information of the different notes (time and duration) are shown as virtual rectangles that fall on the keys of the real piano, and the application renders these virtual objects on the image of the 3D world obtained from Passthrough.

The objective of this work is to determine if programming different techniques using the Passthrough API (even with its current limitations) can create a satisfactory MR experience for users, in our case, applied to learning to play the piano. The Passthrough image can be unstable (moving and curving) depending on the distance from the observer. In this work, we also intended to determine if this limitation affects the user when playing a piano key. To do thus, two display modes were developed (border lines on all the keys (Wireframe) vs. solid color hiding the real keys (Solid)). In the Wireframe mode, only the border lines of the keys are displayed and the above-mentioned issues could hinder the selection of the key. However, in the Solid mode, the keys are completely covered and the user does not have these problems.

RELATED WORK

Different XR systems have been presented to help in learning music and specifically the piano. These systems use different technologies. Huang et al. [1] presented Piano AR, which uses natural feature tracking to identify the keyboard outline and the keyboard area by the structure of the black and white keys. They use a webcam to capture the image. Their work focused on determining the accuracy of the system. Chow et al. [2] presented an AR system, which uses a headset with cameras (Trivisio ARvision-3D) and marker-based tracking. Their system presents a feedback mechanism for note playing accuracy and summary feedback after playing a piece. Rogers et al. [3] developed P.I.A.N.O., a system in which notes are physically projected onto the real piano. To do this, they use a projector that is placed above the piano using a tripod. Desnoyers-Stewart et al. [4] created a system that combines VR headsets (HTC Vive) with a hand tracking device (Leap Motion) and an Arduino-based MIDI keyboard. When users play the real piano, their hands are detected by Leap Motion. The virtual piano and a representation of their hands are shown in the headset. Rigby et al. [5] and Molloy et al. [6] both used VR headsets (HTC Vive) with a mounted stereo camera (Zed Mini). Some systems have been proposed since the appearance of HoloLens (Birhanu and Rank [7]; Molero et al. [8]). Molero et al. [8] used HoloLens and Vuforia to track the piano. They used the piano itself as an image target which is tracked by Vuforia.

The objective of many of the XR systems presented is to achieve learning. Rogers et al. [3] compared P.I.A.N.O, Synthesia[1], and Finale. Their results showed that P.I.A.N.O. provided better initial learning performance, faster progress over one week of practice, induced less cognitive load, provided a better user experience, and led to better perceived quality compared to sheet music notation (Finale) and non-projected piano roll notation (Synthesia). Birhanu and Rank [7] explored piano pedagogy using HoloLens. They showed that the augmented contextualization of MR can be an effective tool for aiding students in the process of learning to play piano. Rigby et al. [5] developed an AR piano tutoring tool for teaching sheet music reading, which overlays played notes on a virtual music sheet. They showed that their tool significantly improves recall of notes by name and correct playing of notes. Molloy et al. [6] presented an AR piano teaching tool that focuses on feedback for rhythm and note accuracy. They report that users had greater note playing accuracy than with Synthesia.

THE PASSTHOUGH MIXED REALITY APPLICATION

API Passthrough

The API Passthrough for Oculus/Meta Quest 2 provides the user with a perception of the real world in 3D and in real time thanks to the capture of the frontal sensors of the device. It was originally designed for its use by the Guardian application, which is the application that runs at the beginning of the device start-up and that allows a safe area of use to be defined, e.g., to avoid colliding with obstacles in the game area. Since the interface was not specifically designed to support MR experiences, it has some limitations and artifacts that limit its use in different applications.

Passthrough works as follows: the real-world image is rendered by a special service on a separate layer which is then mixed internally by the device software. The application has no direct access to the image information, the RGB color channels, the depth (Z-buffer), or to any other information captured by the sensors. In other words, the application has very limited control over how to interact with Passthrough, and, in particular, no computer vision techniques such as object detection can be applied. The current version also does not allow operations that use

[1] https://synthesiagame.com/

depth such as Z-buffering (or similar) using Passthrough, so that any virtual element that is drawn on the scene overwrites the real-world image. The API supports some fixed types of compositing with the Passthrough layer, such as automatic edge detection and color keying (which allows drawing the Passthrough in a very limited number of colors) and is always transparent to the application.

Design

The application was designed taking advantage of the potential of Passthrough and minimizing the problems mentioned above. Our objective was to design a functional MR application with different representations when using Oculus Passthrough for augmenting the piano keyboard and to explore its potential. Our first step was to analyze the main functional characteristics. The application must include interaction using the hands. The interaction should be as natural and intuitive as possible. The application must include visual aids to facilitate learning to play the piano while playing the real piano. The design, development, and validation of the application had the involvement of piano teachers. In a co-creation process, the piano teachers reviewed the different prototypes until the final prototype was achieved. This is the one used in the study. Some of the guidelines derived from the design are the following (Figure 1a):

- ➔ *Guide Background* — It is a dark background that has lines on the octaves and fifths and differences in intensity between two successive keys. This background ends just above the beginning of the real piano keyboard and serves as a guide to guess the position of each falling note.
- ➔ *Colors* — Notes with accidentals (sharp or flat) to be played on a black piano key appear in a darker tone than normal notes. Left-hand notes appear in one color and right-hand notes in another.
- ➔ *C keys* — The keys that correspond to the notes C (DO) are highlighted in color overwritten on the real piano to facilitate the location of the octaves.
- ➔ *Note names* — They appear above the end of the real piano keyboard.
- ➔ *Staff* — The application can optionally include the sliding staff at the top of the interface.
- ➔ *Division of the measures* — The application can optionally show horizontal lines on the Guide Background indicating measures.

Two ways to augment the piano keyboard were defined:

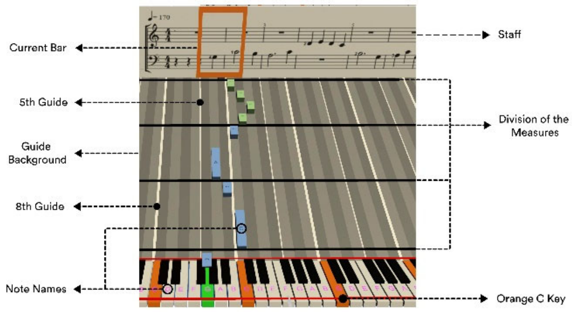
- ➔ *Wireframe* — The application draws border lines on all of the keys, the white keys in white and the black keys in black. The user sees a combination of their real piano with the virtual edges.
- ➔ *Solid* — The app draws the keys in solid color hiding part of the real piano. The user sees the C keys in a different color and the note name for each key.

Overview of the MR application for learning to play the piano

Thanks to Passthrough, the user can see the real piano and the whole environment at the same time as the virtual notes falling on the real keys. In addition, the key that must be pressed at each moment is highlighted with a color on the real piano. The user presses the keys of the real piano playing the melody. Figure 1b shows a user playing the piano with the application, and Figure 1c shows what she is actually seeing on the headset.

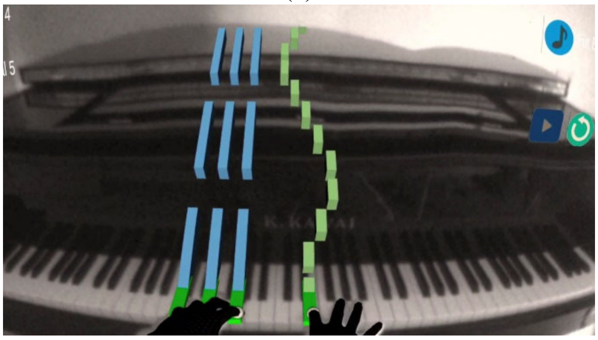
The application begins with a phase of calibration with the real world, which consists of defining the exact position and orientation of the real piano in the virtual space of the device. In this way, the augmented information is perfectly adjusted to the real piano. The user selects the piano type from a list of standard keyboards, indicating the number of keys. In our application, 49, 61, 76, or 88 keys can be selected. Then, the position and orientation in space are indicated using the controllers or the hands. This is first done approximately and then with a fine adjustment, where the position can be slightly modified using virtual buttons.

Once the calibration phase is complete, the application shows a user interface located in front of the piano in such a way that it is accessible without hiding the keys. The first thing the user has to do is to follow the five tutorials in order to become familiar with how the application works. After this familiarization, the user can select a piece of music from the interface. When a piece is selected, a series of rectangles that represent the notes fall on the corresponding piano keys, respecting the time and duration. These rectangles, which we will refer to as falling notes, are the central element of the application since they constitute the main augmented information that guides the user in the interpretation of the piece. Each falling note collides on a physical piano key at the exact moment in which the user has to press that key and keeps colliding for the time that corresponds to the duration of that note.



(a)

(b)



(c)

FIGURE 1. (a) Image showing all of the visual aids that can be selected in our application. (b) Photograph of a user playing the piano with the application and what she is seeing superimposed (blue and green) and the C keys (orange). (c) User’s view using the headset. Passthrough image with augmented information (blue and green).

Architecture and technical features

The application was developed using Unity 2020.3.14 and the first version of the experimental API Passthrough. The application consists of the following modules. The architecture of the application is shown in Figure 2.

The Sync Module

The objective of this module is for the headset to know the exact location of the real piano in the virtual space of the application. Calibration is the most sensitive phase of the whole process because if the piano is not perfectly synchronized in the virtual space of the device, the visual information generated from the falling notes will not intersect in the precise place of the piano keyboard. As mentioned, the API Passthrough does not currently allow access to the information necessary to apply computer vision techniques to the environment and automatically detect the piano or the keys. For this reason, three methods were implemented to solve this problem. The first two methods use the controllers, and the third method uses only the user's hands. The result of the calibration of the three methods is practically the same. Regardless of the method used, after this initial calibration, via menu options manual fine-tuning can be done in millimeter intervals. Depending on the user’s experience with headsets and controllers and the features of the piano, one method may be more advisable than another. The third method was used in the study included in this work.

The first method

The user selects and places a spherical object, called anchor, on the left side of the piano using the controller. Once it is fixed, the process is repeated with the right anchor while the application interactively shows a 3D model of the piano keyboard. This method has the advantage that it is very simple to use and quite fast. The drawback is that it can sometimes be visually confusing to determine if the anchor position is correct.

The second method

The user uses both controllers, rests the left controller on the left side of the piano, and presses the main button to fix it. Then this process is repeated with the right controller. This method has the advantage of being simple, fast, and quite accurate, especially at height. The drawback is that when the controller goes out of the angle of Oculus viewing range, for example, if the piano is long (a grand piano), the device stops tracking the controller and calibration is lost.

The third method

The users use their hands and use the index finger of the left hand to physically press the first key of the piano while doing the pinching gesture with the right hand for a few seconds to confirm the left position. The process is repeated for the right side. This time, the index finger of the right hand is used and pinching is done with the left hand. The advantage of this method is that it is extremely precise thanks to the Oculus hand tracking. In addition, it allows the finger to rest exactly on the position of the key.

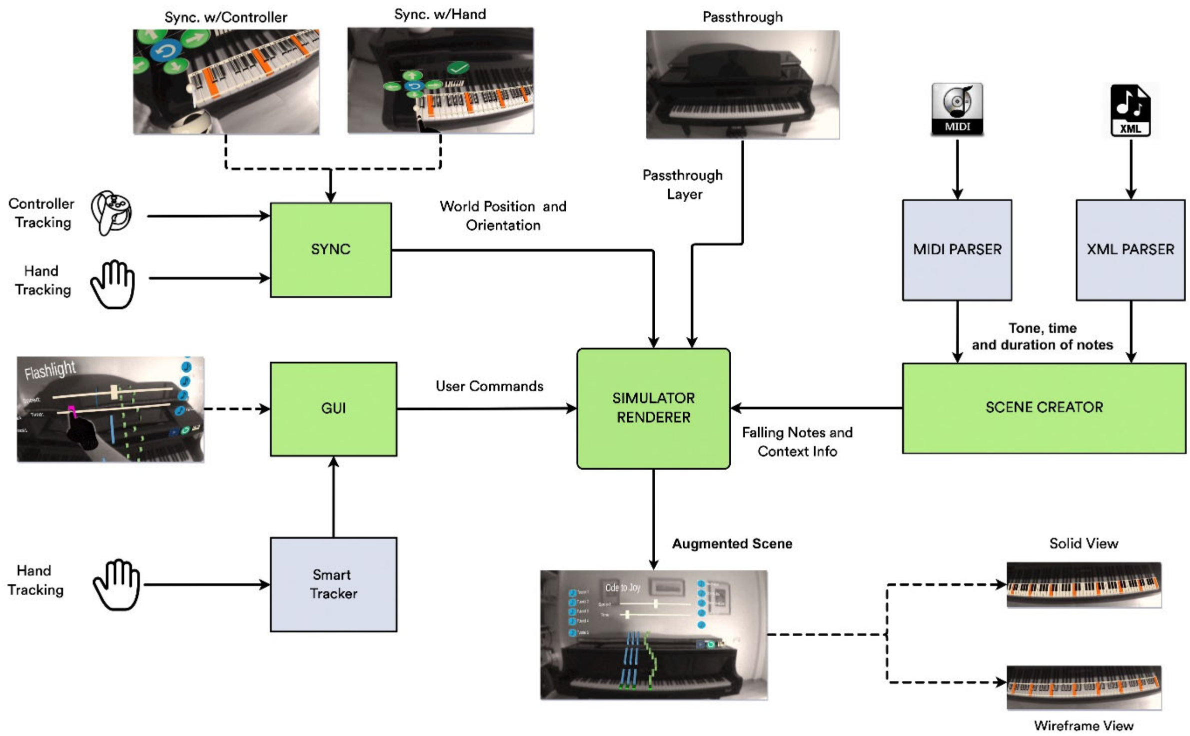


FIGURE 2. Architecture of the application.

The User Interface (GUI)

Oculus guidelines and recommendations[2] for the user interface design are followed. The users only use their hands to interact with the application. The gesture of pressing virtual buttons is done using the index finger. The Oculus SDK API provides the position and orientation of each bone of the fingers of both hands. However, a direct raw use of this information could produce undesirable effects. For example, in the detection of an event such as a button pressed, a simple collision is used between the bounding box that represents the button and the tracked position of the index finger. The natural vibration of the hand and the sampling errors of the device may cause the fingertip to press the button several times. Our implementation of the GUI controls takes into account all of these factors in order to solve this problem. Therefore, the application behaves naturally. The module called Smart-tracker processes the raw information obtained from the Oculus API to adjust it to the current situation of the user interface.

The user interfaces and the behavior of the different elements were implemented from scratch, such as the detection of collisions between the finger and the virtual elements. This implementation was done from scratch for performance reasons of the Unity native collision system.

The MIDI/Music XML Parser

The musical pieces are imported from files with MIDI[3] format or MusicXML[4] format. The information of the notes is extracted from them: frequency, start time, and duration. The application generates the geometry of the

[2] https://developer.oculus.com/resources/hands-design-ui
[3] http://www.music.mcgill.ca/~ich/classes/mumt306/StandardMIDIfileformat.html
[4] https://www.w3.org/2021/06/musicxml40/musicxml-reference

falling notes. The start time determines the position on the Y axis, the frequency (pitch) determines the position on the X axis, and the duration determines its length.

The simulator and renderer

This module is in charge of the animation of the falling notes that must respect the time and the speed of reproduction. The vital aspect of this module is that the user can easily identify in advance on exactly which piano key each one of the falling notes is going to fall in order to be able to press the right key on the real piano without errors and at the right time.

DESCRIPTION OF THE STUDY

Participants

A total of 33 piano students from the Music School U.M. Santa Cecilia of Benicàssim, Spain participated in the study; 18 women and 15 men were distributed in two groups, balancing gender and age range. The two groups were Group Wireframe and Group Solid. Group Wireframe consisted of 16 participants, ranging in age from 7 to 69 years old. Group Solid consisted of 17 participants, ranging in age from 8 to 68 years old. The participants or their parents were informed about the study and its objectives. The participants gave written informed consent prior to the study. The study was conducted in accordance with the declaration of Helsinki and was approved by the Ethics Committee of the Universitat Politècnica de València. Spain.

Measures

Teachers' scores when the student plays the piano. The teachers evaluate how the student plays two pieces (TS1 and TS2 variables).

Questionnaires. There were two questionnaires. Questionnaire 1 is divided into four blocks containing questions from four standard questionnaires: the Simulator Sickness Questionnaire (SSQ) [9], six questions to measure presence [10], four questions based on the Technology Acceptance Model (TAM) [11], and the User Experience Questionnaire (UEQ) [12]. We also included an additional question for self-assessment (SELFA variable) about how the students thought they had played the piano on a scale from 1 to 7, where 1 indicated very poorly and 7 indicated excellent.

The SSQ consists of sixteen symptoms with a 4-point rating scale from 0 (none) to 3 (severe). The symptoms are divided into three subscales: nausea, oculomotor, and disorientation. Each subscale considers seven symptoms. All three subscales include overlapping symptoms from the other subscales. The total score (TotalSSQ variable) and the scores for the subscales (Nausea, Oculomotor, and Disorientation variables) were calculated following the procedure indicated by Kennedy et al. [9].

The presence score was obtained as the mean of the ratings of the presence questions (PRES variable). Two questions were asked in negative, and they were reversed to calculate the mean.

Three of the TAM questions (PU1-PU3 variables) measure whether our application is useful and meets its goal effectively (Perceived Usefulness). One of the TAM questions (PE variable) measures whether our application is easy to understand and use (Perceived Ease of Use).

The UEQ contains twenty-six items on a 7-point Likert scale that are grouped into six scales or variables: Attractiveness, Perspicuity, Dependability, Efficiency, Novelty, and Stimulation.

Questionnaire 2 consists of questions designed for this study to determine the preference of the participants regarding the visualization mode and open comments. These six questions were: PR1- Which one did you like the most (Wireframe or Solid)?; PR2- Which one do you think is better for learning music (Wireframe or Solid)?; PR3- Which one was easier for you to handle (Wireframe or Solid)?; PR4- Would you recommend any to your classmates (Wireframe, Solid, the two equally, none)?; PR5- Would you like one of them to be used by your teacher in class (Wireframe, Solid, the two equally, none)?

Protocol

The participants were counterbalanced and randomly assigned to one of two conditions:

- ➔ Group Wireframe: The students who played the piano visualizing the border lines of the keys to play two pieces first and then used the visualization mode that shows all of the keys to play two other pieces second (at the teacher's discretion they may be the same or different).

- ➔ Group Solid: The students who used the app with the visualization mode showing all of the keys to play two pieces first and then played the piano displaying the border lines of the keys to play two other pieces second (at the teacher's discretion they may be the same or different).

The protocol was the following:

- ➔ The students played two pieces wearing the headset and using the first visualization mode assigned to their group.
- ➔ The participants filled out Questionnaire 1 online.
- ➔ The students played two pieces wearing the headset and using the second visualization mode assigned to their group.
- ➔ The participants filled out Questionnaire 2 online.

The sessions were held at the Music School for one week from Monday to Friday, from 15:30 to 21:00, and each student required about half an hour to complete the task. The standard Quest 2 Strap was changed for the Quest 2 Elite Strap with Battery. The Elite Strap is more ergonomic than the standard strap and has a wheel for easy adjusting. The battery adds weight to the rear, but that weight balances the weight of the front by balancing it, and, it also relieves pressure on the nose, which contributes to greater user comfort. The battery allowed the device to be used during the whole session.

RESULTS

The Shapiro-Wilk test was used to check the normal distribution of the variables. The tests indicated that the sample did not fit a normal distribution. For this reason, we used non-parametric tests. The results were considered to be statistically significant if $p < 0.05$. We used the statistical open source toolkit R[5] to perform the statistical analysis of the data.

User experience

We used the UEQ Data Analysis Tool6 for the analysis of the responses to the UEQ. The results indicate that our application in its two visualization modes is excellent compared to the Benchmark included in the UEQ Data Analysis Tool, and it is in the range of the 10% best results on its six scales. The Mann Whitney U tests were applied to check if the visualization mode affected any of the six variables. The results are shown in the UEQ rows of Table 1. No statistically significant differences were found for any of the six variables.

The Mann-Whitney U tests were applied to check whether or not there were differences for the Perceived Usefulness (PU1-PU3) and Perceived Ease of Use (PE) variables between the two groups of participants. The results are shown in the TAM rows of Table 1. No statistically significant differences were found for any of the four variables.

Presence

The Mann-Whitney U test was applied to check whether or not there were differences for the Presence variable between the two groups of participants. The results are shown in the Presence row of Table 1. No statistically significant differences were found for this variable.

Adverse effects of wearing glasses

The Mann-Whitney U tests were applied to check whether or not there were differences for the SSQ variables between the two groups of participants. The results are shown in the SSQ rows of Table 1. No statistically significant differences were found for this variable.

The mean of TotalSSQ is 14.49 for the use of the Wireframe mode and 7.26 for the use of the Solid mode. The thresholds for classifying the symptoms induced are negligible (<5), minimal (5 – 10), significant (10 – 15), concerning (15 – 20), and bad (>20) [13]. Therefore, the symptoms induced by the Wireframe mode are significant, and the symptoms induced by the Solid mode are minimal.

Gender and age

The Mann Whitney U tests were applied to check if gender affected the variables used (Nausea, Oculomotor, Disorientation, TotalSSQ, PRES, PU1-PU3, PE, Attractiveness, Perspicuity, Dependability, Efficiency, Novelty, Stimulation, SELFA). No statistically significant differences were found for any of these variables when analyzing the two groups separately, except for the PE variable (Perceived Usefulness) when using the Solid mode. In that mode, the women found it easier than men to play pieces with the piano. This analysis is included in the G2 (Solid) row of Table 1. Since there was only a single significant difference of the sixteen variables analyzed when comparing the opinion of men and women in each visualization mode, we conclude that the results were independent of the gender of the participants when using the MR application with the two visualization modes.

The Kruskal Wallis test was applied to determine if age influences the same variables as those used for the gender analysis. From these analyses, it can be concluded that the results were independent of the age of the participants when using the two visualization modes.

[5] http://www.r-project.org

[6] https://www.ueq-online.org

TABLE 1. Mann-Whitney U test for the variables used in the study.

Questionnaire	Variables	G1	G2	U	Z	p	r
UEQ	Attractiveness	2.7;0.7	2.5;1.0	156.0	0.736	0.473	0.128
	Perspicuity	2.8;1.3	2.3;1.3	161.5	0.944	0.355	0.164
	Efficiency	3;0.4	2.7;1.0	153.5	0.680	0.509	0.118
	Dependability	2.5;1.3	2.3;1.3	139.5	0.129	0.912	0.022
	Stimulation	2.5;0.88	2.75;1	147.0	0.407	0.698	0.071
	Novelty	2.5;0.75	3;0.75	120.0	-0.609	0.555	0.106
TAM	PU1	7;3.25	6;1	138.0	0.077	0.954	0.013
	PU2	7;1	7;1	135.0	-0.041	0.984	0.007
	PU3	6;1.25	7;1	104.5	-1.225	0.228	0.213
	PE	6;3	6;1	136.0	0.000	1.000	0.000
Presence	PRES	5.6;1.8	5.7;1.2	133.0	-0.108	0.928	0.019
SSQ	Nausea	4.8;9.5	0;0	175	1.646	0.104	0.287
	Oculomotor	11.4;30.3	0;7.6	174	1.447	0.153	0.252
	Disorientation	6.9;27.8	0;13.9	156	0.791	0.440	0.138
	TotalSSQ	11.2;26.2	0;11.2	174.5	1.460	0.150	0.254
Self-assessment	SELFA	5;0.5	5;1	147.5	0.438	0.675	0.076
Group	**Variables**	**Women**	**Men**	***U***	***Z***	***p***	***r***
G2 (Solid)	PE	7;1	6;2	59	2.376	**0.020**	0.576
Subgroup		**Known piece**	**Unknown piece**	***U***	***Z***	***p***	***r***
G1 (Wireframe)		8;1	7;1	50	1.978	0.054	0.495
G2 (Solid)		9;2	8;2	86	1.717	0.092	0.366

Note: The values in columns G1 (Wireframe Group), G2 (Solid Group), Women, Men, Known piece, and Unknown piece depict the median and the interquartile range, respectively. The numbers in bold indicate significant differences.

Outcomes

The piano teachers selected the first and the second piece to be played by the student. They assessed how the student played each of the two pieces on a scale from 0 to 10. When possible, a known piece and one that the student did not know were selected. In some cases, this was not possible because the students already knew all of the pieces included in the application or did not know any of them. With these limitations, there were eight students in the Wireframe subgroup and eleven students in the Solid subgroup who played a known piece and an unknown piece. The Mann-Whitney U tests were applied to check whether or not there were differences in the scores given by the teachers to the students between playing a known vs. unknown piece when analyzing the two groups separately. The results are shown in the last two rows of Table 1. The subgroup of students who played the known piece using the Wireframe mode and the subgroup of students who played the known piece using the Solid mode were also analyzed. The subgroups that used the two visualization modes for playing the unknown piece were also analyzed. No statistical differences were found for any of these analyses mentioned above.

The Mann-Whitney U test was applied to check whether or not there were differences for the participants own assessment (how they thought they had played the piano) between the two groups of participants. This analysis is included in the Self-assessment row of Table 1. No statistically significant differences were found for their own assessment.

Preferred visualization mode

This section focuses on the users' preferred visualization mode. A total of 72.7% of the users showed their preference for the Solid mode in PR1 and PR3. A total of 66.7% preferred the Solid mode for PR2. In PR5, a total of 48.5% of the users showed their preference for the Solid mode, 33.3% of the users showed their preference for the two modes equally, 12.1% of the users showed their preference for the Wireframe mode, and, finally, only 6.1% of the users selected none (2 students aged 67 and 69).

DISCUSSION

In this work, it has been demonstrated that programming different techniques using the API Passthrough, despite its current limitations, can create satisfactory MR experiences. Therefore, it has enormous potential for learning to play the piano and for other cases in which the current limitations are not incompatible with a satisfactory experience for the user. The limitations of the API Passthrough detected in the current version are the following:

- ➔ It does not allow the detection of objects or the application of computer vision techniques.
- ➔ It does not allow rendering using depth.
- ➔ The image can be unstable (moving and curving), and it is in grayscale.

To assess the accuracy of Passthrough in our application, images were captured using Oculus SideQuest streaming under different lighting settings (including very low natural light and very low artificial light) and at different distances from the camera position to the center of the keyboard (from 0.3 meters to 1 meter). We observed that the accuracy of Passthrough depends on both the distance to the target (in this case, the piano keyboard) and the specific lighting conditions. At distances greater than or equal to 0.6 meters, distortion is not perceived; however, the image appears more blurred and becomes slightly blurrier as the distance increases. Lighting plays an important role, and, in situations with more light, the distortion is somewhat amplified. In some specific cases, direct light reflections (specular reflection) on the keyboard generate a local distortion that can quickly disappear when the point of view changes. The optimal distance is around 0.5-0.6 meters (which, in many cases, coincides with the usual distance of the posture for playing the piano). For all of these reasons, our recommendations are to not stand at a distance of less than 0.4 meters and to avoid direct strong light on the target in order to prevent specular reflections.

In Wireframe mode the user has to look at the piano and is seeing mostly the real piano. Due to these short-distance Passthrough limitations, the students in the Wireframe mode experienced significant adverse symptoms (14.49), but those were not considered bad (>20). This is one of the reasons why, although no significant differences were found for any of the variables analyzed between the two groups, the students mostly preferred the Solid mode as the visualization mode.

These limitations could be overcome in the near future since Meta has recently announced that one of its next glasses will support Passthrough in color. Without a doubt, these improvements would facilitate the creation of much more immersive MR experiences in different fields.

Our application helps to locate the key on the keyboard quickly and efficiently and also to locate start and end of the sound in time. These are valuable and motivating aspects that were identified by both the students and the teachers. The application is valuable for working on some aspects of psychomotricity and spatiality, which are usually skills that need to be developed, especially by people who are just starting out in the study of the instrument. The application is motivating because it helps to develop these skills (spatiality and psychomotricity). The ratings of the students for the entry motivating-demotivating of the UEQ (with values from 1 to 7) show the students' motivation. All of the students considered the two visualization modes to be motivating, selecting option 1 (71%), 2 (18%) and 3 (11%)) for the Wireframe mode, and 1 (75%), 2 (19%), and 3 (6%) for the Solid mode. The students gave higher values to the Solid mode, but there were not significant differences.

From the students' comments and our own observations during the study, the superimposition of help elements directly on the real piano has advantages over other types of systems, such as sheet music notation (Finale) or non-projected piano roll notation (Synthesia). This conclusion is in line with previous works [3]. Another advantage of Passthrough is that the student can see the teacher who can make some indications with her/his hands while the student is playing.

Finally, we mention some opinions included in the comments of the participants. Quite a few students expressed their satisfaction with the virtual hand that was superimposed on the real hand and followed it naturally. Quite a few students indicated that they were not bothered by the fact that the real world was in grayscale, and some of them had not even noticed. Several students indicated that they were able to play pieces they did not know and previously had not felt capable of playing. These comments corroborate the satisfaction collected in the questionnaires. Only two 8-year-olds students reported that the Oculus Quest 2 was heavy for them. The Quest 2 Elite Strap with Battery, which weighs 500 grams, was used. Although, this strap generally provides greater comfort, for smaller children, it would be preferable to use the standard strap.

CONCLUSION

In this work, we have exploited the potential of the Oculus Passthrough and minimize its drawbacks for developing a satisfying MR experience for users applied to learning to play the piano as a use case. We included two visualization modes (border lines on all of the keys vs. solid color hiding the real keys). The results show that the two visualization modes help to play the piano. The subjective ratings of users regarding user experience, presence, and adverse effects of wearing glasses were independent of the visualization mode, gender, and age. However, when students were explicitly asked which one of the two visualization modes they preferred, they mostly selected the Solid mode. This preference indicates that, although not significantly, they detected the limitations of Passthrough at short distances. Therefore, this indicates that short-distance Passthrough limitations must be minimized in order to develop satisfactory MR experiences.

As future work, the details of this work can be used for the development of new applications for different purposes. Our application can be extended in several aspects; for example, an option could be added to hide the virtual hands or to show only the skeleton of the hand. The application in its current state can be used to carry out other studies. A study could determine differences between piano students and people who do not know how to play the piano. Our proposal could also be compared with baseline approaches (e.g, Synthesia) and other systems for learning piano using different technologies (e.g., XR headsets; hand-tracking).

ACKNOWLEDGMENTS

We thank the Music School U.M. Santa Cecilia of Benicàssim, Spain, and its principal David Ramón, Nathalie Paniagua, and all of the people who participated in the study. We would like to thank the editor and reviewers for their valuable suggestions.

REFERENCES

[1] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano AR: A Markerless Augmented Reality Based Piano Teaching System," in *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2011, vol. 2, pp. 47–52, doi: 10.1109/IHMSC.2011.82.

[2] J. Chow, H. Feng, R. Amor, and B. C. Wünsche, "Music Education Using Augmented Reality with a Head Mounted Display," in *Proceedings of the Fourteenth Australasian User Interface Conference - Volume 139*, 2013, pp. 73–79.

[3] K. Rogers *et al.*, "P.I.A.N.O.: Faster Piano Learning with Interactive Projection," in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, 2014, pp. 149–158, doi: 10.1145/2669485.2669514.

[4] J. Desnoyers-Stewart, D. Gerhard, and M. Smith, "Mixed Reality MIDI Keyboard," in *13th International Symposium on CMMR,* 2017, pp. 376–386.

[5] L. Rigby, B. C. Wünsche, and A. Shaw, "PiARno - An Augmented Reality Piano Tutor," in *Proceedings of the 32nd Australian Conference on Human-Computer Interaction*, 2020, pp. 481–491, doi: 10.1145/3441000.3441039.

[6] W. Molloy, E. Huang, and B. C. Wünsche, "Mixed Reality Piano Tutor: A Gamified Piano Practice Environment," in *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, 2019, pp. 1–7, doi: 10.23919/ELINFOCOM.2019.8706474.

[7] A. Birhanu and S. Rank, "KeynVision: Exploring Piano Pedagogy in Mixed Reality," in *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, 2017, pp. 299–304, doi: 10.1145/3130859.3131336.

[8] D. Molero, S. Schez-Sobrino, D. Vallejo, C. Glez-Morcillo, and J. Albusac, "A novel approach to learning music and piano based on mixed reality and gamification," *Multimed. Tools Appl.*, vol. 80, no. 1, pp. 165–186, 2021, doi: 10.1007/s11042-020-09678-9.

[9] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," *Int. J. Aviat. Psychol.*, vol. 3, no. 3, pp. 203–220, 1993, doi: 10.1207/s15327108ijap0303_3.

[10] H. Regenbrecht and T. Schubert, "Measuring presence in augmented reality environments: Design and a first test of a questionnaire," in *Proc. 5th Annu. Int. Workshop Presence*, 2002, pp. 1–7, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.581.2484&rep=rep1&type=pdf.

[11] F. D. Davis, "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts," *Int. J. Man. Mach. Stud.*, vol. 38, no. 3, pp. 475–487, 1993, doi: https://doi.org/10.1006/imms.1993.1022.

[12] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work*, 2008, pp. 63–76.

[13] P. Bimberg, T. Weissker, and A. Kulik, "On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 464–467, doi: 10.1109/VRW50115.2020.00098.

Mariano Banquiero was a researcher at GIGC UTN and INTEC, Argentina. He is currently a researcher at the Universitat Politècnica de València, Spain. His current research interests include computer graphics, extended reality, and video-game engine development. He is one of the corresponding authors of this article. Contact him at marianobanquiero@gmail.com.

Gracia Valdeolivas is an Assistant Professor at the Universitat Jaume I and a teacher at the Music School U.M. Santa Cecilia of Benicàssim. Contact her at valdeoli@uji.es.

Sergio Trincado is a teacher at the Music School U.M. Santa Cecilia of Benicàssim. Contact him at sergio.trincado.ruano@gmail.com.

Natasha García is a teacher at the Music School U.M. Santa Cecilia of Benicàssim. Contact her at natasha.garcia@iepgroup.es.

M.-Carmen Juan is a Full Professor of Computer Science with the Universitat Politècnica de València, Spain, where she has taught since 1996. Her main research interests include computer graphics, extended reality, advanced user interfaces and their applications to learning and psychology. She is one of the corresponding authors of this article. Contact her at mcarmen@dsic.upv.es.