



Revisiting natural user interaction in virtual world

Tam V. Nguyen¹ · Shreyas Raghunath¹ · Kim Anh Phung¹ · Tom Ongwere¹ · Minh-Triet Tran^{2,3,4}

Received: 8 March 2022 / Accepted: 28 November 2022 / Published online: 7 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Natural user interaction in virtual environment is a prominent factor in any mixed reality applications. In this paper, we revisit the assessment of natural user interaction via a case study of a virtual aquarium. Viewers with the wearable headsets are able to interact with virtual objects via head orientation, gaze, gesture, and visual markers. The virtual environment is operated on both Google Cardboard and HoloLens, the two popular wireless head-mounted displays. Evaluation results reveal the preferences of users over different natural user interaction methods.

Keywords Mixed reality · Virtual reality · User interaction · Evaluation

1 Introduction

In recent years, mixed reality (MxR) applications including virtual reality and augmented reality have become very popular in various domains such as education, knowledge dissemination, healthcare, entertainment, and manufacturing. The most common thing in these applications is a display in the form of a smartphone or a headset. These applications mainly focus on providing good graphics for immersive environment. In fact, the user interaction is also important since it connects the users to the virtual world. In practice, the MxR users are able to engage with virtual objects in the applications via various interaction interfaces such as tangible, devices, or sensors.

The existence of aquariums in most regions of the world creates knowledge dissemination opportunities across a diverse and multi-cultural audience. The aquarium provides a great experience to encounter the wonders of marine life. However, the recent COVID-19 pandemic (COVID-19 pandemic 2022) radically transforms the real world. The pandemic has since affected our daily life. Many restaurants,

bars, gyms, parks, and recreational areas are closed. People are wearing masks and avoiding crowds. These directives could lead to an increased demand for mixed-reality applications. Besides, the recent reports show that the need for mixed-reality applications will boost over the next decade (Augmented Reality Gets Pandemic Boost 2022). While most aquariums in the US are closed to the public in compliance with health orders for zoos and aquariums (Columbus Zoo and Aquarium 2022), in this study, we aim to offer ways for people to continue connecting with sea animals and learn about marine life. Figure 1 illustrates the setting of the real-life aquarium and the virtual aquarium. The virtual aquarium offers several new virtual opportunities to meet animals in close proximity. In that, people (viewers) may have virtual meet-ups with a live guest appearance such as a fish, sea lion, penguin, or sea otter. Viewers learn more about their favorite animals via interaction such as gesture, or visual markers. Therefore, a virtual aquarium is an ideal platform for us to evaluate the natural user interaction in virtual environment.

Our contributions are summarized as follows. We first introduce the main components of a feasible solution of a virtual aquarium, which can be utilized as an efficient and interesting approach for education and knowledge dissemination. We integrate various interaction methods in order to enhance the user experience. We later study the impact of different interaction methods in mixed reality applications. In particular, we conduct an extensive user study to evaluate different interaction methods.

✉ Tam V. Nguyen
tamnguyen@udayton.edu

¹ University of Dayton, Dayton, OH, USA

² University of Science, Ho Chi Minh City, Vietnam

³ John von Neumann Institute, VNU-HCM, Ho Chi Minh City, Vietnam

⁴ Vietnam National University, Ho Chi Minh City, Vietnam



Fig. 1 The conventional real-life aquarium (left—image courtesy of Georgia Aquarium) vs. virtual aquarium (right—our work) with the tiger shark

The remainder of this paper is structured as follows. We conduct a brief review of related works in Sect. 2. Then, Sect. 3 introduces the virtual aquarium application with different interaction methods. Next, Sect. 4 presents the experiments and discusses the experimental results. Finally, the paper is concluded in Sect. 5 and we further present some prospective future works.

2 Related work

2.1 Mixed reality applications

In this subsection, we explore different mixed reality (MxR) applications including virtual reality (VR) and augmented reality (AR).

Virtual reality offers an experience of a fully artificial environment to users. The users are totally immersed in such a virtual world and interact with virtual objects via game controllers, sensors, or gestures. As an example, Nguyen and Sepulveda (2016) used gestures to support user interaction inside the virtual environment. Since the users only view the virtual world, there have been many efforts to integrate other human senses. Ranasinghe et al. (2017) integrated thermal and wind modules in order to help users perceive the ambient temperatures and wind conditions in the virtual environment. Similarly, Zhu et al. (2019) developed a thermal haptic feedback in a form of a smart ring. They found that the users are able to feel the temperature accurately. In another work, Tran et al. (2018) studied the users' speed change perception via cycling in VR. In particular, they developed a bike prototype that simulates cycling techniques such as bike pedaling or throttling.

Meanwhile, augmented reality attempts to expand reality by overlaying virtual objects into the real environment. Kato and Billinghurst (2004) released AR Toolkit which overlays virtual objects on top of a binary marker. Later, Fiala (2005)

developed ARTag markers with bi-tonal planar patterns to improve the marker's recognition and pose estimation rate. Liu et al. (2012) introduced an AR system, magic closet, for autonomously pairing and rendering clothes based on the input occasion.

There have been efforts to define mixed reality (Speicher et al. 2019; Mystakidis 2022). In practice, mixed reality applications involve both physical reality and virtual reality. In such a hybrid environment, the virtual objects are overlaid and mapped to the real world (as in augmented reality) and users are allowed to interact with objects via devices, sensors, or gestures (as in virtual reality). Nguyen et al. (2016) introduced a mixed reality application allowing users to design interactive repair manuals by using Moverio smart glasses (Epson 2022). Then, other users can download the designed manuals to fix their own items via pre-designed instructions.

There have been many mixed reality applications in education and training. In fact, the visualization of 3D content and the manipulation of virtual object is an entry level in taxonomies of immersive instructional design methods (Mystakidis et al. 2022; Cruz et al. 2021). Kim et al. (2020) developed a theoretical framework including authentic experience, cognitive and affective responses, attachment, and visit intention with tourism education using a stimulus-organism-response model. Nguyen et al. (2020) developed a MxR system on HoloLens (2022) for nondestructive evaluation training (NDE). The users interact with virtual objects such as ultrasound transducer and inspection objects within an NDE session via gestures and gaze. Butail et al. (2012) created a virtual environment where fish is put in the fish tank. However, the setting is simple with only one fish and there is no user study. Recently, Wißmann et al. (2020) introduced an approach to fish tank virtual reality with semi-automatic calibration support. However, the system only works properly if the glasses are front-facing the monitor. Small head rotations to the left or right are

possible, but more extensive rotations may cause shutter synchronization errors. In addition, they did not conduct the user study to evaluate the user interaction. Araiza-Alba et al. (2021) used 360-degree virtual reality videos to teach water-safety skills to children. However, there is no user interaction with 360-degree videos. Čejka et al. (2020) and Sinnott et al. (2019) proposed AR/VR underwater systems. Čejka et al. developed a prototype run on a smartphone sealed in a waterproof case and uses a hybrid approach (markers and inertial sensors) to localize the diver on the site. Sinnott et al. (2019) created an underwater head mounted display by adapting a full-face scuba mask to house a smartphone and lenses. Here, these systems are impractical since they require actual water environment such as a swimming pool which is not always available to the end users.

2.2 User interaction in virtual world

User interaction in virtual world plays a crucial role in mixed reality applications. User interaction nurtures a contextual user engagement with the virtual content. Previously the VR headsets blocked the natural view of the user to the real environment. Due to this engineering problem, the users were not able to directly interact with the real world. However, the release of modern see-through headsets, such as HoloLens allowed the interaction between users and the physical world. As reviewed in Bekele and Champion (2019), there are four primary interaction methods, namely, tangible, device-based, sensor-based, and multimodal interaction methods.

- **Tangible interaction** involves physical objects to support the user interaction with virtual objects. For example, Araújo et al. (2016) developed a tangible VR using an interactive cube. The cube side has special line patterns for ease of recognition. Later, Harley et al. (2017) added visual and auditory haptic feedback into the cube. The tangible interfaces must be customized, therefore, these interfaces are not available to massive users.
- **Device-based interaction** adopts graphical user interfaces and/or external devices, for example, computer mouse, gamepad, joystick to help the user engage with virtual objects. Here, the visual marker can be considered as a device-based interface. As an example, Henderson and Feiner (2009) used visual markers in an AR prototype for vehicle maintenance.
- **Sensor-based interaction** utilizes sensing devices such as motion tracker and gaze tracker to understand user inputs. For example, Avgoustinov et al. (2011) developed an NDE training simulation by analyzing the gaze and gesture from users.
- **Multimodal interaction** fuses aforementioned methods, i.e., tangible, device-based, sensor-based inter-

action techniques to sense and perceive human interaction. For example, this kind of interaction allows devices along with gestural, and gaze-based to engage in interaction with virtual objects.

Bekele and Champion (2019) compared different interaction methods for cultural learning in virtual reality. However, there was no user study conducted. Recently, Yang et al. (2019) reviewed different interaction methods in the virtual environment. They also discussed the application of gesture interaction system in virtual reality. However, there is no evaluation from actual users. FaceDisplay Gugenheimer et al. (2018) is a modified VR headset consisting of three tangible displays and a depth camera attached to its back. Surrounding people can perceive the virtual world through the displays and interact with the headset-worn user via touch or gestures. However, this touch screen-based interaction is unnatural and inconvenient for both the user wearing headset and surrounding people. Meanwhile, Kang et al. (2020) only analyzed the simple interaction such as moving (translation) in virtual reality.

2.3 Experience of mixed reality users

The MxR experience is often determined by powerful hardware and actuators (Pittarello 2017). The hardware provides an engagement for the visual experience. Actuators provide user inputs and outputs, and support mapping users' actions in their real-life experience to their actions in the virtual experience. The implementation of these experiences in VR and AR technologies has been reported in some studies. For example, there are (i) animal-free aquariums that provide visitors with an authentic in-the-wild experience (Jung et al. 2013), and (ii) storytelling virtual theaters that provide visitors an engaging and educational experience about endangered animal species (Dooley 2017; Daut 2020). However, these technologies are installed and are operating in designated and often expensive areas. We have yet to see work that discusses how such technologies can engage users in a less expensive and personal way. The research that has explored user interaction with MxR tools is still complex and has failed to detail the standards for user engagement and interaction preferences (Pittarello 2017). User engagement, interaction, and preferences are important components of MxR and are relevant in describing the MxR experience. In this paper, we adapt six parameters to explore various interaction modes and capture several facets of the MxR usage when connecting with and learning about marine life. These parameters include ease of use, convenience, preference, engagement, and motivation.

Fig. 2 The system overview of the virtual aquarium used in this research

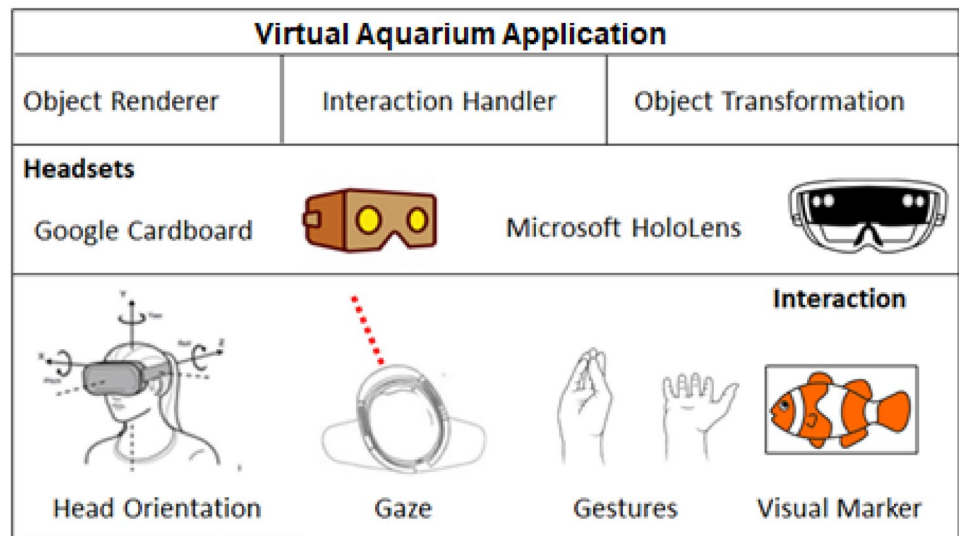


Fig. 3 The headsets in mixed reality applications. From left to right: Google Cardboard, Oculus Rift, HTC Vive, Epson Moverio, and Microsoft HoloLens. Note that Google Cardboard and Microsoft HoloLens are under wireless setting

3 Virtual aquarium application

3.1 Overview

Here we briefly introduce our virtual aquarium framework. We designed our framework following a modular design approach for the extensibility and re-usability purposes in different application domains. The main components of the framework are shown in Fig. 2.

As a quick glimpse, the *object renderer* is responsible for rendering virtual objects and superimposing them into the real or virtual scene. Meanwhile, the *interaction handler* analyzes the user input/interaction and provides the corresponding feedback. For example, users may interact with virtual objects using the gaze, gesture, and visual marker. The *object transformation* is used to perform certain Affine transformation on the virtual objects such as scaling, translating, rotating according to the user interaction. Besides, the *headsets* are used to display the virtual environment to the users. The details of different components are listed in the following subsections.

3.2 Headsets

A headset is one of the essential devices for any mixed reality application. It comprises a tiny display optic in front of one eye (monocular vision) or each eye (binocular vision) to render stereoscopic views and head motion tracking sensors.

Google Cardboard (2022) is the most simple yet effective VR headset (The Best VR Headsets for PC 2022). As shown in Fig. 3, a smartphone serves as the display and the main processor of the headset, and a cardboard box is a holder. To enhance the usability for users, various extra features such as trackers, stereo sound, and controllers have been integrated into other VR headsets, i.e., Rift (2022) and Vive (2022). In most VR headsets, the main computation and visual rendering processes are performed in an external PC/laptop connected to the headsets via wired connections. Besides, the VR users cannot feel the real environment since they are immersed in the virtual world. Thus, it is inconvenient for users to walk naturally in VR, and they should simulate their movement via a gamepad or joystick.

To overcome issues with VR headsets, the Epson Movio glasses (Epson 2022) comprises two components, the binocular and mini-projectors. The binocular has a see-through capability via the transparent display, and mini-projectors are responsible for rendering virtual objects. In this setting, users can see the virtual objects augmented in the real environment. However, its bulky setup, such as a separate heavy battery, main processor, and external controller, prevents users from long working sessions. Therefore, Microsoft introduced HoloLens (2022) which integrates the display, processor, and battery within the same unit. It also has a built-in eye tracker for gaze detection, front-facing cameras for gesture recognition, and a depth sensor for spatial mapping.

From the list above, we opt to use Google Cardboard and HoloLens for two reasons. First, they are both wireless headsets which are more convenient for viewers than wired ones. Second, they represent two extreme settings, i.e., Google Cardboard for the fully immersive virtual environment whereas HoloLens for the mixed reality environment of the real world and the overlaid virtual content.

3.3 Integration of interaction methods

In this work, we specifically focused on the interactions related to head orientation, hand gestures, gaze, and visual markers. Below we provide the details of each interaction method.

3.3.1 Head orientation

When detecting the orientation of the user's head, the headset uses data from accelerometers and gyroscopes for inertial tracking. In particular, the output of the accelerometers is used to find the velocity and then estimate the position. Meanwhile, the angular velocity is measured by gyroscopes to determine the angular position relative to the initial point.

In our system, the important modality for input is user gaze. A user can target and select a virtual object in the mixed reality environment via user gaze. The gaze determines the location where the user is looking at. Here the gaze direction can be formed via the camera position and the center point of the viewport. This can be considered as the fixed gaze.

3.3.2 Human gaze input

Unlike the fixed gaze mentioned above, the modern headset such as HoloLens integrates a built-in eye tracker. During the spatial mapping phase, HoloLens continuously scans through the real surrounding environment to map spatial meshes in the real world with those meshes in the virtual world. Then, the built-in eye tracker detects the human gaze,

namely, the position vector $g(x_g, y_g, z_g)$ and the direction vector $f(x_f, y_f, z_f)$. Here, the gaze direction vector f is a ray casting straight through the gaze vector g .

3.3.3 Gesture input

In our prior work (Nguyen et al. 2020), we show how users engage with virtual content and applications by using their hand gestures. The users may initiate the interactions via simple gestures, for example, raising and tapping down the index finger. Users manipulate the virtual objects or navigate the virtual environment using the relative motions of their hands.

3.3.4 Visual marker input

Visual markers can be in the form of binary markers or natural images. They are pre-scanned models of feature points (the key points detected in the markers). The integrated computer vision component such as dos Santos (2016) performs marker detection and marker tracking. When the marker is detected and tracked, the marker's feature points are used to estimate the pose of the marker from the user viewpoint. Thanks to the estimated pose, virtual objects can be augmented (superimposed) into the scene on the visual markers.

3.4 System implementation

3.4.1 Object renderer and transformation

The virtual aquarium is composed of three sets of rendered objects. These include fish species, non-fish species, and landscape items. Figure 4 illustrates some 3D models used in our virtual aquarium. Each object has an origin point and a transformation matrix. The matrix stores and manipulates the position, rotation, and scale of the object.

We follow Huth and Wissel model (1992) to simulate the fish movement. Each fish swims either individually or in a school. Every fish swims within the school that has the same behavior model. The motion of the model fish school is not affected by external influences. Here, we consider the position p_i , orientation o_i , and velocity v_i for each fish i . The position $p_i(x_i, y_i, z_i)$ at each time-step Δt is defined based on the Huth and Wissel (1992):

$$p_i(t + \Delta t) = p_i(t) + \Delta t \times v_i(t + \Delta t), \quad (1)$$

where $v_i(t + \Delta t)$ and $o_i(t + \Delta t)$ are defined as follows:

$$v_i(t + \Delta t) = o_i(t + \Delta t) \times v_i(t), \quad (2)$$

$$o_i(t + \Delta t) = \begin{bmatrix} \cos \alpha_i(t) & -\sin \alpha_i(t) & 0 \\ \sin \alpha_i(t) & \cos \alpha_i(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \times o_i(t). \quad (3)$$



Fig. 4 The underwater 3D models used in our virtual aquarium

Algorithm 1 Attentive Object Retrieval

Input: User gaze: g , and f ;
 n virtual objects: $\{obj_1, \dots, obj_n\}$
Output: The attentive object.
Process flow:

```

1:  $result \leftarrow null$ 
2:  $min_d \leftarrow -1$ 
3: for  $i=1 \rightarrow n$  do
4:   Compute  $h_i, d_i$  by Eq. 4
5:   if  $h_i \neq null$  and  $obj_i.isAnimated()$  then
6:     if  $min_d \leq d_i$  then
7:        $result \leftarrow obj_i$ 
8:     end if
9:   end if
10: end for
11: Return  $result$ .

```

To facilitate the implementation, we permit each fish to swim at a certain depth z_i . Meanwhile, the turning angle, α_i , is determined using the same behavior model described by Huth and Wissel (1992).

3.4.2 Interaction handler

Regarding the user gaze, we compute the hit point $h(x_h, y_h, z_h)$. The hit point is the intersection between the gaze vector g with the direction f and the object's mesh:

$$h = g + d \times f, \quad (4)$$

where d is the distance between the gaze and object hit point. Algorithm 1 describes the pseudo code to identify the attentive virtual object (aquarium animals). Note that the attentive object must contain animations and gaze handler implementation. Then, the attentive object will provide the feedback, for example, it performs certain animation and a

narrative sound raises to introduce the animal. Regarding the visual marker, we opt to natural images for the markers instead of using binary images. In particular, we adopt Vufo-ria Engine dos Santos (2016) as mentioned earlier.

3.4.3 System integration

We adopt Unity3D engine¹ to implement the proposed system. The engine provides packages for mixed reality development and also supports cross-platform features. These packages allow developers to run the applications onto different devices, including smartphones (used in Google Cardboard) and HoloLens.

We first insert the 3D models, i.e., fish, non-fish, and landscape object instances, into the Unity3D scenes. Each object instance is attached with a C# programming script. For example, each object is assigned to a random trajectory. We also group several schools of fish to simulate the actual aquarium. Regarding the interaction feedback, we display the pop-up information and play the voice for the models designed based on the ray cast hit point h . For the visual marker, we adopted dos Santos (2016) to detect and track the marker. Note that our framework can be easily adopted to different education domains. For example, the underwater 3D models can be replaced as models of paintings or sculptures for a virtual museum. Another application example is the virtual heritage where monuments, buildings and objects with historical, archaeological, or anthropological value can be integrated into our system. Figure 5 illustrates the user interface of the virtual aquarium system in different interaction modes, namely, Google Cardboard, HoloLens with gaze-only, HoloLens with gaze and gesture, HoloLens with multimodal of gaze, gesture, and visual markers.

¹ <https://unity3d.com>.

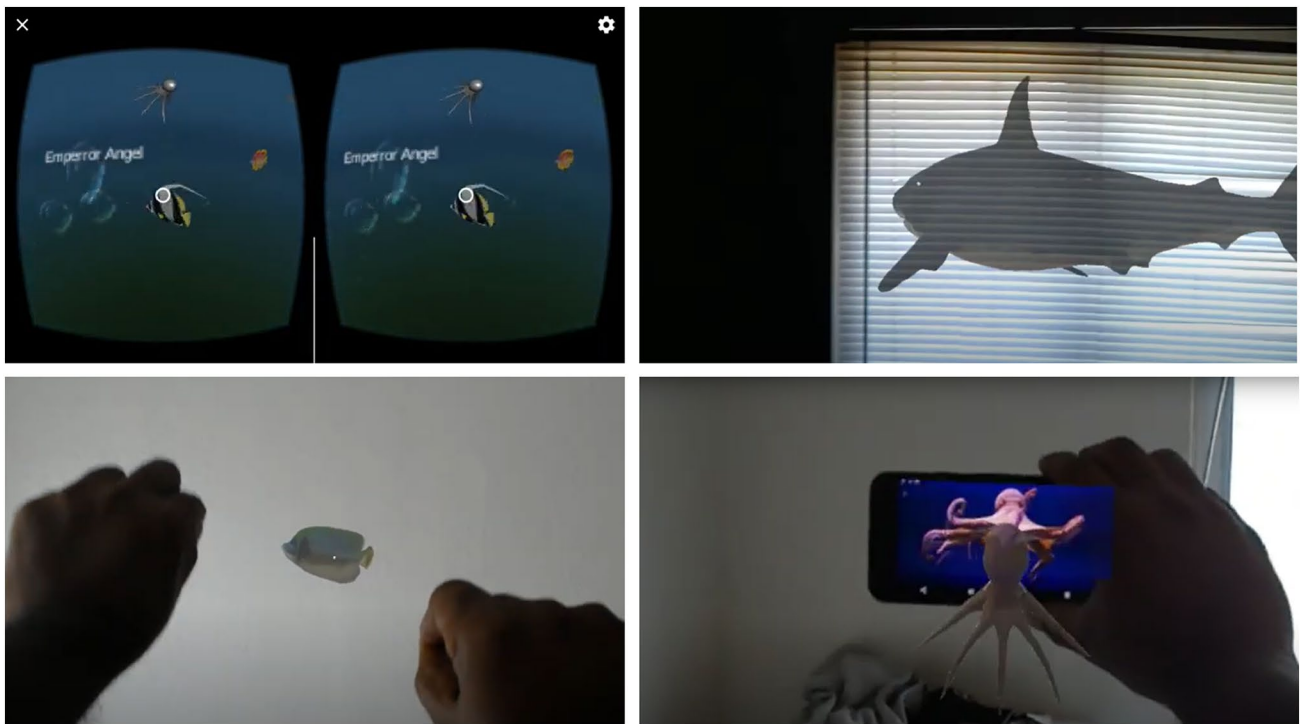


Fig. 5 The interface of different interaction methods. From left to right, top to bottom: Google Cardboard, HoloLens with gaze only, HoloLens with gaze and gesture, HoloLens with gaze, gesture, and visual markers

4 Evaluation

4.1 Participants and experimental settings

Our study received approval from the IRB at the University of Dayton. We follow (Kaur 1997) and (Suárez-Warden et al. 2015) for the design methodology and sample size, respectively. The participants who have no or limited prior knowledge in using mixed reality headsets (e.g., Google Cardboard and HoloLens) are eligible to take part in the user study. The mixed reality application was designed to be simple enough for the novice to contribute to without compromising the integrity of the results. In total, 51 people participated in this study, 18 of these participants identified themselves as female. The participants are university students and staff, whose ages range from 18 to 60 ($\mu = 29.4$). We provided participants the instructions for the experiment after they completed the consent form. Table 1 lists the configuration comparison of different interaction modes. Each participant took part in a session, namely, a 5-min trial followed by 1-min break for each interaction method. We observed no cybersickness from participants after every 5-min short trials. We then showed the questionnaire and asked for feedback regarding the following perspectives:

- Ease of use: How easy is the method?
- Convenience: How convenient is the method?

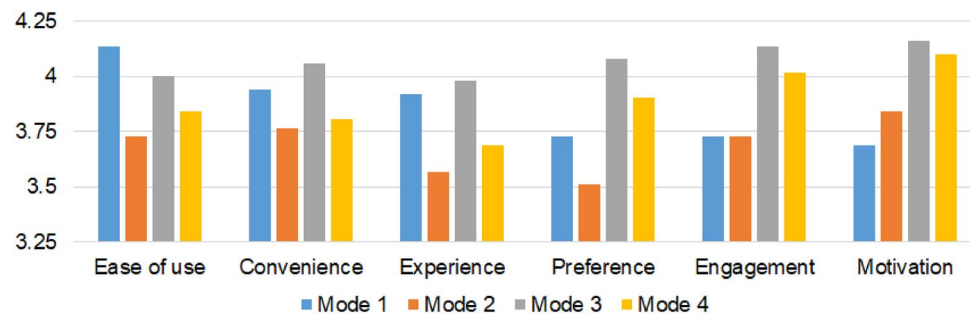
- Experience: How much does the interaction mode help you experience the environment?
- Preference: How much do you prefer a certain interaction mode over the others?
- Engagement: How much does the method engage the user?
- Motivation: How much does the method increase the involvement of the user?

The ‘*ease of use*’ is the most popular criterion in literature (Venkatesh 2000; Ahn et al. 2017; Gopalan et al. 2016). Meanwhile, ‘*convenience*’ and ‘*experience*’ were included in Nguyen and Sepulveda (2016) and Kim et al. (2020). In addition, the ‘*preference*’ criterion was studied in Nguyen et al. (2018) and Cicek et al. (2021). Also, the criteria

Table 1 The configuration comparison of different interaction modes in our experiments

	Device	Head orientation	Gaze	Gesture	Visual marker
Mode 1	Google Cardboard	✓	c		
Mode 2	HoloLens	✓	✓		
Mode 3	HoloLens	✓	✓	✓	
Mode 4	HoloLens	✓	✓	✓	✓

Fig. 6 The average rating scores from the user study evaluation



‘engagement’ and ‘motivation’ were included in Gopalan et al. (2016), Huang et al. (2021) and Bekele and Champion (2019). Therefore, we included all the aforementioned criteria in our study. The participant rated each interaction mode on a 5-point Likert scale (Likert 1932) from the best (5) to the worst (1) for each criterion.

4.2 Experimental results

Figure 6 shows the average scores of different interaction modes for the aforementioned criteria. Mode 1 (Google Cardboard) is highly rated for *ease of use*. There is not much instruction on using Google Cardboard. The participants simply wear the Google Cardboard and use head rotation to observe the virtual environment. Meanwhile, the participants need to learn how to use HoloLens, such as focusing gaze or using gestures. Note that the consecutive wrong gestures that are not recognized by HoloLens frustrate the users.

Regarding the *convenience* and *experience*, Mode 3 (HoloLens with gaze and gesture) and Mode 1 achieve the highest and the second highest rates, respectively. The two modes are convenient to users. Meanwhile, Mode 2 (HoloLens with gaze only) is not as convenient as Mode 1 since the device is heavier and there is not much different in terms of functionality. However, Mode 3 (HoloLens with gaze and gesture) is preferred due to the usage of human gestures.

In terms of *engagement* and *motivation*, Mode 3 and Mode 4 (HoloLens with gaze, gesture, and visual marker) obtain the top-2 rates. The main reason that Mode 3 outperforms Mode 4 can be explained via the usage of the visual marker itself. Actually, the use of visual markers is very interesting to participants. However, it interrupts the user experience. Indeed the participant needs to use one hand to hold the visual marker. Therefore, it is inconvenient to use the other hand for gesture-based interaction. Meanwhile, Mode 1 and Mode 2 achieve the lowest rates due to the limited interaction.

Regarding the *preference*, the participants favor Mode 3 and Mode 4 over Mode 2 (HoloLens with gaze only) and Mode 1 (Google Cardboard). The participants appreciate the usage of gestures and no strict spatial requirement when using HoloLens.

Next, we apply a statistical significance test to verify if the user ratings of two interaction modes are equivalent. This is regarded as the null hypothesis, for example, $H_0 : r(CV_{Mode1}) = r(CV_{Mode2})$, where $r(CV_{Mode1})$ and $r(CV_{Mode2})$ are the user ratings with regards to the convenience from Mode 1 and Mode 2, respectively. In other words, the significance test verifies whether there is no difference between two sets of user ratings. Given a null hypothesis, we compute the p-value, namely, the probability of obtaining the observed ratings if the null hypothesis is true. In this work, we compute p-values via the Wilcoxon signed rank test (Wilcoxon 1992).

Table 2 presents the p-values of every pair of interaction mode ratings in terms of ease of use (in the lower diagonal) and convenience (in the upper diagonal). We note that the p-values of (EU_{Mode1}, EU_{Mode3}) and (EU_{Mode1}, EU_{Mode4}) are ≤ 0.05 meaning that the null hypothesis is rejected to the two interaction modes. Therefore, the differences between Mode 1 and Mode 2/Mode 4 are statistically significant. Likewise, the null hypothesis is rejected to (Mode 1 and Mode 2) and (Mode 1 and Mode 4) in terms of EU. However, the p-value of (Mode 1 and Mode 3) with regard to EU is > 0.05 meaning that the null hypothesis is true. In other words, Google Cardboard and HoloLens (with gaze and gesture) share the similar ease of use. All p-values with regard to CV are > 0.05 meaning that the four interaction modes have similar convenience ratings.

For the remaining criteria, Tables 3 and 4 show p-values of every pair of mode ratings (X_A and X_B), where X is the

Table 2 Wilcoxon signed rank test with p-values of every two compared mode ratings w.r.t. ease of use and convenience

	Convenience (CV)			
	Ease of use (EU)			
Mode 1	–	0.608	0.676	0.189
Mode 2	0.014	–	0.239	0.444
Mode 3	0.475	0.068	–	0.117
Mode 4	0.023	0.635	0.043	–
	Mode 1	Mode 2	Mode 3	Mode 4

p-values ≤ 0.05 are marked in boldface

Table 3 Wilcoxon signed rank test with p-values of every two compared mode ratings w.r.t. experience and preference

	Preference (PR)		Experience (XP)	
	Mode 1	Mode 2	Mode 3	Mode 4
Mode 1	–	0.701	0.063	0.827
Mode 2	0.006	–	0.003	0.537
Mode 3	0.616	0.033	–	0.050
Mode 4	0.031	0.419	0.011	–

p-values ≤ 0.05 are marked in boldface

Table 4 Wilcoxon signed rank test with p-values of every two compared mode ratings w.r.t. engagement and motivation

	Motivation (MV)		Engagement (EN)	
	Mode 1	Mode 2	Mode 3	Mode 4
Mode 1	–	0.108	0.043	0.097
Mode 2	0.960	–	0.310	0.537
Mode 3	0.010	0.004	–	0.831
Mode 4	0.136	0.055	0.492	–

p-values ≤ 0.05 are marked in boldface

rating in terms of experience (XP), preference (PR), engagement (EN), and motivation (MV), and A and B are the two interaction modes considered. The differences between Mode 3 (HoloLens with gaze and gesture) and other modes are statistically significant. In particular, the null hypothesis is rejected to (Mode 3 and Mode 2) and (Mode 3 and Mode 4) in terms of experience and preference. In addition, the null hypothesis is also rejected to (Mode 3 and Mode 1) in terms of engagement, motivation, and preference. The results clearly show that the usage of Mode 3 (HoloLens with gaze and gesture) for interaction in mixed reality is preferred by the participants.

4.3 Discussion

As mentioned earlier, the demand for mixed reality applications was boosted during the time of the pandemic (Augmented Reality Gets Pandemic Boost 2022). In our experiment, we show that users appreciate the usage of mixed reality applications for virtual aquariums. In particular, we show how the use of HoloLens with gaze and gesture-based interaction was favored by the participants. The participants highly rated the ability to observe and interact with the superimposed objects in the virtual aquarium. The user study indicates that the users prefer the addition of natural interaction such as gesture.

There needs to be an increase in the use of MxR applications in the form of VR and AR, so that experiences, such as a headset that provides 360°o imagery to the user's eyes and immersive audio to their ears, is possible. AR creates visual and auditory overlays on top of reality (Augmented Reality Gets Pandemic Boost 2022). During the pandemic, our MxR system (combining VR and AR) could create a new experience through headsets and visual makers that can change based on user preferences. This is an important component of the MxR system and is confirmed in our results section, where all of our participants agreed that head orientation is required for all interaction modes. Furthermore, our participants appreciated the ability to navigate inside the virtual aquarium which is difficult in the real-life aquarium.

We also found that the addition of visual markers is not convenient to the end users. In particular, when visual markers are being used, our results suggest that the visual marker should not be used to trigger the display of virtual objects. Instead, the visual markers should be used for the side information such as manual or in-app instruction. When visual makers are being used, future MxR tools should create an immersive interactive experience for the user and should add to the user's sense of agency and immersion (in a virtual aquarium for example). The participants also recommended applying our MxR system in a more complicated environment (for example manufacturing, architecture, and healthcare).

5 Conclusions and future work

In this paper, we revisit the assessment of natural user interaction in virtual world. In particular, we develop a virtual aquarium on two popular wireless headsets, namely, Google Cardboard and HoloLens. Different from other works in literature, we evaluate different user interaction methods such as head orientation, gaze, gesture and visual markers. Our analysis first shows that the ease of use is highly appreciated across the two headsets while the user's experience, preference, engagements and motivations vary from one mode to another. Our second finding is that the users prefer the addition of natural interaction such as gesture. Another finding is that the addition of visual markers is not convenient to the end users.

In the future, we will investigate methods to further improve the current system. In particular, we aim to extend this work from the aquarium to different virtual environments. Additionally, future studies should integrate visual markers in an appropriate form to increase engagement and convenience. Finally, we believe this work could attract more future research looking to assess the natural user interaction in mixed reality applications.

Funding This research is supported by National Science Foundation (NSF) under Grant no. 2025234 and Vingroup Innovation Foundation (VINIF) in project code VINIF.2019.DA19.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Ahn J, Choi S, Lee M, Kim K (2017) Investigating key user experience factors for virtual reality interactions. *J Ergon Soc Korea* 36:267–280
- Araiza-Alba P, Keane T, Matthews B, Simpson K, Strugnelli G, Chen WS, Kaufman J (2021) The potential of 360-degree virtual reality videos to teach water-safety skills to children. *Comput Educ* 163:104096
- Araújo B, Jota R, Perumal V, Yao J, Singh K, Wigdor D (2016) Snake charmer: physically enabling virtual objects. In: *Proceedings of the international conference on tangible, embedded, and embodied interaction*, pp 218–226
- Augmented Reality Gets Pandemic Boost (2022) <https://www.wsj.com/articles/augmented-reality-gets-pandemic-boost-11611866795>. Retrieved 28 Feb 2022
- Avgoustinov N, Boller C, Dobmann G, Wolter B (2011) Virtual reality in planning of non-destructive testing solutions. In: Bernard A (ed) *Global product development*, pp 705–710
- Bekele MK, Champion E (2019) A comparison of immersive realities and interaction methods: cultural learning in virtual heritage. *Front Robot AI* 6:91
- Bekele MK, Champion E (2019) A comparison of immersive realities and interaction methods: cultural learning in virtual heritage. *Front Robot AI* 6:91
- Butail S, Chicoli A, Paley DA (2012) Putting the fish in the fish tank: immersive vr for animal behavior experiments. In: *2012 IEEE international conference on robotics and automation*. IEEE, pp 5018–5023
- Čejka J, Zsíros A, Liarokapis F (2020) A hybrid augmented reality guide for underwater cultural heritage sites. *Pers Ubiquit Comput* 24:815–828
- Cicek I, Bernik A, Tomicic I (2021) Student thoughts on virtual reality in higher education—a survey questionnaire. *Information* 12:151
- Columbus Zoo and Aquarium canceling events over COVID-19 coronavirus concerns (2022). <https://www.dispatch.com/story/news/2020/03/13/columbus-zoo-aquarium-canceling-events/1530300007/>. Retrieved 28 Feb 2022
- COVID-19 pandemic (2022) <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Retrieved 28 Feb 2022
- Cruz A, Paredes H, Morgado L, Martins P (2021) Non-verbal aspects of collaboration in virtual worlds: a CSCW taxonomy-development proposal integrating the presence dimension. *J Univers Comput Sci* 27:913–954
- Daut M (2020) Immersive storytelling: leveraging the benefits and avoiding the pitfalls of immersive media in domes. In: *Handbook of research on the global impacts and roles of immersive media*, pp 237–263
- Dooley K (2017) Storytelling with virtual reality in 360-degrees: a new screen grammar. *Stud Australas Cine* 11:161–171
- dos Santos AB, Dourado JB, Bezerra A (2016) Artoolkit and qualcomm vuforia: an analytical collation. In: *Symposium on virtual and augmented reality*, pp 229–233
- Epson Moverio glasses (2022). <https://epson.com/moverio-augmented-reality>. Retrieved 28 Feb 2022
- Fiala M (2005) Artag, a fiducial marker system using digital techniques. In: *Computer vision and pattern recognition*, pp 590–596
- Google Cardboard (2022) <https://arvr.google.com/cardboard/>. Retrieved 28 February 2022
- Gopalan V, Zulkifli AN, Abubakar J (2016) A study of students motivation based on ease of use, engaging, enjoyment and fun using the augmented reality science textbook. *J Fac Eng* 31:27–35
- Gugenheimer J, Stemasov E, Sareen H, Rukzio E (2018) Facedisplay: towards asymmetric multi-user interaction for nomadic virtual reality. In: Mandryk RL, Hancock M, Perry M, Cox AL (eds) *Proceedings of the 2018 CHI conference on human factors in computing systems*, CHI 2018, Montreal, QC, Canada, April 21–26. ACM, p 54 (2018)
- Harley D, Tarun AP, Germinario D, Mazalek A (2017) Tangible VR: diegetic tangible objects for virtual reality narratives. In: *Proceedings of the 2017 conference on designing interactive systems*, pp 1253–1263
- Henderson SJ, Feiner S (2009) Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: *IEEE international symposium on mixed and augmented reality*, pp 135–144
- HTC Vive (2022) <https://www.vive.com>. Retrieved 28 Feb 2022
- Huang W, Roscoe RD, Johnson-Glenberg MC, Craig SD (2021) Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *J Comput Assist Learn* 37:745–758
- Huth A, Wissel C (1992) The simulation of the movement of fish schools. *J Theor Biol* 156:365–385
- Jung S, Choi YS, Choi JS, Koo BK, Lee WH (2013) Immersive virtual aquarium with real-walking navigation. In: *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*, pp 291–294
- Kang HJ, Shin J, Ponto K (2020) A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality. In: *IEEE conference on virtual reality and 3D user interfaces, VR 2010*, Atlanta, GA, USA, March 22–26, 2020. IEEE, pp 275–284
- Kato H, Billinghurst M (2004) Developing AR applications with artoolkit. In: *IEEE and ACM international symposium on mixed and augmented reality (ISMAR)*, p 305
- Kaur K (1997) Designing virtual environments for usability. In: *International conference on human-computer interaction*, pp 636–639
- Kim MJ, Lee CK, Jung T (2020) Exploring consumer behavior in virtual reality tourism using an extended stimulus-organism-response model. *J Travel Res* 59:69–89
- Likert R (1932) A technique for the measurement of attitudes. *Arch Psychol* 22(140):55
- Liu S, Nguyen TV, Feng J, Wang M, Yan S (2012) Hi, magic closet, tell me what to wear!. In: *Proceedings of ACM multimedia conference*, pp 1333–1334
- Microsoft HoloLens—Mixed Reality Technology (2022) <https://www.microsoft.com/en-us/hololens>. Retrieved 28 Feb 2022
- Mystakidis S (2022) Metaverse. *Encyclopedia* 2:486–497
- Mystakidis S, Christopoulos A, Pellas N (2022) A systematic mapping review of augmented reality applications to support STEM learning in higher education. *Educ Inf Technol* 27:1883–1927
- Nguyen TV, Kamma S, Adari V, Lesthaeghe T, Boehnlein T, Kramb V (2020) Mixed reality system for nondestructive evaluation training. In: *Virtual reality*, pp 1–10
- Nguyen TV, Mirza B, Tan D, Sepulveda J (2018) ASMIM: augmented reality authoring system for mobile interactive manuals. In: *Proceedings of international conference on ubiquitous information management and communication*, pp 3:1–3:6

- Nguyen TV, Tan D, Mirza B, Sepulveda J (2016) MARIM: mobile augmented reality for interactive manuals. In: Proceedings of ACM conference on multimedia conference, pp 689–690
- Nguyen TV, Sepulveda J (2016) Augmented immersion: video cutout and gesture-guided embedding for gaming applications. *J Multimed Tools Appl* 75:12351–12363
- Nguyen TV, Sepulveda J (2016) Augmented immersion: video cutout and gesture-guided embedding for gaming applications. *Multimed Tools Appl* 75:12351–12363
- Oculus Rift (2022) <https://www.oculus.com>. Retrieved 28 Feb 2022
- Pittarello F (2017) Experimenting with playvr, a virtual reality experience for the world of theater. In: Proceedings of the 12th biannual conference on Italian SIGCHI chapter, pp 1–10
- Ranasinghe N, Jain P, Karwita S, Tolley D, Do EY (2017) Ambiotherm: Enhancing sense of presence in virtual reality by simulating real-world environmental conditions. In: Proceedings of the 2017 CHI conference on human factors in computing systems, pp 1731–1742
- Sinnott C, Liu J, Matera C, Halow S, Jones A, Moroz M, Mulligan J, Crognale M, Folmer E, MacNeilage P (2019) Underwater virtual reality system for neutral buoyancy training: development and evaluation. In: 25th ACM symposium on virtual reality software and technology, pp 1–9
- Speicher M, Hall BD, Nebeling M (2019) What is mixed reality?. In: Brewster SA, Fitzpatrick G, Cox AL, Kostakos V (eds) Proceedings of the 2019 CHI conference on human factors in computing systems, CHI 2019, Glasgow, Scotland, UK, May 04–09, 2019. ACM, p 537
- Suárez-Warden F, Rodríguez M, Hendrichs N, García-Lumbreras S, Mendivil EG (2015) Small sample size for test of training time by augmented reality: an aeronautical case. *Procedia Comput Sci* 75:17–27
- The Best VR Headsets for PC (2022) <https://thewirecutter.com/reviews/best-desktop-virtual-reality-headset/>. Retrieved 28 Feb 2022
- Tran TQ, Tran TDN, Nguyen TD, Regenbrecht H, Tran M (2018) Can we perceive changes in our moving speed: a comparison between directly and indirectly powering the locomotion in virtual environments. In: Proceedings of ACM symposium on virtual reality software and technology (VRST), pp 36:1–36:10
- Venkatesh V (2000) Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Inf Syst Res* 11:342–365
- Wilcoxon F (1992) Individual comparisons by ranking methods. In: Breakthroughs in statistics, pp 196–202
- Wißmann N, Miśiak M, Fuhrmann A, Latoschik ME (2020) A low-cost approach to fish tank virtual reality with semi-automatic calibration support. In: 2020 IEEE conference on virtual reality and 3d user interfaces abstracts and workshops (VRW). IEEE, pp 598–599
- Yang L, Huang J, Feng T, Hong-An W, Guo-Zhong D (2019) Gesture interaction in virtual reality. *Virtual Real Intell Hardw* 1:84–112
- Zhu K, Perrault ST, Chen T, Cai S, Peiris RL (2019) A sense of ice and fire: Exploring thermal feedback with multiple thermoelectric-cooling elements on a smart ring. *Int J Hum Comput Stud* 130:234–247

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.