



Development and evaluation of a collaborative virtual reality system for tour guide training

Downloaded from: <https://research.chalmers.se>, 2025-09-25 06:39 UTC

Citation for the original published paper (version of record):

Le, K., Ly, D., La, T. et al (2025). Development and evaluation of a collaborative virtual reality system for tour guide training. *Virtual Reality*, 29(3). <http://dx.doi.org/10.1007/s10055-025-01206-0>

N.B. When citing this work, cite the original published paper.



Development and evaluation of a collaborative virtual reality system for tour guide training

Khanh-Duy Le¹ · Duy-Nam Ly¹ · Thanh-Thai La¹ · Cuong Nguyen² · Morten Fjeld^{3,4} · Tam V. Nguyen⁵ · Minh-Triet Tran¹

Received: 29 March 2024 / Accepted: 15 July 2025
© The Author(s) 2025

Abstract

Tour guiding plays a key role in turning sightseeing into memorable experiences. Tour guides, especially inexperienced ones, must practice extensively to reach proficiency. Skill-sets typically include knowledge about sights, in-situ presentation, and a convincing ability to interact with and engage tourists. These skills require on-site training with live tourists. However, with modest budgets, such setups may be out of reach and trainees have to compromise with off-site or simulated classroom setups. To tackle this problem, we present the development and evaluation of VRGuideMaster; a VR system enabling its users to practice tour guiding with 360-degree travel videos. With VRGuideMaster, a tour-guide trainee equipped with a HMD can rehearse presenting while immersed in a virtual environment constructed from 360-degree tourist site videos. Additionally, in the virtual environment, the trainee can communicate and interact with video streams of tourists, who remotely join in through their mobile and personal devices. The result of a user study ($n = 12$) comparing VRGuideMaster with a baseline emulating the practice of showing 360-degree videos via conventional video conferencing interfaces, shows that overall, VRGuideMaster was more effective in supporting tour-guide trainees in practicing their tour-guiding skills.

Keywords Virtual reality · Tour guide training · 360 VR · VR collaboration

1 Introduction

Tour guiding is an essential part in the tourist industry. Tour guides need to keep track of the physical surroundings while monitoring tourist behavior. At the same time, they must adapt their communication style to keep the tour interesting. Such skills require extensive prior training. In addition, as there might be changes in the tour site (e.g., new landmarks,

installations, restaurants, stores, etc.), tour guides also need to update their knowledge regularly. Therefore, giving tours needs to be frequently practiced by tour-guides at all levels of experience.

In the training of tour guides, trainees typically learn to conduct field trips to the place of interest in the presence of target tourists (Fig. 1). The goal given is to familiarize themselves with the actual environment of the tourist site

✉ Khanh-Duy Le
lkduy@fit.hcmus.edu.vn

✉ Morten Fjeld
morten.fjeld@uib.no

Duy-Nam Ly
ldnam@selab.hcmus.edu.vn

Thanh-Thai La
ltthai18@clc.fitus.edu.vn

Cuong Nguyen
cunguyen@adobe.com

Tam V. Nguyen
tamnguyen@udayton.edu

Minh-Triet Tran
tmtriet@hcmus.edu.vn

¹ VNUHCM - University of Science, 227 Nguyen Van Cu street, Ho Chi Minh City 70000, Vietnam

² Adobe Research, San Francisco, USA

³ t2i Lab, University of Bergen, Fosswinkels gate 6, Bergen, Norway

⁴ t2i Lab, Chalmers University of Technology, Lindholmsplatsen 1, Gothenburg, Sweden

⁵ Department of Computer Science, University of Dayton, 300 College Park, Dayton, OH 45469, USA

Fig. 1 An illustration of current practice of field-trip based tour guide training. A trainee practices presenting about the tourist site combining multiple communication streams (voice, gaze, hand pointing gestures) to interact with tourists. The trainee uses a paper-based itinerary (colored in blue for highlighting purpose) to keep track of the places along the trip



while practicing their observation and improvisation skills. However, such a setup can be extremely costly, especially in overseas destinations. It can be difficult for smaller tour agencies to organize such training on a frequent basis. This is also problematic for trainees, who often have a tight budget for such field trips. The lack of training could lead to a lack of experience and preparedness of tourism professionals. To address this problem, our work contributes a solution that allows tour-guide trainees to effectively: (a) rehearse their presentation in front of realistic representations of tourist sites, and (b) improve their ability to communicate and interact with tourists in a lively manner. Hence, one contribution of our work is a novel collaborative virtual reality (VR) based training system, VRGuideMaster, enabling trainees to practice without having to take costly onsite field trips. A second contribution of our work is a set of empirical findings from a comparative evaluation of the concept underlying VRGuideMaster.

The prevalence of 360-degree videos (hereafter: 360 videos) and the recent rapid development of VR opened the opportunity to assist in this respect. Travel 360 footage of popular destinations around the world can now easily be found on video-sharing platforms like YouTube. They are rich sources not only for online traveling but also for tour guide professionals to quickly acquire up-to-date knowledge about places new to them. The immersion of viewing 360 videos in VR has been shown to be effective for training, especially in public speaking skills (Stupar-Rutenfrans et al. 2017; Flobak et al. 2019). Leveraging these technologies, we present VRGuideMaster, a VR system that allows users to practice giving tours without having to travel to the actual physical destination. Conceptually, VRGuideMaster allows trainees to view 360 videos of the site while presenting it to

remote tourists who can join in the training session using their personal computers, laptops or mobile devices like tablets. The trainees can thus effectively familiarize themselves with the place through the real omni-directional images presented in the 360 video while honing their ability to maintain their awareness of the tourists and adapt accordingly. This paper is built on our previous work (Ly et al. 2022), which reports design requirements elicited from three need-finding interviews, the concept and a pilot study with an early prototype of VRGuideMaster. However, it is still unclear how effective VRGuideMaster is, as compared to an existing practice in practicing tour-guiding skills. In this paper, we present the complete design and development of the fully functional prototype of VRGuideMaster based on insights from a previous pilot study (Ly et al. 2022) and an empirical study of the prototype. To evaluate the effectiveness of the proposed system, we conducted a user study with twelve students ($n=12$) majoring in tourism where we compared VRGuideMaster with a baseline. The baseline system adopts a conventional video conferencing interface, where users could present a 360 video to remote viewers, much like using Zoom or Google Meet. The baseline condition was designed to emulate current practices in virtual tour guide training. The results of the user study show that when using VRGuideMaster, tour-guide trainees reported better spatial perception of the tourist site. Using VRGuideMaster, it was also easier for tour-guide trainees to non-verbally refer to and be aware of virtual tourists' references to objects and landmarks in the place. Participants also reported feeling more confident to give real-world tours. We conclude with potential directions to improve 360 video systems for more effective tour-guiding training in the future.

2 Related work

VRGuideMaster extends previous training systems that leverage VR or 360 videos by trying to overcome their limitations. Its design was also based on previous research on interactions with 360 videos as well as VR collaboration.

2.1 Virtual reality for tourism

360 videos have been widely adopted by the tourism industry for promotion (Rahimizhian et al. 2020; Wu and Lai 2022; Prasetyo and Hati 2022) and virtual travel purposes (Rifai et al. 2021) especially during and after the most virulent stages of the COVID-19 pandemic (Yang et al. 2021). Wu and Lai (2022) reported that viewing 360 videos of a mountain resulted in the positive influence on the sense of immersion, emotional involvement and enjoyment of viewers, leading to their higher intention to take a walking tour at the mountain. Willems et al. (2019) also found that compared to traditional media, 360 VR generated higher senses of telepresence and engagement which in turn positively affected their enjoyment. Škola et al. (2020) explored storytelling-based VR systems supporting virtual tours to cultural heritage sites. The systems combine 360 videos of the heritage sites with 3D graphical elements or pre-captured images of actors to guide the viewers through the sites in a predefined flow and itinerary. The COVID-19 pandemic restricted physical traveling of many people, which led to a new form of touristic traveling called distant local-guided tours. In a distant local-guided tour, one or multiple tourists can remotely travel to a tourist site through their personal computers where they can view the landscapes of the site accompanied with the presentation and personal communication of a local tour guide (Seyitoğlu and Atsız 2022). Supporting this new concept of touristic travel, Nassani et al. (2021) developed a system which equips a local guide with a mobile interface coupled with a 360 camera to live stream the panoramic video of the site and communicate with remote tourists. However, the interface could not allow the tour guide to perceive tourists as if they were in the same spatial environment, and thus also prevented the guide's ability to monitor and respond to their facial expressions, postures, and gestures. Overall, these works focused on supporting tourists to remotely experience a touristic site, but they were not suitable for tour guide trainees to practice their tour-giving skills.

2.2 Training in VR

360-degree videos have been extensively employed in VR systems for training operational skills and decision making

in, for example, sports (Hebbel-Seeger 2017; Panchuk et al. 2018; Kittel et al. 2020), learning foreign languages (Repetto et al. 2021), medical education and training (Izard et al. 2017; Patel et al. 2020; Chang et al. 2022), and academic pedagogy (Ferdig and Kosko 2020; Walshe and Driver 2019; Ibrahim-Didi 2015). VR systems presenting audiences in pre-recorded 360 videos were also shown to be an effective training tool for reducing anxiety in public speaking (Stupar-Rutenfrans et al. 2017; Flobak et al. 2019). Tseng et al. (2013) developed a web-based VR environment where users can view panoramic pictures of a tourist site while practicing their tour guiding presentation skill. Li and Hu Na and Weihua (2012) proposed a web-based interface consisting of a panoramic image of a touristic site accompanied with textual guidance and self-made commentaries to support users to practice presenting about it. However, in these systems, trainees can only view the pre-recorded videos. They cannot interact with the content or perceive behavioral responses from tourists, and are therefore unable to develop their adaptive presenting skills.

Instead of using 360 videos, other VR systems utilize 3D models to offer interactive training environments. Pittarello et al. (2020) developed a VR environment where actors can interact with 3D models of stage objects and characters while practicing. 3D-based VR has also been shown to be effective for public speaking training (Palmas et al. 2019; Takac et al. 2019). Wang and Wang (2019) developed a VR system allowing users to freely navigate in the 3D environment of a park to practice giving tours, although, as with other such models, there was no involvement of tourists or audiences in this environment to allow them to realistically practice response and improvisation. In general, compared to 360 videos, although 3D contents provide more navigation flexibility in VR, they have lower presentation realism. They also require much higher cost for production and maintenance. Eiris et al. (2020) showed in a use case of hazardous identification training in construction environments that students perceived a 360-degree panorama of the training site to be more realistic than 3D models. Investigating VR viewers' sense of presence between handcrafted 3D and panorama based virtual environments, Schäfer et al. (2021) concluded that panoramic media can be used as a low-cost alternative for 3D models in cases where the users do not need the full freedom to navigate in the virtual environment. In guided tours, tour guides and tour tourists often follow a predefined trip itinerary. Therefore, we opted to choose 360 videos as the medium to represent a place in VRGuideMaster as they provide trainees and tourists with realistic appearances of the site at a low cost while not necessarily allowing them to fully freely explore it.

2.3 Collaborative interaction in VR

VR opened the opportunity for users to jointly view and interact with virtual contents either in collocated or remote settings. Maintaining social awareness among collaborators is essential in collaborative systems. A Facebook 360 demonstration illustrated the use of simplified 3D avatars representing collaborators co-viewing a 360 video Facebook Social VR Demo¹. However, simplified avatars cannot easily express the facial expressions of collaborators adequately. Leveraging textured 3D meshes of body scan (Li et al. 2015) can realistically convey collaborators' expressions but this requires specialized capture hardware like a depth-camera. CollaVR (Nguyen et al. 2017) is an in-headset platform for collaboratively viewing and annotating 360 videos. To maintain spatially mutual awareness among collaborators, CollaVR highlights the collaborators' areas of interest on 360 videos to indicate their current attention and maintain awareness among them. Tourgether360 (Kumar et al. 2022) provides an interface which incorporates a timeline encapsulating both spatial and temporal information of the 360 video and viewers concurrently watching the video. Such a timeline visualization allows the viewers to maintain their mutual awareness on where they are on the video timeline and what they are seeing. 360-video VR collaboration was also explored in heterogeneous device settings. Henrikson et al. (2016) support collocated collaboration between a VR user and a user using a mobile tablet in creating and previewing storyboards of 360 videos. TransceiVR (Thoravi Kumaravel et al. 2020) also supports collaboration between a VR user and a mobile user in working on 3D virtual contents. Similarly, WebTransceiVR (Lyu et al. 2022) is a web-based collaborative platform which allows distributed users on heterogeneous devices to remotely watch shared VR content from different perspectives. In line with WebTransceiVR, WebXR¹ is a web-based commercial product allowing users to create shared VR environments based on 3D assets where multiple users can join from different

devices, converse and interact with the virtual environment. In these environments, participants are also represented as stylized or abstract 3D avatars. In VRGuideMaster, we enable collaborative communication between the tour guide in VR and a number of tourists through their mobile devices or personal computers. For training realism, we make use of the webcam on mobile devices, or on personal computers, to conveniently involve users into training sessions as virtual tourists.

3 Design and implementation of VRGuideMaster

Conceptually, VRGuideMaster involves two types of users. The primary users are tour-guide trainees who need to utilize the system to improve their professional skills such as presenting about tourist sight, communicating and interacting with tourists. The secondary users are those who remotely join in as tourists who the tour-guide trainees can interact with to practice their skills. To support these two types of users, VRGuideMaster provides them with two respective interfaces specifically designed to accommodate their necessary activities. Fig. 2 illustrates the overall concept of VRGuideMaster. In this section, we describe the design and implementation of these two interfaces.

3.1 Tour guide trainee interface

Fundamentally, the tour guide trainee interface is a VR environment where a user can immerse him/herself in the landscape of a tour site and give presentation about it to tourists, who are virtually co-present in the same space. The tour guide trainee interface primarily consists of a spherical surface displaying tour sites' 360 videos, the representations of tourists and a control panel for the trainee to play back the videos in different ways. The interface also enables the tour guide trainee to monitor places along the trip through

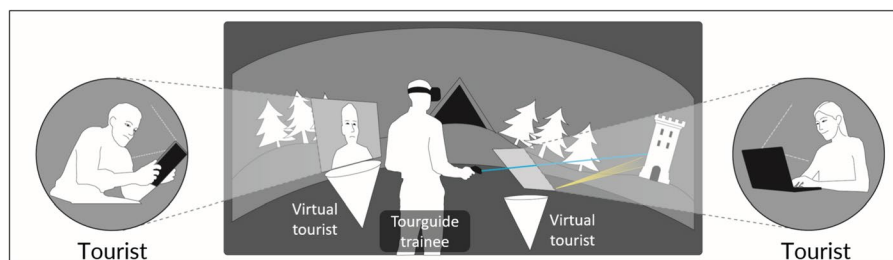


Fig. 2 Concept of VRGuideMaster: a tour guide trainee (center) can view and practice giving tours with 360 travel videos presented in VR. Remote users (left and right) can join a training session as tourists, through their mobile device (left) or their personal computer (right).

¹ <https://immersiveweb.dev/>

Combining built-in front-facing cameras and virtual tourist VR avatars, such devices can capture and stream virtual tourist facial expressions back to the trainee, thereby offering instant visual feedback

a graphical itinerary, triggered by a widget on the trainee's hand in the virtual environment. Moreover, the interface also integrates visualizations to convey non-verbal communication cues of the trainee as well as multiple interaction techniques to help the trainees maintain his/her awareness on the tourists' interactions. The different components of the tour guide trainee interface were designed and implemented as follows:

3.1.1 Tour sites representation

To provide tour-guide trainees with a realistic representation of tour sites, we utilize 360 videos of the tour sites, which can be easily found on several online video platforms or captured by off-the-shelf 360 cameras. The 360 videos were rendered on a 3D spherical skybox with a radius of 30 m (Fig. 3). The tour guide was positioned at the center of the sphere and not supported to move around the environment. Rather, they could see and easily interact with the 360 videos, e.g., through distal pointing, from the center of the skybox.

3.1.2 Visualization of remote tourists

To help trainee users practice their ability to observe and communicate with tourists, VRGuideMaster enables other users to remotely participate in a training session as tourists using their mobile devices or personal computers. Video streams of the tourists' facial expression captured by the front-facing camera built into their devices will be displayed in the VR environment, enabling the trainee user to perceive tourists' facial expressions and behaviors while practicing guiding (Fig. 2). We chose to leverage the use of mobile devices or personal computers for these users for two reasons. First, mobile devices and personal computers are widely prevalent, allowing anyone to act a tourist for

the trainee user anywhere at anytime. Second, the built-in facing camera of those devices can capture real images of the tourist, which can allow the trainee user to realistically perceive their facial expressions and behaviors.

The video stream of each virtual tourist was rendered on a vertical 100 cm × 100 cm rectangular plane with a 3D cone underneath representing its body. The avatars of the audiences could move around the VR environment according to their navigation on their mobile or personal device. The audience's video stream could perform 6-DoF rotation depending on their current view, visually indicating their current attention. To avoid the avatars appearing as floating inside the skybox, we created a ground plane for the VR environment by mapping the bottom spherical half of the 360 video onto a horizontal plane placed at the middle of the skybox. Thereby, the audiences' avatars would cast shadows, thus mitigating the floating-in-the-air effect.

3.1.3 Pointing and marking

As deictic gestures (e.g., pointing) are essential in presentation, especially in giving tours, VRGuideMaster provides a tool to support trainee users to convey such communication cues through interactions with touch controllers. When the trainee user presents, he/she can use a handheld touch controller to point onto the 360 video, emitting a ray from the virtual representation of the controller towards the target. The trainee user can press the trigger on the controller to pin a holo marker at a location on the 360 video to attract the audiences' attention to an object (e.g., a house or a sculpture).

In 360 voyage videos, objects in a scene usually move as the videos play due to the camera's movement in the capturing process. Thus, if a holo marker placed on an object does not spatially adapt to this, it can cause a user's confusion due to the misalignment between the actual region

Fig. 3 The VR interface of tour guide trainees where 360 videos of tour sites are displayed on a spherical skybox combined with a ground plane constructed from the bottom half of the video. Tourists remotely joining in the training sessions are represented as 3D avatars with their video streams as the heads of the avatars showing the tourists' realistic facial expressions



Fig. 4 Marking feature with automatically tracking marker: **a** tour guide aims pointing ray at a target region; **b** a holo marker is initially placed at the target; **c** the holo marker continues attaching to the target as the video plays

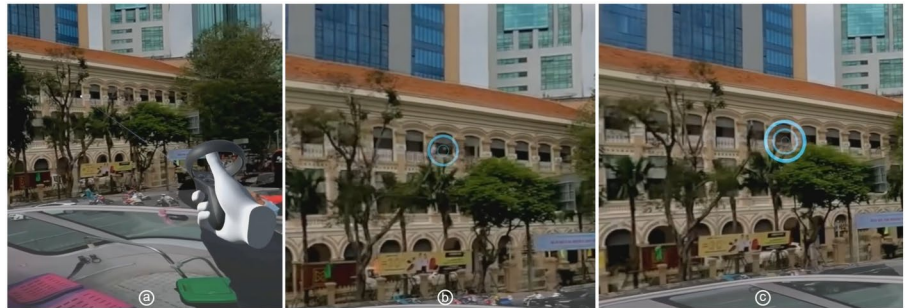


Fig. 5 **a** Two user interface widgets attached to two hands of the tour guide's avatar to launch the video control panel (left hand) and trip itinerary (right hand); **b** video control panel where the tour guide can play, pause, fast forward and fast review the current 360 video. When

the video is playing, the label on the top of the panel displays the next destination in the itinerary; **c** the representation of the itinerary displayed on the right side of the tour guide indicating the past destinations and the next ones

of interest and the holo marker's position. To address this, we made holo markers automatically stick to the corresponding region of interest when the video played. More specifically, we employed optical flow to track the region of interest across the video frames. Now, when the pointing ray of the tour guide reached a point on the video, the system would extract key points in a circular region with a radius of 100 pixels centered at the pointed position. Then the system would extract visual features from these key points, to be used by the optical flow tracking algorithm to estimate the location of the object of interest in the following frames in order to visualize the holo marker (Fig. 4b, c). The tracking would stop when the ratio of the number of key points remaining in a frame over the total number of initial key points was lower than a certain threshold; or the holo marker had lasted for more than five seconds. In our system, after several testings during the prototyping phase, we set this threshold to be five percent as this provided quite reliable tracking outcomes.

3.1.4 Video control and navigation

VRGuideMaster provides a control panel consisting of three buttons: one for playing and pausing the video; the other two for temporally navigating the video backward or forward (Fig. 5b). The trainee user can pause the video to present about particular places or things at that place in the video. This resembles tour guiding in reality when the guide gathers tourists in front of a certain place to give more details about it. The forward feature supports the trainee user to quickly skip certain parts of the videos where there

are no interesting things to talk about. Each time the trainee user presses the button to trigger this feature, the video will forward to the next 5 s. The backward feature enables the trainee to review any point of the tour to rehearse or improve it. We added this feature because VRGuideMaster is a platform for practicing, not a virtual guiding platform, thus the users should be able to repeatedly rehearse for higher perfection.

Besides that, the system can also support users to fast-forward to the next destination in the trip itinerary rather than in five-second increments. We thus refined VRGuideMaster to provide the users with another video fast-forwarding technique. In this technique, the user can perform a long press on the fast-forward button, which will trigger a tool tip to pop up on the video timeline, showing the name of the next place in the itinerary. Once the user selects the tooltip, the video will automatically jump to the timestamp of the place. Besides that, we also refined the video control panel interface by showing the next destination in the itinerary when a video is playing (Fig. 5c).

3.1.5 Tour guide representation

The system uses a 3D avatar to represent the tour guide trainee in VR. We chose to use animated 3D avatars instead of 2D images or 3D construction of the tour guide, despite the higher representation fidelity offered by the latter, as we would like to keep the hardware setup of the system simple and compact, easily acquired and deployable in different contexts. Additionally, when giving tours in VR, the tour guide's face will be occluded by the headset, posing a

challenge for capturing and representing it realistically in the VR environment. Therefore, using animated 3D avatars is a more appropriate alternative.

We provided two 3D avatar alternatives (one male and one female) selected from the Oculus Meta Avatars package (Fig. 6). The tour guide trainee can choose one of the 3D avatars to represent him/her in the VR environment depending on their preferences before starting a training session. The avatars had arms and hands but not legs. Inverse kinematics were supported in these avatars, enabling the mapping of the tour guide trainee's gestures, conveyed through interactions with the handheld touch controllers corresponding with arm and hand movements. The avatars were also configured to convey certain simulated facial expressions based on the tour guide trainee's interaction. The avatars can automatically perform lip movements depending on the audio signals obtained from the user's VR headset. More advanced, this feature can recognize laughter when the user is laughing and represent it on the avatar accordingly. In the VR environment, the tour guide trainee will view things from the first-person perspective of the avatar, thus can see the avatar's body, arms, and hands but not its face.

3.1.6 Trip itinerary

The main use of a trip itinerary is to allow a tour guide to keep track of the list of places of interest along the trip, to better prepare his/her presentation. On a paper-based itinerary, the guide often needs to mark off the places as they have been passed in order to keep track of their progress. To make this monitoring task more efficient, we implemented a dynamic trip itinerary in the VR interface. This appeared as a vertical list of the places to be visited along one side of the tour guide. The list only displayed up to five places in the trip including most recent, the current and the next three following places. From several informal tests, we found that this number should be sufficient for the tour guide to prepare the presentation far enough ahead while not too lengthy to cause a cluttering and overwhelming visualization. Each place in the list was associated to its corresponding starting video frame, so when the video had played over a place, the list would scroll up, putting the next point of interest in the center. The interface also changed font size and color

luminosity to distinguish between past and coming places and further emphasize the next one (Fig. 5c). We also prototyped and tested other design alternatives for the trip itinerary, such as a horizontal strip placed at the user's eye level or at the chest level, but found that for accommodating a similar number of places, a horizontal strip representation could lead to excessive head movement efforts of the guide to view all the text displayed. Through further testing, we chose the scrollable vertical list representation as most optimal for allowing the guide to effectively monitor the progress of the trip with little head navigational effort.

In a real tour, a tour guide often holds the trip itinerary as a piece of paper in his/her hand to glance at it when needed. We adopted this metaphor to design the interaction for tour guides to open the trip itinerary feature in the VR interface. In the VR interface, the trainee will see a graphic widget in the shape of a sticky note, analogous to the physical itinerary paper, on the right hand of his/her 3D avatar. When the guide wants to open the list of places, he/she performs a grab gesture by pressing all the buttons on the touch controller being held by the corresponding real hand. Similarly, when the list is already displaying, the tour guide can close it by performing a grab gesture.

In the current prototype, the list of places in a 360 video's itinerary need to be manually prepared in advance using a text editor. An itinerary file consists of multiple lines of text, corresponding to the places in the itinerary. Each line consists of the place's name, the starting time and the ending time of the place's appearance in the video. The itinerary interface will utilize these timestamps to represent of the list of places accordingly.

3.1.7 Supporting spatial awareness of verbal communication

When tour guides conduct tours in the real world, sometimes they need to rely on spatial information through their speech to efficiently coordinate the communication. For example, when a tourist standing outside the field of view of the tour guide asks a question, the guide will first use the voice direction to determine the tourist's location, and then vision to more precisely locate the speaker. To support this communication mechanism in the virtual world,

Fig. 6 Two avatar alternatives (one male (left) and one female (right)) representing tour guides in the virtual environment as seen on the tourist's interface

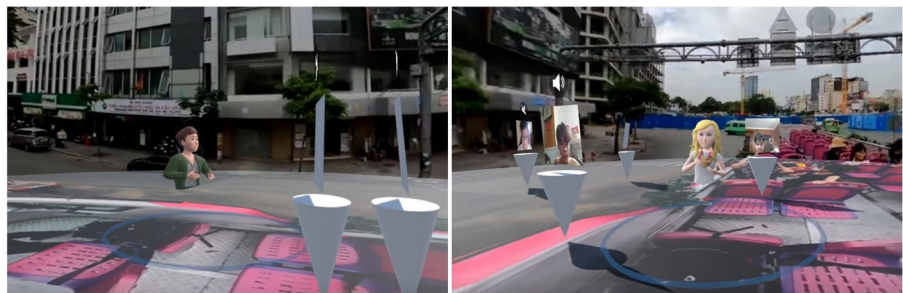


Fig. 7 User interface for tourists to refer to a region of interest using pointing gesture: pointing highlighter on the tourist's interface (left) and cone-shaped pointing beam aiming at the target as shown on the tour guide's VR interface (right)



Fig. 8 User interface components to help tourists maintain their awareness of the tour guide's communication and interaction: picture-in-picture preview window showing the tour guide's current view on the 360 video (left) and animated arrow pointing towards the region on the 360 video where the tour guide is pointing (right)



VRGuideMaster enables spatial audio rendering of tourists' verbal communication. To achieve this, every tourist 3D avatar will be equipped with an audio source component, which will render the audio signal coming together with the video stream of the corresponding tourist. This feature should constitute a more realistic virtual environment to help tour guide trainees.

3.2 Tourist interface

Aside from the VR interface for tour guide trainees, we also developed an interface running either on personal computers or mobile devices like tablets, smartphones whereby a tourist can remotely join in the guide's virtual environment. The interface allows the tourist to effectively view tour sites, follow and interact with the tour guide. When the tourist navigates in the virtual environment to view different regions in the 360-degree video, his/her video-based avatar in the VR interface also changes its direction accordingly to reflect their current view.

Displaying the tour site environment: To help tourists view the tour site, the same spherical 360-video skybox as constructed for the tour guide trainee interface was also used on the tourist interface. Tourists can also freely navigate (e.g. move and turn) on the ground plane to view the 360-degree video from different perspectives on their computer or tablet. On this interface, tourists can also view the animated avatar of the tour guide.

Support for communication and interaction with tour guides: The tourist interface also enables tourists to communicate with the tour guide through video and audio sharing as well as non-verbal communication cues. When a tourist wants to express his/her attention on an object in the 360 video, aside from verbally asking or describing about it, he/she can point directly on the interface (e.g,

using mouse-click interaction on desktop or laptop or touch interaction on tablet) (Fig. 7 (left)). Concurrently, the VR interface also displays a cone-shape beam emitting from the corresponding avatar of the tourist towards the object or position of interest on the video to inform the tour guide and other tourists about the pointing gesture (Fig.7 (right)).

While a tourist can freely navigate around the skybox to explore the sight shown on the 360 video, he/she should be aware of what the tour guide is looking at and be able to quickly turn to it when needed. Compared to the tour guide's field of view via a VR headset, the screen of desktops, laptops or tablets provides the tourist with a much smaller field of view of the virtual environment, making several activities of the tour guide easily appear out of the screen of the tourist's device. To support this, we adopted the Outside-In interface proposed by Lin et al. (2017), which leverages spatial picture-in-picture previews to help VR users maintain their awareness on out-of-sight regions of interest. In particular, when the system detects no overlap between the tourist's view on the video and the one of the tour guide, the preview window showing the current view of the tour guide will be displayed on the nearest side of the tourist's view frustum (Fig. 8 (left)). If the tourist wants to quickly see what the tour guide is viewing, he/she can press on the preview window, which will trigger a rotation of the scene to the tour guide's current view. This feature is necessary to avoid tedious interactions in case the tourist has to manually navigate the video using mouse or touch panning gestures.

Similarly, the tourist can also easily lose his/her sight of the tour guide's pointing on the video due to the limited screen size of his/her device. To mitigate this, VRGuideMaster provides a similar feature as the aforementioned preview window to help the tourist stay informed about the tour guide's pointing. More specifically, when the pointed target is outside the tourist's field of view, the tourist interface will

display an animated arrow at the side nearest to the target to attract the tourist's attention (Fig. 8 (right)). The tourist can also press on this arrow to rapidly navigate to the target being pointed out by the tour guide.

3.3 Implementation

Both the tour-guide and tourist interface were developed in Unity using C#. Oculus development SDK was integrated into the development of the tour-guide interface to make it compatible with the Oculus Rift S headset we used to run this prototype. The system uses Agora Unity SDK Real-time Voice and Chat Inside Your Game for video and audio streaming among tour-guide and tourist interfaces. Besides that, the Photon Unity Networking library The Ease-of-use of Unity's Networking was used to synchronize states and perspectives of the virtual environment on the tour guide and tourist interfaces. We utilized the OpenCV for Unity library OpenCV for Unity in implementing the optical flow tracking of the region-of-interest marking feature. For rendering spatial audio on the Oculus Rift S VR headset, we integrated the Oculus Spatializer Unity package Oculus Spatializer Features into the tour-guide interface. Finally, we leveraged the Meta Avatars SDK for Unity Overview of Meta Avatars SDK to create 3D avatars of tour guide with vivid hand gestures and lip movements.

4 User study

We conducted a user study to evaluate the effectiveness of VRGuideMaster in supporting tour-guide trainees in practicing their skill of presenting about a tourist site while maintaining their awareness on the tourists' behavior and interacting with them. We compared VRGuideMaster with a baseline interface which resembled the practice in which a tour-guiding trainee used a common video conferencing interface, such as Zoom or Google Meet to share his/her screen showing a 360-degree video while presenting about a touristic place to remote tourists, acting as tourists. We chose this setting as the baseline based on the insights that we learned from the need-finding interviews. The lecturers and trainees reported that giving presentations to the class via Zoom or Google Meet was a common choice for

practicing or taking examinations in tour guiding, especially when co-located gathering could not happen. In addition to that, if those presentations made use of 360 images to provide an extended view of the place, they were often highly appreciated by the tourists (i.e. classmates and trainers pretending to be tourists). We would like to examine whether compared to this conventional setting, VRGuideMaster would be perceived to be more effective by tour-guide trainees in helping them practice their tour-giving skills.

4.1 Baseline condition

We compared VRGuideMaster to the training setup where a tour-guide trainee and tourists participated in a video conferencing session, using common tools, such as Zoom or Google Meet, and the tour-guide sharing his/her screen showing the 360-degree video of the touristic side while presenting about it. Unfortunately, when playing 360-degree videos, Zoom and Google Meet only allow users to record and share screens at a rate of 10–15 frames per second. This has an impact on the viewer's experience. As a result, to ensure a fair comparison in our study, we reused VRGuideMaster's architecture to create a new baseline interface mimicking the conventional window-based layout of those existing video conferencing tools with the following core features:

Communication via video and audio: a tour-guide trainee and tourists can talk to and see video streams showing facial expressions of the others. The video streams are shown in rectangular windows displayed as a vertical list, which is common layout on Zoom or Google Meet in screen-sharing mode when the shared screen accommodates a major space on the screen (Fig. 9). To avoid obstructing the view of users, we also allowed them to minimize the list of faces or move it to anywhere on their screen. Moreover, each user could turn on or off his/her video or voice when necessary.

Co-viewing a 360-degree video: simulating the tour-guide trainee sharing his/her screen showing a 360-degree video to tourists, the baseline interface provides a feature which allows the tour-guide trainee and the tourists to remotely view a 360 video together on their devices (Fig. 9). Similar to screen-sharing in conventional video conferencing where only the sharer can interact with the video, only

Fig. 9 Interfaces used in the baseline condition: interface for tour guides where they can play, pause and navigate 360 videos as well as observe tourists' video streams (left); and interface for tourists which shows the 360 video being shared by the tour guide as well as the video streams of tourists and the tour guide (right)



the tour-guide trainee can navigate the 360 video temporally and spatially while the tourists will follow the current view of the tour-guide trainee. Compared to screen sharing in Zoom or Google Meet, this baseline system offers a higher streaming frame rate (25–30 fps) as we utilized an in-house streaming server. This is comparable to the frame rate of the VRGuideMaster prototype in this study to ensure comparable conditions in this respect.

In this condition, we also provided users with a paper itinerary note to help them prepare their presentation. This is a common practice in conventional tour-guiding either in training or in a real tour, where tour guides often glance at the list of places on the paper to know what the next places are and prepare themselves with necessary information to talk about. The itinerary note showed a vertical list of places but not their corresponding timestamps in the video to simulate actual itinerary notes in the live tour.

4.2 Task

Each participant in the user study was tasked to perform virtual tour guiding with a group of five tourists using VRGuideMaster and the baseline interface. In each condition, the participant would present with a 360 video consisting of multiple locations in a city. To maintain similar task difficulties between the two conditions, we used two cuts from an original 360 video of the local city where the research team was located. Each video cut was approximately eight minutes and covered five tour locations. Each video was accompanied with a script providing details of the touristic locations appearing in it. The scripts were prepared in advance by the research team based on information of the locations that the team had collected and composed from the internet. The scripts were sent to the participants via email one week before their respective experiment date. The participants just needed to remember information of the locations but not word by word as provided by the scripts. We also encouraged the participants to personalize their presentation with personal details they had about the locations to resemble realistic tour-guiding practices.

In each trial, a participant would guide a group of five tourists. To foster interactivity, we requested each tourist to ask three questions in each presentation. As the trainee might present about the locations slightly differently due to their ways of personalizing the tour, we did not constrain the tourists on when, where and what they should ask about, as long as they ensured their total of three questions as required. The tourists were volunteers we recruited through emails and personal connections. In total, we recruited seven volunteers through email broadcast and snowball sampling. Three among those participated with all the test subjects, while the other four could not due to their availability. Two

of these took part in the user studies with the first five test subjects while the others were involved in the studies with the last seven test subjects. The volunteers were 3rd and 4th-year university students (age: 20–21) consisting of five males and two females with no experience in the tourism industry.

As we had two interface conditions and two 360 videos in total, the combinations of the videos and the interface conditions were counter-balanced using a Latin-Square approach to mitigate learning and order effects.

4.3 Participants

We recruited twelve trainees (four males and eight females) majoring in tour guiding or tourism management and in their senior or final year of their under-graduate study. Their ages were between 21 and 22 years old. They possessed basic to moderate computer literacy. Only three of them had tried VR a few times before. As for 360 videos, all participants rarely watched or used them for their professional practices. Among the participants, eight had taken part in practical tours. Among those, five had a lot of experience in guiding tours. The others were less experienced but had already taken basic courses on tour guide skills during their undergraduate study.

4.4 Design and procedure

We performed a within-subject user study where each participant needed to perform a tour-guiding task with both the baseline and VRGuideMaster condition. After confirming their participation in the user study (via email or messaging platforms), they received the videos used in the user study together with their corresponding scripts one week before the participation date. The videos consisted of two relatively long ones to be used in experimental tasks and a shorter one for familiarization with the interface conditions. The participants were informed in the emails that the videos would be used as tour guiding materials in the experiment. We also instructed them that they could utilize basic information in the scripts to present about the places in the video but did not have to strictly follow it. He/she was encouraged to personalize their information presentation to make it engaging.

When a participant arrived at the experimentation place, he/she was greeted, then completed a consent form and a questionnaire about age, occupation, experience with tour guiding, mobile devices and VR technologies. Following this the participant was introduced to the first trial. Next, he/she was shown the first interface to be used and given ten minutes to become familiar with it. For the familiarization, the participant performed a mock-up tour guide using the short video with the attendance of the tourists. After the

familiarization, he/she started on the first trial. Tourists could pose questions during this presentation. After finishing the first task, the participant answered a questionnaire consisting of eight Likert-scale questions about different usability and experience aspects of the interface condition. There was a 10-minute break before starting the second trial, using the second interface condition, which also began with the same familiarization video as used in the first trial, and after performing the task with a second video of tour sites the participant answered the same questionnaire. Finally, a post-study interview was conducted to gather the participant's opinions on the two interface conditions.

4.5 Apparatus

The user study was conducted in a laboratory room ($W = 6 \text{ m} \times L = 8 \text{ m} \times H = 3 \text{ m}$) on the university campus. Inside the room, an area of $W = 2 \text{ m} \times L = 3 \text{ m}$ was reserved for participants to stand and move around in VR if needed in the VRGuideMaster condition. We used a Windows PC equipped with a Core i7 processor, 32-GB RAM and a NVIDIA Geforce 1080 graphic card for running the systems for tour-guide trainees (user study participants), both in the VRGuideMaster and the baseline condition. In the baseline condition, the participants were seated on a chair with a 45-cm height and the interface was shown on a 27-inch display placed on a table with an 80-cm height. The PC was wired to the local network system. The recorded internet speed was around 100 megabits per second. Participants viewed the VRGuideMaster interface on an Oculus Rift S headset which was wired to the PC used in the experiment.

We used an in-house server to store the 360-degree videos used in the study for streaming to the PCs of the participants and the devices of volunteer tourists in order to assure a reliable network performance. The volunteer tourists all used laptops to run the tourist's system even though we did not constrain them to only use these, they could also have chosen PCs or tablets. All the volunteer tourists reported that their laptops had at least a Core-i3 processor and 4GB RAM, and that their laptops were wirelessly connected to the internet with speeds around 100 megabits per second.

4.6 Measures of effectiveness

To measure the effectiveness of the interfaces for practicing tour guiding, we collected participants' ratings of eight statements using a Likert-scale. This aimed at investigating how participants perceived the effects of the interfaces on different aspects of their perception and experience when practicing their tour guiding skills. The eight measures of effectiveness targeted:

1. Perceiving spatial environment of the site,
2. Ease of observing visitors and their behaviors,
3. Ease of comprehending visitors' questions,
4. Ease of directing tourists' attention,
5. Monitoring the trip's itinerary,
6. Observing tourists' actions and expressions while presenting,
7. Feeling mental performance caused by visitors' behaviors and expressions, and
8. Confidence in giving tours after practicing with the system.

Participants rated their level of agreement on a 7-point scale (1 = "Strongly Disagree" and 7 = "Strongly Agree"). Each item in the following list includes the name of the measure of effectiveness, followed by the statement rated by participants:

1. **Site Perception (SP):** I had a realistic view of the spatial environment of the tour site about which I needed to present.
2. **Ease of Tourist Observation (ETO):** I could easily observe visitors and their behaviors (e.g., where visitors were looking, their expressions).
3. **Ease of Tourist Referencing Perception (ETRP):** When visitors asked or mentioned a point or an object during the presentation, I could easily know what they were referring to or asking about.
4. **Ease of Attracting Tourists (EAT):** I could easily direct tourists' attention to a point or an object during a presentation.
5. **Itinerary Monitoring (IM):** I could easily know what the next place in the video was to prepare the presentation about it.
6. **Actual Observation on Tourists (AOT):** I was observing visitors' actions and expressions while presenting.
7. **Perceived Mental Effect (PME):** I felt that visitors' expressions and behaviors affected my mental state (e.g., pressure, excitement, concentration) when presenting.
8. **Self Confidence (SC):** I felt confident in my ability to guide tours (understanding places, interacting with tourists, or responding to situations) after practicing with the system.

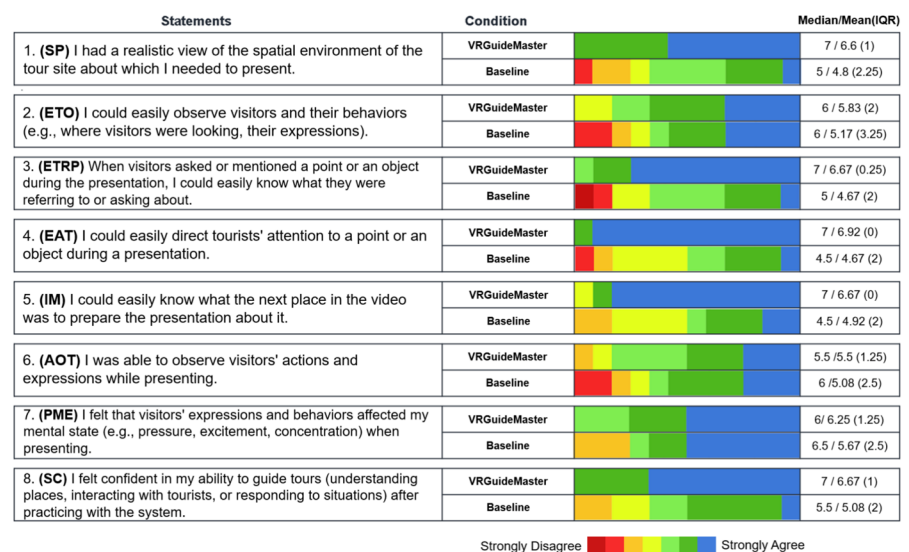
This questionnaire was iteratively proposed and refined by the research team based on the necessary skills tour guide trainees often need to practice that we learned from the need-finding interviews (Ly et al. 2022). We also asked the tourism lecturers in the need-finding interviews to validate the questionnaire. Overall, the lecturers confirmed that the questionnaire fundamentally covered most of the aspects

needed for a training solution for tour guiding despite missing criteria on tour guide trainees' knowledge of the tour site such as understandings of cultures or local specialties. However, they agreed with us that this aspect might not belong to the scope of this study, as it depended more on how much the trainees had studied about the site by themselves in advance, through books, movies or direct exposures to local living environments through onsite visits, rather than impacted by the training solutions in our studies.

We also collected qualitative feedback of participants through semi-structured interviews to obtain more details on how they perceived the strengths and weaknesses of the interfaces as well as points for improvements. Below are some primary questions asked to the participants which could have led to other follow-up questions depending on the participants' answers:

1. How do you feel about the realism of the tour sites presented on the interfaces?
2. What do you think about your ability to navigate the tour sites using the interfaces?
3. What do you think about the presence of the tourists on the interfaces?
4. What do you think about the communication and interaction between you and the tourists during the tour-guiding sessions using the two systems?
5. What do you think about your ability to observe tourists' behaviors and expressions using the two systems?
6. What do you think about the usefulness and convenience of the trip itinerary in each system?
7. According to you, what are the advantages and disadvantages of each system for tour-guiding training?
8. What do you think about the effectiveness of the two systems for tour-guide training?
9. If possible, how would you improve the systems?

Fig. 10 Detailed ratings of participants for the two interface conditions in the post-study questionnaire. The table also provides the median value, the mean value and the interquartile range (IQR) of the participants' ratings for each interface in each statement in the questionnaire



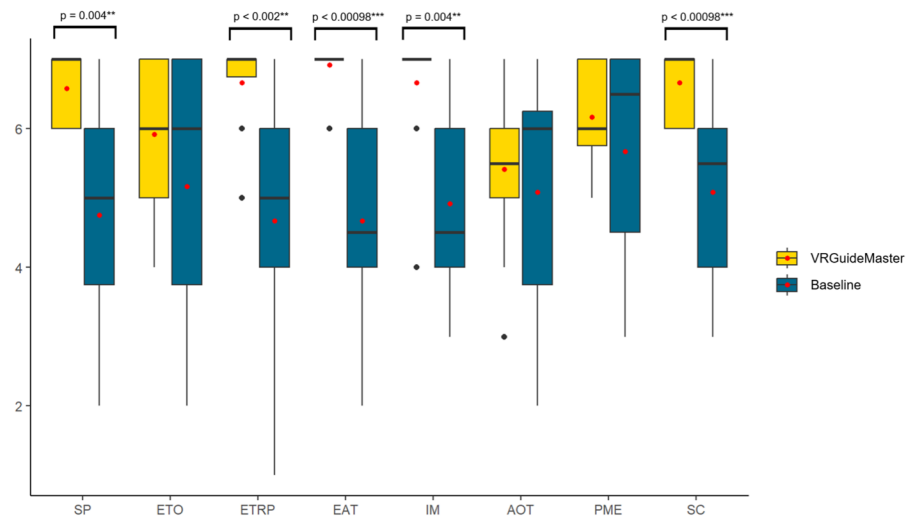
4.7 Results

4.7.1 Quantitative data

Figure 10 provides an overview of all participants' ratings for the questionnaires in both conditions. The interquartile ranges (IQRs) in the figure illustrated that the participants' ratings for the baseline condition were much more spread out than those for VRGuideMaster. We performed the Wilcoxon signed rank test to examine if there were significant differences in the participants' ratings between the two interface conditions. Overall, participants rated VRGuideMaster significantly more effective for practicing tour guiding than the baseline in five over eight questions in the questionnaire.

More specifically, with VRGuideMaster, participants reported that they had a more realistic view of the spatial environment of the tour site about which they were presenting than when using a conventional video conferencing tool combined with screen-sharing showing 360 videos (6.6 vs 4.8, $Z = 2.87$, $p = 0.0039$, $r = 0.59$) (SP). When tourists asked about or referred to a point or an object during their presentation using VRGuideMaster, it was significantly easier for the participants to know what the tourists were referring to or asking about (6.67 vs 4.67, $Z = 2.98$, $p = 0.002$, $r = 0.61$) (ETRP). Similarly, using VRGuideMaster, the trainees could also direct tourists' attention to an object or a landmark more effectively than when using the baseline interface (6.92 vs 4.67, $Z = 3.05$, $p = 0.00098$, $r = 0.62$) (EAT). Participants also reported that compared to the baseline, using VRGuideMaster was significantly easier for them to know what the next place in the itinerary was in order to prepare for presenting about it (6.67 vs 4.92, $Z = 2.87$, $p = 0.0039$, $r = 0.59$) (IM). Additionally, participants also rated significantly higher scores of self-confidence in their tour-guiding ability after training with

Fig. 11 Boxplot visualizations of participants' ratings for post-study questionnaires (The red circle in the center of each box indicates the mean value while the black horizontal line represents the median value). Questions where there are statistically significant differences in participants' rating between two conditions are annotated with corresponding p values



VRGuideMaster compared to the baseline (6.67 vs 5.08, $Z = 3.07$, $p = 0.00098$, $r = 0.63$) (SC).

Participants' ratings in the other three questions in the questionnaires also depicted slightly higher effects of VRGuideMaster compared to the baseline condition, even though Wilcoxon signed rank tests did not yield significant results. More specifically, participants found that VRGuideMaster could help them marginally more easily perceive tourists' expressions and behaviors during their presentations (5.83 vs 5.17, $Z = 1.32$, $p = 0.21$) (ETO). Similarly, participants' ratings showed a slightly better ability to observe tourists' expressions and behaviors during their VRGuideMaster presentation than the baseline condition (5.5 vs 5.08, $Z = 1.11$, $p = 0.34$) (AOT). Likewise, they also reported that seeing tourists' behaviors and expressions in VRGuideMaster affected their mentality more than in the baseline (6.25 vs 5.67, $Z = 1.14$, $p = 0.25$) (PME). Figure 11 provides the boxplot visualizations of participants' ratings for both interface conditions in the post-study questionnaire.

We also performed Wilcoxon signed rank tests to examine if the scenario videos used in the study or their experimentation orders had any effects on participants' ratings on the questionnaire. The test did not yield any significant differences between the videos or the video orders on participants' ratings for all the eight questions.

4.7.2 Qualitative feedback

With the baseline interface, all participants found it similar to the conventional Zoom interface which they were already very familiar with. Thus, they did not have much to comment about it. In contrast, VRGuideMaster was found to provide novel experiences with several interesting aspects for tour-guiding training that they would like to discuss.

All participants reported that viewing 360 videos in VR provided more a realistic impression and immersion of the

tour sites than viewing them with the baseline interface ("The VR interface definitely offered a much more realistic experience of being at the tour site." (P5)). Seeing tourists' 3D avatars in the virtual environment was also deemed to be beneficial to tour guide trainees ("Of course, seeing the tourists navigating in the same space feels more realistic than seeing their videos on the Zoom interface." (P5)). Nevertheless, tourists' gazes shown on the video feeds of the avatars were deemed to be somehow unnatural to the participants. P2, P3, and P7 reported that they felt that the tourists seemed not to look at them, thus "the communication was not really natural" (P3).

Additionally, participants' feedback highlighted significant advantages of VRGuideMaster compared to the baseline interface. Particularly, P7, a student who had quite extensive professional tour guiding experiences thanks to his part-time jobs at some local tourism companies, reported "even though giving tours in the VR interface (VRGuideMaster) was not still the same as in actual field trips, but the realism was already around 70 percent compared to the real life, which was much better compared to the video conferencing". He also commented that he really hoped to see the deployment of VRGuideMaster in tourism institutions and training centers ("This system will be a game-changing solution for the current landscape of tour-guide training." (P7)).

Regarding the value of directional audio rendering, some participants (P4, P7, P9) commented that the audio directions actually helped them quickly locate the tourists who were standing outside of their views and asking a questions. P7, a student with extensive hands-on tour guiding experience, said that such behaviour was really similar to what he often encountered in reality. However, the audio directions were perceived not very accurately rendered ("Sometimes I heard the voice coming right behind my back but the speaker was actually a bit to the right behind me" (P9)).

5 Discussion

In this section, based on the quantitative result and qualitative feedback from the study, we will discuss the effects of VRGuideMaster's different features on the participants as well as its potential application in other domains.

5.1 Immersion and spatial interactions improves efficacy in navigation and communication

Overall, the study results reflected clear benefits of VRGuideMaster's design features in providing an effective environment for training different tour-guiding skills. Viewing 360-degree videos of tour sites through a VR headset provides a more immersive way to observe their spatial arrangement and realistic appearance thanks to a wider view frustum than that of a conventional computer screen. In addition, VRGuideMaster also supports an intuitive and efficient way of spatial navigation based on head movements compared to mouse-based dragging interactions in the baseline condition. As a result, this might have led to less disruptive viewing experiences of 360 video content during the training session. We argue that this contributed to their subjective perception of more effectively understanding the spatial environment of a site.

As we expected, compared to the baseline where tourists could only follow the tour guide's perspective on the screen sharing of the 360 videos, the pointing feature offered by the tourists' interfaces provided them with a more expressive and intuitive way to communicate with the tour guide. As a result, this feature allowed the guide to be more easily aware of what a tourist was saying and to respond quickly. Similarly, the pointing feature in the VR interface was an effective tool to help tour guides efficiently direct tourist attention towards an object or a location. The spatial audio rendering of tourists' voices in the VR interface also played a useful role in more efficient communication. While there were still limitations in the spatial rendering of the audio, based on the feedback of the participants, we argue that the spatial audio feature had contributed to making the VR interface more realistic. With the continuous advancement of VR headsets, not only in terms of visual rendering but also auditory capabilities, we expect the accuracy of audio spatial rendering could be addressed in future devices.

VRGuideMaster focused on creating an immersive environment based on 360-degree videos due to their prevalence, realism, and affordable production costs. However, a noticeable limitation of these contents is their flat 2D representation which might affect users' immersion. With the emergence of novel 3D construction techniques such as Gaussian Splatting (Fei et al. 2024), especially utilizing 360 images and videos as input references (Rey-Area

and Richardt 2025), in the future the cost for producing 3D environments with high realism levels can be significantly reduced. Therefore, in the next steps, we will consider extending VRGuideMaster to incorporate 3D meshes generated by these techniques and conducting additional studies comparing such 3D environment representations to 360 videos in terms of their effectiveness for tour guide training.

In this study, the superior performance of VRGuideMaster compared to the baseline condition, which is a desktop-based video conferencing interface, might have been the result of the collective effects of immersion in the VR environment, the different interaction designs incorporated in the system, and even probably the novelty effect of the technologies on the participants. To better examine the separate effects of these factors on tour guide trainees' performance and experience, future studies should consider isolating them in different interface conditions such as VR environments on desktops or headsets, VR environments with and without spatial-awareness support or itinerary management features. In the future, we might consider conducting a long-term user study where each trainee can practice with the system multiple times. With this study design, due to the frequent exposure of the trainees with the VR technology in an extended period, this should enable us to eliminate the technological novelty from users' perception and experience with the system.

5.2 Virtual itineraries support efficient destination tracking

Compared to the conventional practice of using a paper-based itinerary, the virtual itinerary of the VR interface provides the participants a more convenient way for keeping track of the agenda. The virtual itinerary offers live visual updates on the list of places along the trip, automatically scrolling through places visited and moderating text sizes, thereby reducing the cognitive and physical efforts of the tour guide. In addition, by overlaying the list of places on the 360 videos, the virtual itinerary provides tour guides the ability to maintain their awareness of the trip's progress while still observing what is happening on the 360 video of the site. Compared to the conventional setting where the guide had to switch between the 360 video on the screen and the paper, this virtual itinerary feature is far more efficient.

5.3 Merits and drawbacks of tourists' video streams

The reason why VRGuideMaster did not significantly help participants better perceive facial expressions or behaviors of tourists was because the 3D avatars sometimes turned their back towards the tour-guide when the corresponding tourist was watching certain parts of the 360 videos. The baseline

condition using a conventional videoconferencing interface, on the other hand, always showed the video feeds of the tourists on a 2D panel, thus their faces were always visible to the guide. However, participants commented that even though tourists' facial expressions and behaviors were not always visible, this resembled the reality as in real life, sometimes they could only see the backs of the tourists, even when asking questions. In those cases, the guide could also only hear the tourists' voice and see their hand pointing towards the object of interest, which is similar to what happened in VRGuideMaster when the participants saw the backs of some avatars and their pointing highlighters. Nevertheless, a limitation of the tourist avatars' representation is that they look almost the same from the back, which created challenges for the tour guide to visually differentiate the tourists to each other. In the future, we can consider allowing tourists to visually personalize their avatars in order to make them more distinguishable.

The conventional video conferencing interface displayed the tourists' video streams on two panels, thus their faces and behaviors were always visible to the tour guide trainees. However, the participants reported that they had to frequently switch their attentions between the 360 video and the tourists' video feeds while giving the tour. Moreover, during certain time periods in their presentation when they needed to focus on the scenery or objects in the 360 video, they forgot to look at the video panel to observe the expressions and behaviors on the tourists. With VRGuideMaster, participants reported that when they were presenting about certain objects or locations on the 360 video, it was easy and quite natural for them to glance at the video feeds of the tourists' 3D avatars nearby to keep track of their expressions and behaviors as they usually do in the real life. Thus by integrating tourists' video feeds as 3D avatars in the VR environment of the 360 videos, VRGuideMaster provided spatially contextual awareness for a tour guide to seamlessly switch their attention between the scenery in the 360 video and the tourists in the VR environment.

Regarding the "unnatural gaze" of the tourists as perceived by some participants, in fact, it was caused by the spatial disparity between the built-in camera of commodity laptops or mobile devices and the device's display, which has already been reported in previous research works on remote collaboration (Gemmell et al. 2000; Giger et al. 2014; Le et al. 2019). For example, when a tourist was looking at the face of the tour guide's avatar, his/her gaze as conveyed on their video feeds in the VR environment appeared to look down at the avatar's body instead of at the face. Such distortions of the tourists' conveyed gaze might have reduced the effect of the tourists' video feeds on the mental state of the tour guide. However, compared to the conventional video conferencing interface, in VRGuideMaster, the 3D avatars could externalize the current attention of the

corresponding tourists more clearly through their orientation and even position in the VR environment. We argue that this could lead to a higher presence of the tourists, which could explain for the slightly higher ratings of their effect on the tour-guide trainee's mentality during the training sessions. In the future of VRGuideMaster, the distorted gaze issue can be mitigated by leveraging advances in computer vision to correct the looking direction of the tourists' eyes in the videos (Gemmell et al. 2000; Giger et al. 2014). Corrected gaze might improve the presence of the tourists as perceived by the tour-guide trainee, which can potentially lead to more effects of their presence on the tour-guide trainee's mentality, making the training more realistic. Further empirical studies will be needed to explore this in the future.

5.4 Possible improvements of tour-guide trainee's representation and interaction

In the current implementation of VRGuideMaster, the tour-guide trainee is represented as a stylized 3D avatar in VR. Besides that, trainee users are only given two choices of 3D avatars with pre-defined appearances. This might hamper the communication and interaction between the trainee and tourists during a tour-guiding session. First, with tourists, the lack of personalization of the avatar's representation might reduce their engagement with the tour guide trainee during the trip. On the other side, the tour guide trainee might not feel comfortable to freely perform different facial expressions as some particular features on their physical appearance might not be adequately reflected on the generic avatar. However, with recent technical capabilities of latest MR headsets such as Apple Vision Pro, Meta Quest 3 or Meta Quest Pro combined with the rapid advancements of generative artificial intelligence, such issues could be mitigated. The generic stylized avatars could be replaced by more personalized ones, which can be generated by either requiring the trainee to upload his/her portrait image or scanning his/her face using the built-in camera of the device before entering the VR environment. Such imagery data can be used to generate a 3D avatar that realistically resembles the physical appearance of the trainee (Cheng et al. 2024; Zhang et al. 2023). Besides that, the trainee user's complex facial expressions could also be adequately captured by these headsets, enabling more realistic subtle expressions of the trainee's avatar in VR. This might enhance the trainee's confidence to externalize various expressions as well as improve tourists' engagement. Moreover, the eye tracking features on these headsets also enable more precise gaze representations of the trainee's avatar in VR (Waisberg et al. 2024). In addition to that, the hand pose detection and tracking features on these devices can also be leveraged to allow trainee users to perform natural bare-hand

interactions with the tourist site in VR rather than relying touch controllers as in the current prototype.

5.5 Potential applications in other domains

Besides tourism, the concept of VRGuideMaster could also be applied to training in domains like geology, archaeology, environmental sciences, medicine, and logistics. Field trips in these domains are often expensive and can also contain threats to inexperienced practitioners or harmful to physical artefacts at the site. Using the various features of VRGuideMaster, a trainer could safely explain, instruct and help trainees familiarize themselves with the environment and potential threats of the site prior to a field trip. This can thus help them better prepare themselves to avoid undesirable incidents when they are onsite.

6 Limitations

Although participants' subjective data has demonstrated the positive effects of VRGuideMaster on user experience in practicing tour guiding, further studies are needed to investigate how VRGuideMaster might affect other objective factors. For example, future studies could collect tour-guide trainees' gaze data to examine if when using VRGuideMaster, tour-guide trainees look at tourists when giving tours more than they do when using a conventional video conferencing tool. Furthermore, even though in our study the participants reported to be significantly more self-confident in their skills after practicing using VRGuideMaster, we believe that it would be interesting to involve tour-guide trainers or coaches to evaluate the trainees' tour-giving performance in different conditions. This would provide more objective insights into the effects of VRGuideMaster on the trainees' skills. From the system's perspectives, trip itineraries were still manually prepared in a separate software. It would be more convenient for tour guide trainees and trainers if VRGuideMaster or similar solutions could offer a tool to efficiently author trip itineraries while using the system.

Although the sample size of this study ($n = 12$) seems to be quite small, all statistical tests which resulted in significances also yielded large effect sizes ($r > 0.5$). This suggested that even for a limited quantity of participants, the effects of the interface were strong enough to be statistically detected. Nevertheless, future studies might consider a large sample size to validate our findings in this study. In this user study, we did not interview the volunteer tourists regarding their experience when attending different tour guiding sessions using different interfaces due to their repetitive involvement in the user study with multiple participants. This can lead to learning effects with the interface

conditions and the scenario videos, potentially causing biases in their perception of the interface effects. Likewise, the requirement in which each tourist had to pose at least three questions in each session might also have forced them to be unnaturally attentive to the participants' performance regardless which tourist interface they were using. Future studies could consider evaluating the effectiveness of the tour-guiding interface on the trainee's performance from tourists' perspectives. Such user studies should consider tourists as primary experimental subjects, ensuring their natural involvement in the tour guiding sessions and avoiding learning effects of the tour agenda on them.

Additionally, volunteer trainees in this user study were tourism students, who mostly lacked of professional hands-on experience. In the future, we will consider evaluating the systems with professional tour guides to develop more comprehensive understandings on how the system affects tour guide professionals with broad levels of experience. Besides that, while we focused only tour guide training where trainees learn and develop their professional knowledge and skills through interacting with the tourist site environment and live tourists, it could be beneficial to compare these approaches with a more conventional method where trainees watch pre-recorded videos of real-world guided tours.

7 Conclusion

Tour guides, especially tour guide trainees, need to intensively practice their knowledge about tour places as well as their presentation and communication skills with tourists in order to develop or maintain their professional proficiency. Current practices of training in field trips can provide trainees with realistic contexts to effectively train those skills, but often come with financial and logistic barriers for many trainees and thus reduce their professional readiness to compete in the job market. Towards addressing this issue, we designed VRGuideMaster; a collaborative virtual reality solution based on 360 videos to open the opportunity for tour guide trainees to practice their professional knowledge and skills without having to take field trips. Our empirical user study yielded several benefits of VRGuideMaster over a common method for practicing tour giving, leveraging the combination of a conventional video conferencing tool and 360 videos in supporting different aspects in practicing tour guiding. Overall, VRGuideMaster allowed tour guide trainees to better perceive the space and appearance of tour sites and more realistically communicate with tourists, thus more effectively practicing presentation and communication skills, and generally feeling more confident about their tour guiding skills after training with the system. To further improve the usability of VR systems for tour guiding training, future

systems might consider improving tourists' gaze as conveyed on their video feeds or utilizing novel hardware which can render spatial audio more accurately. Looking ahead, further possibilities emerge. One more step of improvement might be aiming to reduce the live participation of real tourists. This could be achieved by leveraging the latest advancements in generative artificial intelligence to generate AI-based tourists as substitutes to allow trainees to practice more conveniently, without having to look for real people to attend as tourists.

Acknowledgments This research is funded by Vietnam National University, Ho Chi Minh City (VNU-HCM) under grant number DS2024-18-01. We also thank the volunteers for their participation in the interviews and the user study.

Author contributions K-D. L., D-N. L., M. F., T. V. N., and M-T. T. worked on conceptualization and ideation of this research. K-D. L., D-N. L., and T-T. L. worked on implementation of the proposed system. K-D. L., D-N. L., T-T. L., T. V. N., and M-T. T. performed the system testing. K-D. L., C. N., M. F., T. V. N., and M-T. T. designed and conducted the user study. K-D. L., D-N. L., T-T. L., and T. V. N. worked on the manuscript draft. C. N., M. F., and M-T. T. revised the manuscript. All authors reviewed the manuscript prior to the submission.

Funding Open access funding provided by University of Bergen (incl Haukeland University Hospital)

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Chang C-Y, Sung H-Y, Guo J-L, Chang B-Y, Kuo F-R (2022) Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. *Interact Learn Environ* 30(3):400–416. <https://doi.org/10.1080/10494820.2019.1661854>
- Cheng R, Wu N, Varvello M, Chai E, Chen S, Han B (2024) A first look at immersive telepresence on apple vision pro. In: *Proceedings*

- of the 2024 ACM on Internet Measurement Conference, pp. 555–562
- Eiris R, Gheisari M, Esmaceli B (2020) Desktop-based safety training using 360-degree panorama and static virtual reality techniques: a comparative experimental study. *Autom Constr* 109:102969. <https://doi.org/10.1016/j.autcon.2019.102969>
- Facebook Social VR Demo - Oculus Connect 2016. <https://www.youtube.com/watch?v=YulgyKLt3s>
- Fei B, Xu J, Zhang R, Zhou Q, Yang W, He Y (2024) 3d gaussian splatting as new era: A survey. *IEEE Transactions on Visualization and Computer Graphics*
- Ferdig RE, Kosko KW (2020) Implementing 360 video to increase immersion, perceptual capacity, and teacher noticing. *TechTrends* 64(6):849–859. <https://doi.org/10.1007/s11528-020-00522-3>
- Flobak E, Wake JD, Vindenes J, Kahlon S, Nordgreen T, Guribye F (2019) Participatory design of VR scenarios for exposure therapy. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. <https://doi.org/10.1145/3290605.3300799>
- Gemmell J, Toyama K, Zitnick CL, Kang T, Seitz S (2000) Gaze awareness for video-conferencing: a software approach. *IEEE Multimed* 7(4):26–35. <https://doi.org/10.1109/93.895152>
- Giger D, Bazin J-C, Kuster C, Popa T, Gross M (2014) Gaze correction with a single webcam. In: *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. <https://doi.org/10.1109/ICME.2014.6890306>. IEEE
- Hebbel-Seeger A (2017) 360 degrees video and VR for training and marketing within sports. *Athens J Sports* 4(4):243–261
- Henrikson R, Araujo B, Chevalier F, Singh K, Balakrishnan R (2016) Multi-device storyboards for cinematic narratives in VR. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 787–796. <https://doi.org/10.1145/2984511.2984539>
- Ibrahim-Didi K (2015) Immersion within 360 video settings: capitalising on embodied perspectives to develop reflection-in-action within pre-service teacher education. *Res Dev Higher Educ Learn Life Work Complex World* 38:235–245
- Izard SG, Méndez JAJ, García-Peñalvo FJ, López MJ, Vázquez FP, Ruisoto P (2017) 360 vision applications for medical training. In: *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*, pp. 1–7. <https://doi.org/10.1145/3144826.3145405>
- Kittel A, Larkin P, Elsworth N, Lindsay R, Spittle M (2020) Effectiveness of 360 virtual reality and match broadcast video to improve decision-making skill. *Sci Med Footb* 4(4):255–262. <https://doi.org/10.1080/24733938.2020.1754449>
- Kumar K, Poretski L, Li J, Tang A (2022) Tourgether360: collaborative exploration of 360 videos using pseudo-spatial navigation. *Proc ACM Hum-Comput Interact* 6(CSCW2):1–27. <https://doi.org/10.1145/3555604>
- Le K-D, Avellino I, Fleury C, Fjeld M, Kunz A (2019) Gazelens: Guiding attention to improve gaze interpretation in hub-satellite collaboration. In: *Human-Computer Interaction-INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part II* 17, pp. 282–303. https://doi.org/10.1007/978-3-030-29384-0_18. Springer
- Li H, Trutoiu L, Olszewski K, Wei L, Trutna T, Hsieh P-L, Nicholls A, Ma C (2015) Facial performance sensing head-mounted display. *ACM Trans Graph (ToG)* 34(4):1–9. <https://doi.org/10.1145/2766939>
- Lin Y-T, Liao Y-C, Teng S-Y, Chung Y-J, Chan L, Chen B-Y (2017) Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews. In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 255–265. <https://doi.org/10.1145/3126594.3126656>

- Ly D-N, La T-T, Le K-D, Nguyen C, Fjeld M, Tran TN-D, Tran M-T (2022) 360tourguiding: Towards virtual reality training for tour guiding. In: Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–6. <https://doi.org/10.1145/3528575.3551436>
- Lyu H, Vachha C, Chen Q, Pyrinis O, Liou A, Thoravi Kumaravel B, Hartmann B (2022) Webtransceivr: Asymmetrical communication between multiple vr and non-vr users online. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1–7. <https://doi.org/10.1145/3491101.3519816>
- Na L, Weihua H (2012) Virtual reality applications in simulated course for tour guides. In: 2012 7th International Conference on Computer Science & Education (ICCSE), pp. 1672–1674. <https://doi.org/10.1109/ICCSE.2012.6295385>. IEEE
- Nassani A, Zhang L, Bai H, Billingham M (2021) Showmearound: Giving virtual tours using live 360 video. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–4. <https://doi.org/10.1145/3411763.3451555>
- Nguyen C, DiVerdi S, Hertzmann A, Liu F (2017) Collavr: collaborative in-headset review for VR video. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 267–277. <https://doi.org/10.1145/3126594.3126659>
- Oculus Spatializer Features. <https://developer.oculus.com/documentation/unity/audio-spatializer-features/>
- OpenCV for Unity. <https://assetstore.unity.com/packages/tools/integration/opencv-for-unity-21088>
- Overview of Meta Avatars SDK. <https://developer.oculus.com/documentation/unity/meta-avatars-overview/>
- Palmas F, Cichor J, Plecher DA, Klinker G (2019) Acceptance and effectiveness of a virtual reality public speaking training. In: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 363–371. <https://doi.org/10.1109/ISMAR.2019.00034>. IEEE
- Panchuk D, Klusemann MJ, Hadlow SM (2018) Exploring the effectiveness of immersive video for training decision-making capability in elite, youth basketball players. *Front Psychol* 9:2315. <https://doi.org/10.3389/fpsyg.2018.02315>
- Patel D, Hawkins J, Chehab LZ, Martin-Tuite P, Feler J, Tan A, Alpers BS, Pink S, Wang J, Freise J et al (2020) Developing virtual reality trauma training experiences using 360-degree video: tutorial. *J Med Internet Res* 22(12):22420. <https://doi.org/10.2196/22420>
- Pittarello F, Pagini V, Zuffellato L (2020) Design and evaluation of an educational virtual reality application for learning how to perform on a stage. In: Proceedings of the International Conference on Advanced Visual Interfaces, pp. 1–8. <https://doi.org/10.1145/3399715.3399919>
- Prasetyo, RD, Hati SRH (2022) The impact of marketing with 360-degree videos on tourist willingness to travel during the covid-19 pandemic. In: International Academic Conference on Tourism (INTACT) "Post Pandemic Tourism: Trends and Future Directions" (INTACT 2022), pp. 166–188. https://doi.org/10.2991/978-2-494069-73-2_13. Atlantis Press
- Rahimizhian S, Oztüren A, Ilkan M (2020) Emerging realm of 360-degree technology to promote tourism destination. *Technol Soc* 63:101411. <https://doi.org/10.1016/j.techsoc.2020.101411>
- Real-time Voice and Chat Inside Your Game. <https://www.agora.io/en/unity/>
- Repetto C, Di Natale AF, Villani D, Triberti S, Germagnoli S, Riva G (2021) The use of immersive [CDATA[360^{\circ}]] 360° videos for foreign language learning: a study on usage and efficacy among high-school students. *Interactive Learning Environments*, 1–16. <https://doi.org/10.1080/10494820.2020.1863234>
- Rey-Area M, Richardt C (2025) 360^{\circ}[CDATA[360^{\circ}]] 3D photos from a single 360^{\circ}[CDATA[360^{\circ}]] input image. *IEEE Transactions on Visualization and Computer Graphics*
- Rifai MB, Agustina AR, Rachmalya DD, Sarina E, Araffi SS, Agustina A (2021) Virtual tour at the louvre museum Paris and Pattaya floating market using virtual reality technology based on youtube 360. *Int J Adv Tour Hosp* 1(2):304–339. <https://doi.org/10.30645/jurasik.v8i1.613>
- Schäfer A, Reis G, Stricker D (2021) Investigating the sense of presence between handcrafted and panorama based virtual environments. In: Proceedings of Mensch und Computer 2021, pp. 402–405. <https://doi.org/10.1145/3473856.3474024>
- Seyitoğlu F, Atsız O (2022) Distant local-guided tour perceptions and experiences of online travellers. *J Vacat Mark*. <https://doi.org/10.1177/13567667221135198>
- Stupar-Rutenfrans S, Ketelaars LE, van Gisbergen MS (2017) Beat the fear of public speaking: mobile 360 video virtual reality exposure training in home environment reduces public speaking anxiety. *Cyberpsychol Behav Soc Netw* 20(10):624–633. <https://doi.org/10.1089/cyber.2017.0174>
- Takac M, Collett J, Blom KJ, Conduit R, Rehm I, De Foe A (2019) Public speaking anxiety decreases within repeated virtual reality training sessions. *PLoS ONE* 14(5):0216288. <https://doi.org/10.1371/journal.pone.0216288>
- The Ease-of-use of Unity's Networking with the Performance & Reliability of Photon Realtime. <https://www.photonengine.com/pun>
- Thoravi Kumaravel B, Nguyen C, DiVerdi S, Hartmann B (2020) Transceivr: Bridging asymmetrical communication between VR users and external collaborators. In: Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, pp. 182–195. <https://doi.org/10.1145/3379337.3415827>
- Tseng S-P, Huang MW, Liu HJ, Chung CC, Chiu CM (2013) A virtual reality based training system for cultural tourism. In: International Conference on Web-Based Learning, pp. 272–277. https://doi.org/10.1007/978-3-662-46315-4_29. Springer
- Waisberg E, Ong J, Masalkhi M, Zaman N, Sarker P, Lee AG, Tavakoli A (2024) The future of ophthalmology and vision science with the apple vision pro. *Eye* 38(2):242–243
- Walshe N, Driver P (2019) Developing reflective trainee teacher practice with 360-degree video. *Teach Teach Educ* 78:97–105. <https://doi.org/10.1016/j.tate.2018.11.009>
- Wang L, Wang L (2019) Design and implementation of three-dimensional virtual tour guide training system based on unity3d. In: 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), pp. 203–205. <https://doi.org/10.1109/CISCE.2019.00054>. IEEE
- Willems K, Brengman M, Kerrebroeck HV (2019) The impact of representation media on customer engagement in tourism marketing among millennials. *Eur J Mark* 50(9):1988–2017. <https://doi.org/10.1108/EJM-10-2017-0793>
- Wu X, Lai IKW (2022) The use of 360-degree virtual tours to promote mountain walking tourism: stimulus-organism-response model. *Inf Technol Tour* 24(1):85–107. <https://doi.org/10.1007/s40558-021-00218-1>
- Yang T, Lai IKW, Fan ZB, Mo QM (2021) The impact of a 360 virtual tour on the reduction of psychological stress caused by Covid-19. *Technol Soc* 64:101514. <https://doi.org/10.1016/j.techsoc.2020.101514>
- Zhang Z, Giménez Mateu LG, Fort JM (2023) Apple vision pro: a new horizon in psychological research and therapy. *Front Psychol* 14:1280213
- Škola F, Rizvić S, Cozza M, Barbieri L, Bruno F, Skarlatos D, Liarokapis F (2020) Virtual reality with 360-video storytelling in cultural heritage: study of presence, engagement, and immersion. *Sensors* 20(20):5851. <https://doi.org/10.3390/s20205851>

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