# The Owl 2.0: An Enhanced Immersive Telepresence Low-Cost Prototype for Hybrid Events

JUAN TORRES-ARJONA, BALDOMERO RODRÍGUEZ, MARTA ORDUNA, PABLO PÉREZ, ESTER GONZÁLEZ-SOSA, and ÁLVARO VILLEGAS, Nokia eXtended Reality Lab, Spain

In this paper, the Owl 2.0 is described. It is a 360-video streaming prototype for immersive remote communication. We believe the Owl 2.0 might be an alternative to enhance the experience of small-to-medium hybrid events with a semi-balanced number of both local and remote people. The prototype is a full end-to-end system that incorporates a 360-degree camera to capture and transmit, in real-time, audio and 360 video to other users using commercial VR devices. Thus, it allows remote users to teleport to another real place. Among their advanced features, the Owl 2.0 incorporates: i) the possibility of sharing digital content in the immersive scene with other remote users; ii) the representation of remote users with virtual avatars displayed on a tablet attached to the 360 camera or any smartphone using the provided app, so that local people can be also aware of remote users; iii) the virtual hands' representation of each remote user; and iv) a magnifying glass option that helps to get fine details of the immersive scene. Finally, we depict several relevant specific use cases where the Owl 2.0 might be successful such as a small workshop, social gatherings, remote attendance to classes in education, etc. Further, this prototype is available for trial by the scientific community upon request.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing devices**; • **Information systems** → **Multimedia streaming**.

Additional Key Words and Phrases: 360 video, videoconferenring, streaming, immersive

**ACM Reference Format:**
Juan Torres-Arjona, Baldomero Rodríguez, Marta Orduna, Pablo Pérez, Ester González-Sosa, and Álvaro Villegas. 2023. The Owl 2.0: An Enhanced Immersive Telepresence Low-Cost Prototype for Hybrid Events. 1, 1 (October 2023), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 OVERVIEW

Recent works in the field showcase the interest of the eXtended Reality (XR) communities with immersive telepresence communication systems [5]. *Show me around*, the system proposed by [4], was designed for immersive Virtual Tours. One of the participants in the area of interest, commonly the guided tour, holds a 360° camera connected to a laptop computer with audio headphones and a microphone to capture the whole scene. This 360° video is integrated into the open-source video conferencing platform Jitsi, which allows a large number of users ($20 - 30$) to access this immersive content through a 2D screen. There are other related works such as *Exleap*, an immersive telepresence system done by [1] that gives users the possibility to move around the area of interest through the use of different 360° cameras, which

Authors' address: Juan Torres-Arjona, juan.torresarjona@nokia.com; Baldomero Rodríguez, baldomero.rodriguez.ext@nokia.com; Marta Orduna, marta.orduna@nokia.com; Pablo Pérez, pablo.perez@nokia.com; Ester González-Sosa, ester.gonzalez@nokia.com; Álvaro Villegas, alvaro.villegas@nokia.com, Nokia eXtended Reality Lab, Madrid, Spain, 28050.


(a) (b) (c)

Fig. 1. Owl prototype: (a) architecture; (b) physical device; (c) examples of different remote user representations: avatar of an immersive remote user & 2D video of a 2D remote user.

are placed at strategic positions, creating the experience of leaping between places. To achieve this, only HMDs are required.

More recently and in the same line, [6] proposed a web based system for virtual tour applications that combines the use of 360° images with 360° videos. 360° images allow the user to move from one place to another from a remote location. When a particular place is also provided with a 360° camera, users can see a live 360° video. Additionally, this system also supports multiple users, who are represented by very simple avatars comprised of two circles within a larger one, representing eyes and head, respectively. Concerning representation of remote users for local users, *Think Fast* [7] represents the remote user's avatar on a tablet attached to the 360° camera; or even by rendering his/her upper body on top of the robot so that the local people can see him/her using AR glasses.

Concerning commercial solutions, *Avatour*[1]is a remote collaboration platform designed for industrial applications. The prototype consists of a 360° camera attached to a smartphone, which in turn is placed on a stick held by the local user. The remote user can join the session using a 2D screen or a Head Mounted Display (HMD), and he/she has the possibility to make annotations or record sessions.

The system that we developed, the Owl 2.0, is focused on both the remote scenario and the local people placed there, allowing communications very similar to face-to-face. It is a bi-directional real-time immersive telepresence system that focuses on the real-time capture, delivery and rendering of real spaces, including the people that are present in them [2]. It incorporates a 360-degree camera that captures and transmits, in real-time, audio and 360 video to those who are connected to the session using commercial VR devices. In turn, the VR devices transmit the user's hands and head movements which are displayed in the remote locations. The simplified architecture of the Owl 2.0 is shown in Fig. 1(a). The camera is plugged into a Single Board Computer with a touch screen in charge of the capturing and streaming logic, a 4G/5G modem, a commercial hands-free speaker/microphone and a power bank as the power supply for all the devices. All the components are placed together in a 3D-printed custom housing, which can be either mounted on a tripod or placed on a flat surface.

[1]https://avatour.co/
[2]The Owl has been the recipient of the ISMAR 2021 Contest Honorable Mention https://ismar2021.vgtc.org/program/awards/index.html

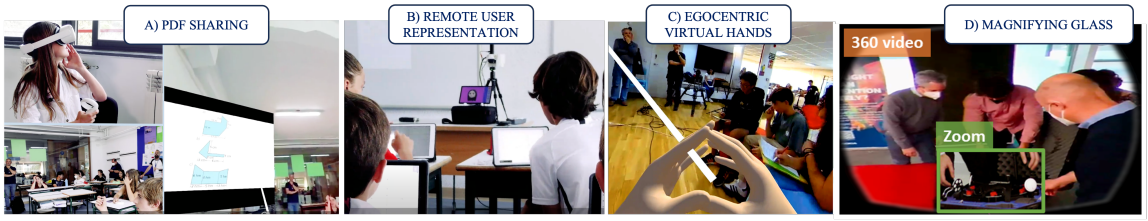


Fig. 2. The Owl 2.0 has some additional features for enhancing the experience: (a) content sharing with other remote users; (b) representation of the remote user through avatars so that local people can see how they move their head, hands and mouth; (c) representation of remote users in the immersive scene, so that all remote users can see each other with avatar representations (for other immersive users) or 2D screens (for 2D remote users); and (d) representation of a magnifying virtual glass to see details not provided by the 360 camera.

In the next section, we explain the technical details of the system including the aspects related to advanced features, the types of users, the immersion feeling, the physical interaction and other convenience issues. Finally, we describe representative real experiences including several use cases, and discuss the main knowledge learned from them.

## 2 TECHNICAL CHARACTERISTICS

The Owl 2.0 was built upon the previous version Owl. For more details, we refer the readers to [2, 3]. Together with the transmission of the multimedia data captured by the prototype, the Owl 2.0 has relevant features that enhance the immersive experience:

- **Content sharing**: the Owl 2.0 brings the possibility to share visual content such as slides, images or videos. Through the VR controller, the user has the freedom to place the content at any point of the 360° scene and scale it (see Fig. 2(a))
- **Representation of remote users**: Remote users (section 2.1) are represented through cartoon avatars, as depicted in Fig. 1. They can be customized to match user's identity in terms of shape and color of the face, shape and color of eyes and eyebrows, nose shape, mouth shape, lip color, ear shape, hair style, beard style, freckles and glasses. Head pose and voice of each remote participant are used to control the avatar's head, hands and lips in real-time. These avatars are used to represent remote users for local users by placing a tablet with the avatar(s) next to the Owl 2.0 prototype and also to represent remote users in the immersive scene, so that all remote users can see each other (Fig. 2(b)). For the case of 2D remote users, a 2D video of the user is shown instead.
- **Egocentric virtual hands**: Remote users can see within the HMD a virtual representation of his/her own hands (Fig. 2(c)) in real-time. This provides to the remote users a better sense of being in a remote location. This is achieved by using the tracking data provided by the HMD.
- **Magnifying glass**: Sometimes the resolution of the displayed video in the HMDs is not enough to distinguish fine details of the immersive scene (e. g. text), or the 360 camera cannot show some parts of the scenario (e. g. due to occlusions). To cope with these situations, it is possible to use the camera of an additional smartphone in the remote location that points at the desired place. The captured video is then streamed to the remote devices and it is displayed in the HMD at a configurable floating location (Fig. 2(d)).

Once the main features of the Owl 2.0 have been presented, we get deeper into the technical details, starting by identifying the types of users that can participate in the experience. Later, we continue with the technologies involved in the immersion, how the users interact with the system, and its convenience.

### 2.1 Users

Within the system, we distinguish three types of users:

(1) **Local users:** Those present in the same location as the Owl 2.0.
(2) **Immersive remote users:** The participants who feel immersed in the physical location where the Owl 2.0 is placed, through the use of a HMD. The system allows multiple remote users to connect to the same immersive video session while talking with each other and with the local users. These remote users can see the local environment and interact with the participants in real-time. They can freely move their heads visualizing the scene from the acquisition perspective of the camera.
(3) **2D remote users:** With the idea of increasing accessibility to this system when there is no possibility of using a HMD, users can also join the session using regular smartphones or tablets. Despite not being an immersive experience, it still allows the user to navigate in real-time through the 360° scene (manually shifting the view) and better observe the non-verbal language of local users without any additional cost.

### 2.2 Immersion

Captured video from the 360 camera is compressed and streamed in H.264 format with a customizable bitrate and a video resolution of up to 4K. To get a good quality of experience a bitrate of 8 Mbps or higher is recommendable, although it can operate at lower bandwidth. A backend is in charge of orchestrating the video streams requested by multiple users using RTP transmission. It includes a session control handling the required signaling, coded in Python, and a scalable video processing module coded in C. The backend is executed on a far edge server which allows scaling the service to a large number of users as well as enhancing the real-time video streams with additional Augmented Reality (AR) and Artificial Intelligence (AI) capabilities by adding processing power in dedicated GPU hardware. For audio communication, the open-source software Mumble[3] is used. There is a server instance of Mumble running in the edge and all the devices that send and receive audio run a client instance locally and are connected to the same session in the server.

### 2.3 Physical interaction

The Owl 2.0 has the ability to allow communications in a more realistic way than traditional communication platforms, transmitting more information related to body language, gaze, etc. and allowing remote users to freely move their head along the yaw, roll and pitch axes (3 degrees of freedom, 3 DoF). Thanks to the content sharing feature (Section 2), any remote user can share digital content that is visible to all remote users joining the immersive experience. Also, with the help of a local user joining the immersive session with a 2D user, a zoom area can overlay the immersive scene (magnifying glass feature), providing the remote with a more fine-detail view of a particular local area (very useful with text). Last, the remote user is provided with a virtual representation of their hands, crucial as feedback when talking, but also important to increase their sense of presence (being physically in the remote place).

### 2.4 Convenience

This prototype has a form factor of 15X25X12 cm (width X height X depth). It can be used either with a short tripod to be placed on a table or a chair (for those experiences whether the remote user is expected to be seated), or with a larger tripod simulating the average height of human eyes.

[3]https://www.mumble.info/


(a) Poster presentation (b) Working groups

Fig. 3. Use cases of the Owl

From the technical point of view, the architecture of the system is scalable regarding the number of immersive remote users. However, due to the asymmetric nature of the experience, we expect the Owl 2.0 to work better in small to medium events, where the number of local and remote users is approximately balanced.

In terms of connectivity, the Owl 2.0 is portable and can operate wireless or wired. In wireless operation, the prototype has a 10000 mAh battery that provides enough power to feed all the devices, and data transmission can be done via WiFi or using a 4G/5G connection through an attached modem (included). In wired operation mode, the prototype can be connected to the main electricity supply and the Ethernet interface can be used for the data.

For remote users, an application is provided for Meta/Oculus HMDs, typically Quest2. 2D users can install an application developed for Android devices.

## 3 EXPERIENCE

We envision this system to be particularly useful for two scenarios: a professional or private social gathering group, and the presentation of a poster in a conference. Both are introduced below.

### 3.1 Working/social gathering group

To evaluate its capabilities, it should be tested with at least two people locally and up to two people remotely. On the local side, it is necessary to consider that the location where the Owl 2.0 is placed is the remote user's acquisition perspective. Therefore, it is necessary to locate it at a height equivalent to the eyes of a seated person.

Note that the communication is strongly asymmetric for the local users, who cannot view the remote ones directly. To help with this, the tablet presents the avatars (see Figure 1) whose lip, head and hand movements are synchronized with those of the remote users. This also facilitates a more natural interaction between local users: when talking to remote users they can look at avatars instead of the prototype.

### 3.2 Poster presentation

In this configuration, it is considered that a remote participant is attending a poster presentation in a scientific conference. The remote participant must be able to see the poster, ask questions to the local presenter and discuss with other attendees, either local or remote. This means that the location of the Owl 2.0 is modified since it has to be close to the poster on which the presentation is based. Ideally, the Owl 2.0 should be located where the attendee would be if he/she was there in person, close to the poster itself. As an additional tool, if the resolution of the video shown in the display is

not sharp enough to clearly see the details of the poster, the local presenter can use the magnifying glass (section 2) to help the remote user see the poster more clearly. In this experience, the hands seen in the cartoon avatar synchronized with the movements of the remote presenter become relevant, since they make communication more effective.

## ACKNOWLEDGMENTS

This work was done as part of the COMODIA Project, partially funded by the Ministry of Economy and Digital Transformation. Number of record: 2021/C005/00144164.

## REFERENCES

[1] Atsushi Izumihara, Daisuke Uriu, Atsushi Hiyama, and Masahiko Inami. 2019. ExLeap: Minimal and highly available telepresence system creating leaping experience. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1321–1322.

[2] Redouane Kachach, Sandra Morcuende, Diego Gonzalez-Morin, Pablo Perez-Garcia, Ester Gonzalez-Sosa, Francisco Pereira, and Alvaro Villegas. 2021. The Owl: Immersive Telepresence Communication for Hybrid Conferences. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 451–452.

[3] Redouane Kachach, Diego González Morín, Francisco Pereira, Pablo Perez, Ignacio Benito, Jaime Ruiz, Ester González, and Alvaro Villegas. 2021. A multi-peer, low cost immersive communication system for pandemic times. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 691–692.

[4] Alaeddin Nassani, Li Zhang, Huidong Bai, and Mark Billinghurst. 2021. ShowMeAround: Giving Virtual Tours Using Live 360 Video. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–4.

[5] Pablo Pérez, Ester Gonzalez-Sosa, Jesús Gutiérrez, and Narciso García. 2022. Emerging Immersive Communication Systems: Overview, Taxonomy, and Good Practices for QoE Assessment. *Frontiers in Signal Processing* 2 (2022). https://doi.org/10.3389/frsip.2022.917684

[6] Aizawa Shuta, Sakurai Shouta, and Duc Nguyen. 2023. Design and Implementation of a Web-based Multi-user Virtual Reality System for Virtual Tour of Remote Places. In *Proceedings of the 15th International Workshop on Immersive Mixed and Virtual Environment Systems*. 34–36.

[7] Jingxin Zhang, Nicholas Katzakis, Fariba Mostajeran, Paul Lubos, and Frank Steinicke. 2020. Think Fast: Rapid Localization of Teleoperator Gaze in 360° Hosted Telepresence. *International Journal of Humanoid Robotics* 17, 03 (2020), 1950038.