University of Kent

Kent Academic Repository

Ruiu, Pietro, Giorgi, Ioanna, Casu, Filippo, Mascia, Lorenzo, Bates, Chloe, Grosso, Enrico and Masala, Giovanni Luca (2026) *Towards immersive communication: A user-centered investigation of life-sized 3D telepresence.* Computers in Human Behavior Reports, 23 (101280). ISSN 2451-9588.

Downloaded from
https://kar.kent.ac.uk/117876/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/10.1016/j.chbr.2026.101280

This document version
Publisher pdf

DOI for this version

Licence for this version
CC BY-NC (Attribution-NonCommercial)

Additional information

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Computers in Human Behavior Reports 23 (2026) 101280

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Towards immersive communication: A user-centered investigation of life-sized 3D telepresence

Pietro Ruiu [a,*], Ioanna Giorgi [b], Filippo Casu [a], Lorenzo Mascia [a], Chloe Bates [b], Enrico Grosso [a], Giovanni Luca Masala [b]

[a] Department of Engineering, University of Sassari, Sassari, SS, 07100, Italy
[b] University of Kent, Kent, Canterbury, England, CT2 7NZ, UK

A B S T R A C T

This paper envisions a future where people engage as immersive life-sized, three-dimensional representations, with remote interactions feeling as natural and encompassing as physical presence. For this to become ubiquitous, it is imperative to understand the attitudes of potential users and adoption patterns. To this aim, we present the first empirical investigation where real end users experience an immersive telepresence system as a medium for real-time communication, similar to existing digital communication media. To maximize ecological validity when exploring user perceptions, we designed a realistic prototype system that allowed participants to experience a tangible and credible approximation of what future immersive systems could deliver. In our design, participants could perceive their communicating partner as a life-sized 3D photorealistic representation in real time through mixed reality headsets. This allowed perceptions, user experience, and behavioural outcomes to emerge more authentically, instead of relying on hypothetical vignettes.

Our mixed-methods analyses showed that participants perceived the immersive call as more realistic, engaging and socially present, given a sense of shared space. Their willingness to adopt it increased markedly, approaching parity with video calls despite the technology's much earlier maturity. Participants identified additional benefit for professional and clinical use, and great potential as a credible alternative to in-person consultations, provided sufficient fidelity and reliability. Concerns about headset comfort, motion sickness and technical stability were raised, but also willingness to upgrade in the future, indicating that current barriers relate to practical implementation rather than user acceptance of the underlying paradigm. The study demonstrates the promise of this innovation and highlights the importance of user-centred design for its ongoing development.

1. Introduction

Advances in high-speed internet and digital communication tools have transformed how people interact remotely, making it possible to convey rich visual, auditory and social information across distances (Stieger et al., 2023). Modern video-conferencing platforms allow users to exchange facial expressions, gestures and vocal cues with minimal delay, supporting more extensive non-verbal, visual, and contextual cues, compared to leaner media, such as text-based systems (Short et al., 1976). As such, they provide better resources for conveying interpersonal immediacy, enabling users to experience their communication partner as real, immediate, and socially salient (i.e., social presence) (Short et al., 1976). However, they may still fall short of the experiential depth afforded by physical co-location, or at least the subjective sense of sharing the same space and time with another person, including feeling as if one's actions are reciprocally perceived (i.e., co-presence) (Stieger et al., 2023). While the constructs of social presence and co-presence show conceptual overlap and are sometimes used interchangeably in the literature, both capture key experiential dimensions of mediated interaction that influence engagement, trust, and communication quality.

Immersive telepresence has thus emerged as a niche approach that promises to go beyond conventional video, as it leverages embodied, spatial and scale-consistent representations that can simultaneously target the experiential dimensions of social and co-presence in remote interactions. Such features are now possible following growing traction in the representation of the human body in virtual environments. At the same time, 3D reconstruction techniques have advanced rapidly, largely propelled by progress in artificial intelligence (Lombardi et al., 2019; Saito et al., 2020; Wang et al., 2024). These innovations have markedly improved the accuracy and efficiency of generating three-dimensional models from various data sources, including images, scans, and videos (Khirodkar et al., 2024; Ruiu et al., 2025; Xie et al., 2024). As a result, 3D reconstruction has become increasingly precise, automated, and accessible, fostering its adoption in diverse fields such as healthcare (Lo et al., 2023; Ruiu, Nitti, et al., 2024), Ruiu, Mascia, & Grosso,

* Corresponding author.
E-mail address: pruiu@uniss.it (P. Ruiu).

https://doi.org/10.1016/j.chbr.2026.101280
Received 26 November 2025; Received in revised form 6 August 2026; Accepted 25 August 2026
Available online 29 August 2026
2451-9588/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

2024ntertainment (Shi et al., 2020; Xie et al., 2024), and biometrics (La Cava et al., 2023; Ruiu et al., 2023).

These innovations have also enabled new lines of research on telepresence in social and media platforms to recreate experiences of physical co-presence across distance, primarily through the use of Mixed Reality (XR) technologies (Fadzli et al., 2023). XR, also known as extended reality, refers to a technology that combines physical and virtual worlds, allowing real and digital objects to coexist and interact in real-time. It relies on elements of both Augmented Reality (AR) and Virtual Reality (VR), enabling users to engage with their physical environment while simultaneously interacting with digital content. Unlike AR, which overlays virtual objects onto the real world, XR allows for more complex interactions where virtual objects can interact with real-world objects and vice versa. This creates immersive experiences where digital content is not only viewed but can be manipulated and integrated within the user's physical space, offering applications across various industries, such as healthcare, education, gaming, and design (Fairchild et al., 2016; Fu et al., 2022; Maas & Hughes, 2020; Seaborn et al., 2016; Workforce & Training and Education Knowledge Management Team, 2025). For live telepresence, depth cameras are used to capture individuals and transmit a real-time stream of their 3D digitization, allowing them to appear and interact within another person's environment (Ishigaki & Ismail, 2022; Ruiu, Mascia, & Grosso, 2024; Yang et al., 2024). This real-time 3D streaming enables immersive and interactive communication experiences, which enhance the sense of presence and co-location across distance (Alexiadis et al., 2013; Orts-Escolano et al., 2016; Rhee et al., 2020). It is projected that multiple such solutions will be developed and commercialized in the coming years (Partarakis & Zabulis, 2024; Suh & Prophet, 2018).

However, despite the recognised transformative potential of such technologies across multiple application domains, empirical studies investigating how end users experience and perceive these technologies for real-time digital communication remain relatively underexplored in the current literature (Suh & Prophet, 2018), despite increasing empirical attention to immersive technologies. In particular, existing evidence (see Section 2) focuses on users' interactions within immersive systems, but does not examine these systems as communication media. Thus, an important open question that warrants investigation is whether the added value offered by immersive experiences for digital communication is truly appreciated by users, especially when weighed against the inconvenience of current hardware requirements (Sai et al., 2024). This is particularly relevant when compared to traditional video-based communication, which, while less immersive, remains highly accessible and familiar. End users can provide more accurate insights into the potential adoption and value of a technology, whereas prototype developers may underestimate its usability or appeal. To address this gap, the present study empirically investigates how users perceive and evaluate immersive telepresence as a medium for real-time digital communication. In this work, the term ‘real-time digital communication’ refers to synchronous, bidirectional interaction between geographically separated users mediated by immersive systems, such as VR conferencing platforms, synchronous social VR environments, or volumetric and holographic telepresence systems.

To the best of our knowledge, this work introduces the first empirical paradigm in which immersive systems function as a medium for real-time, human-to-human communication, mirroring the dynamics of existing digital communication media. In our study, end users interact through a novel, custom-designed system that closely approximates real-world communication. Specifically, our system captures a user's form, reconstructs their volumetric representation, and transmits it to a remote communication partner, who perceives them as spatially co-present through AR headsets in real time. Although at a moderate stage of technological maturity, the designed system provides an ecologically valid, credible approximation of how future immersive communication platforms may operate, instead of relying on hypothetical vignettes or untested prototypes. It enables users to fully experience the communication innovation, allowing authentic interactions and realistic user perceptions to emerge, thereby providing richer, more informative insights into the technology's potential.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature and situates our work within the existing body of research. Section 3 presents the study's aims, research questions, and two thematic clusters on perceptions, value judgments, and psychological constructs related to immersive communication. Section 4 outlines the study's overall methodological approach, detailing how the real-time immersive communication system was implemented, how data were gathered, and the ethical principles guiding the research. In Section 5, we report and synthesize both quantitative and qualitative findings. Finally, Section 6 discusses the implications of the results and their broader significance.

2. Related work

Research on telepresence systems has consistently attracted significant attention, particularly from industry, including major technology companies. In recent years, a growing body of literature has emerged presenting a variety of systems based on diverse technological solutions and methodological approaches.

Early contributions to immersive telepresence primarily focused on technical feasibility and system architecture rather than user-centred empirical validation (Alexiadis et al., 2013). The system proposed by Microsoft in (Orts-Escolano et al., 2016), called Holoportation, is a 3D capture technology that reconstructs high-quality 3D models of people, which are then compressed and transmitted in real-time. Key features include the use of custom cameras for depth and RGB capture, which allows users to interact with remote participants as though they are physically present, with applications in telemedicine and mixed reality environments. Holoportation shares similar basic elements with our solution (i.e. depth cameras, headsets, real-time transmission). However, compared to our system, it relies on expensive custom devices and tailored software that are now outdated. Our experiment, on the other hand, utilizes off-the-shelf low-cost devices, enhancing accessibility for user studies.

In (Lawrence et al., 2024) Google proposed a fully distributed immersive teleconferencing system that aims to replicate face-to-face meetings, ensuring realistic visual and auditory interactions. The system employs structured light technology using infrared cameras and projectors, allowing for high-fidelity 3D reconstruction, motion parallax, and real-time gaze preservation among remote participants, thereby enhancing the immersive experience of remote conferences. Unlike our study, this system does not offer immersive experiences, thereby reducing the level of emotional engagement.

(Gunkel et al., 2021) presents VRComm, which is a web-based system for social communication in virtual reality, featuring real-time photorealistic representations generated from RGB-D video streams. The proposed technology comprises an end-to-end framework that includes data acquisition, transmission via WebRTC, and rendering within virtual environments, supported by a multipoint control unit (MCU). The evaluation encompasses both technical performance metrics, such as latency, CPU/GPU usage, and bandwidth consumption, and realistic user sessions involving participants located in different countries. However, the study does not assess the user's subjective experience in terms of satisfaction, sense of immersion, or co-presence. Instead, the evaluation is primarily focused on the system's technical performance during realistic usage scenarios.

The study in (Balani & Tümler, 2021) examines usability and user experience (UX) across four extended reality (xR) platforms — PC-based VR, HoloLens 2, VR Cardboard, and AR Smartphone — within the context of a biomedical training application. The research employs two standardized instruments to assess UX: the System Usability Scale (SUS) and the User Experience Questionnaire (UEQ). The resulting metrics are derived from the specific dimensions evaluated by these tools.

Compared to our study, this work is primarily a comparative analysis of different technological platforms, rather than an evaluation of a specific technology in relation to user experience. Its focus lies in benchmarking usability and UX across multiple xR solutions, rather than investigating how users perceive and interact with a particular system in depth.

In (Langa et al., 2022), the authors present a user study involving 32 participants engaged in interactive VR sessions to evaluate both technical performance (latency, resource usage, and bandwidth) and subjective experience (using SUS, Social VR, SSQ, NASA-TLX, and interviews). The study focuses on a social VR platform for multi-user holographic meetings (holomeetings), employing real-time point cloud representations captured through low-cost RGB-D cameras. The system comprises acquisition, compression, orchestration, and rendering within shared virtual environments, with the goal of assessing its technological readiness and communicative effectiveness. The study emphasizes system effectiveness and usability in context, rather than exploring how users experience or interpret this specific category of technology. Also in this case, the evaluation primarily focuses on the quality and performance of the implemented solution, rather than on an in-depth investigation of users' perception of the underlying technology itself.

A renewed interest in the Metaverse has stimulated a growing body of empirical work examining immersive technologies across various domains (Xi et al., 2024). conducted a laboratory experiment investigating user acceptance of augmented and virtual reality shopping environments in the retail context. Their study provides valuable empirical insights into how users evaluate immersive shopping experiences. However, the experimental design primarily focused on individual interaction with reconstructed 3D environments and did not explicitly address the social dimension of immersive communication. In other words, participants evaluated spatial and experiential aspects of the environment rather than engaging in real-time social interaction mediated through immersive systems. Similar empirical efforts can be observed in the educational domain (Álvarez et al., 2024). developed and evaluated a virtual reality platform for teacher training, focusing on classroom climate management. Although the system included virtual characters and interactive scenarios, the experience was structured around simulation and training, with a strong gamified component. The emphasis was therefore on pedagogical effectiveness and user acceptance within a controlled instructional setting, rather than on telepresence-oriented, real-time digital communication among human users. Furthermore (Kazakou et al., 2025), examined teachers' acceptance of augmented reality smart glasses using an extended technology acceptance framework. While this study contributes to understanding the adoption dynamics of immersive wearable devices, it investigated a different technological configuration, namely AR smart glasses rather than head-mounted VR systems designed for immersive telepresence. The interaction paradigm and use context thus differ substantially from synchronous, socially mediated immersive communication environments. Taken together, these surveyed studies demonstrate increasing empirical attention to immersive technologies. However, they either focus on individual user interaction within immersive systems rather than social communication or investigate alternative hardware configurations. Empirical research specifically addressing how end users experience and perceive immersive head-mounted systems for real-time digital communication and telepresence remains comparatively limited. The present study aims to address this gap by explicitly examining user experience and acceptance within socially mediated immersive communication contexts.

3. The present study

3.1. Study aims

How users perceive innovations can strongly influence their behaviour, as these perceptions serve as benchmarks for deciding whether to adopt the technology (Technology Acceptance Model, TAM) (Marangunić & Granić, 2015). The TAM model traditionally emphasizes perceived usefulness and ease of use as core factors in explaining computer user behaviour. However, new tools and innovations are not evaluated in a vacuum. For niche innovations like immersive technologies to become mainstream, early adoption depends on users perceiving a clear, non-trivial relative advantage over existing solutions, that is, being perceived as superior to the established practices/tools they aim to replace, in line with the Diffusion of Innovations Theory (DIT) (Rogers et al., 2014). Such perceptions are more meaningfully assessed in concrete, already-established usage contexts where incumbent technologies are well understood and accepted as the state of the art. Video calls are currently the most prevalent and versatile form of virtual communication across everyday social (Bleakley et al., 2022), professional (Lee, 2022) and medical contexts (Razavi et al., 2022). These domains reflect widely adopted, operational real-world conditions in which future immersive systems would also need to operate. As such, they provide appropriate and timely benchmarks for evaluating whether immersive technologies offer sufficient value to motivate early uptake. Therefore, this study frames immersive communication within these domains instead of targeting general, context-free user perceptions. The aim is to capture ecologically valid insights into the potential benefits and adoption considerations within concrete and operationally meaningful usage contexts, rather than evaluating value in principle.

To assess whether users might perceive a relative advantage in immersive communication, this study explores user perceptions across three dimensions: (i) behavioural and experiential outcomes, such as perceived social presence, engagement, naturalness, enjoyment, immersion and connectedness; (ii) overall gain, reflecting participants' perception of added value compared with familiar technologies; and (iii) psychological constructs, like confidence using the technology, its perceived usability and fit. These constructs are specifically operationalized and mapped to validated scales in Section 5, where they generate the outcomes for each of the three dimensions. While the degree of technology adoption in any setting depends on its maturity and fit (Mbatha, 2024), the investigated dimensions represent foundational perceptions that allow the early majority to recognise potential in the technology. In this work, these perceptions are captured through a novel immersive communication system developed specifically for this study, which closely represents the experience that future immersive communication technologies are expected to offer (Section 4.1).

3.2. Research questions

The research questions of this study are organized by thematic cluster. User perceptions were collected while controlling for participants' familiarity with and potential bias toward mainstream digital communication, using video calls as a baseline. This ensures that perceived advantages of immersive communication are evaluated relative to established practices, providing a contextual reference for interpretation rather than being confounded by novelty effects or participants' prior experience with digital communication. Our approach is therefore aligned with the DIT (Rogers et al., 2014), which primarily informed our experimental and system design to capture relative advantage and interpret the innovation's adoption trajectories. These constructs are formulated and operationalized through thematic Cluster A. To investigate the psychological and situational factors shaping adoption, we applied constructs from the Technology Acceptance TAM (Marangunić & Granić, 2015) and extensions (e.g., Unified Theory of Acceptance and Use of Technology, UTAUT (Al-Momani & Ramayah, 2025) in thematic cluster B. Together, these frameworks allow us to examine early exposure effects and provide a structured lens for interpreting participants' responses.

3.2.1. Cluster A: perceptions and value judgements of immersive communication

Immersive communication provides a three-dimensional interactive

experience, making user engagement feel more lifelike and intuitive compared to conventional multimedia, like video calls, which are limited to flat screens and stereo sound (Barabas & Bove, 2013). Immersive communication thus often leads to a heightened sense of presence and user engagement. However, it also introduces challenges, like the need for specialized headsets, potential technical glitches, and a learning curve associated with new technology (Alsamhi et al., 2025). Building on these opportunities and challenges, we aimed to explore how users perceive immersive communication and its overall value as an emerging alternative to mature digital communication media, through the following questions:

RQ-A1: To what extent, if at all, do users perceive added value in immersive calling contextualized against mainstream digital communication?

RQ-A2: How do users perceive the experiential qualities of immersive calls relative to familiar communication methods within established usage contexts?

RQ-A3: How do users evaluate the potential benefits or drawbacks of immersive calls compared to familiar digital communication methods, and how do these evaluations contribute to perceived added value?

3.2.2. Cluster B: psychological constructs affecting adoption and use of immersive communication

Alongside overall perceptions and value judgements, technology adoption is also shaped by psychological and contextual factors (TAM). Consumer technology trends predict if and how people interact with new tools across everyday contexts, displaying a general tendency to adopt tools or technologies similar to those they frequently use and are familiar with (Kopplin, 2022). In particular, individuals with greater experience using emerging technologies are more likely to embrace related innovations due to self-perceived confidence and perceived potential of the technology (Nilashi & Abumalloh, 2024, pp. 1–29). In our study, we investigated whether these effects could be observed for immersive communication through the following research hypothesis:

RH-B1: Individuals with prior experience with immersive technologies and/or higher self-reported confidence using digital tools will demonstrate greater willingness to adopt immersive communication across various contexts.

However, adoption is highly situational, meaning it is shaped by the specific environments and contexts in which the technology is used (Lemay et al., 2019). Despite the transformative potential of immersive technology across sectors (Haleem et al., 2022), even a technically superior system may face resistance from users with low acceptance, hindering its uptake (Rouidi et al., 2022). Different drivers affect acceptance or adoption in social, professional and clinical contexts, with computerization in health being one of the most challenging (Shvetsov & Alsamhi, 2024). Thus, this cluster explored the following research questions:

RQ-B1: What factors influence users' preferences for or against the adoption of immersive communication for social, professional and healthcare use, and how do these factors differ between contexts?

Cluster B, thus, attempts to operationalize dimensions of extended TAM (Song et al., 2025) through an empirical perspective on factors reflecting the complexity of adoption decisions of modern innovations, beyond simple usability. Both clusters employed a mixed-methods approach, corroborating quantitative questionnaire data with qualitative open ended participant responses.

4. Methodology

4.1. Implementation of real-time immersive communication system

The system developed for this experiment is composed of a network of off-the-shelf devices, which are integrated to perform specific functions without relying on expensive custom hardware or proprietary software. The main features include scalability, ease of use, and affordability, making it accessible for a wide range of applications while maintaining effective performance. As shown in Fig. 1, the system consists of three main components: (i) the HoloLens 2 headset, which is used to visualize the rendered 3D representation of the person in the real environment of the user, (ii) the Unity application, which manages the shared environment and synchronizes the data across the network, and (iii) the Azure Kinect device, which is used to capture data from the remote user for 3D reconstruction in the virtual environment.

4.1.1. Materials and methods

HoloLens 2 is an advanced mixed reality headset developed by Microsoft that combines augmented reality (AR) with spatial computing, enabling users to interact with 3D objects in their physical environment (Balakrishnan & Guo, 2024; Park et al., 2021). Unlike traditional virtual reality (VR) headsets HoloLens 2 uses special waveguide lenses to project volumetric representations directly into the user's field of view. These lenses are designed to allow light from the 3D renderings to pass through while maintaining transparency, creating a seamless blend between the real world and virtual elements. This unique configuration helps reduce motion sickness, which is often a problem with other headsets. In contrast to other mixed reality or virtual reality devices, which rely on external cameras to capture the surrounding environment and overlay virtual objects onto the video feed, HoloLens 2 projects immersive 3D models directly into the user's line of sight. This direct projection reduces the disconnect between the real and virtual worlds, offering a more natural and comfortable experience that minimizes the risk of discomfort often associated with video-based augmented reality. Another of its key strengths is the ability to track hand gestures, voice commands, and eye movements, enabling intuitive control without the need for external peripherals.

More recent immersive devices do not use holographic displays, such as those of Hololens2, instead, they reconstruct the surrounding environment through video passthrough and overlay 3D elements within that mediated view. However, since the proposed system has been developed using Unity, it can be easily adapted to any headset compatible with the libraries employed during development. This ensures a high degree of flexibility and portability across a wide range of immersive devices, facilitating future extensions and deployments.

Azure Kinect is a depth sensing and motion capture device developed by Microsoft, which integrates advanced sensors and cameras to capture high-resolution depth, color, and infrared data, facilitating interactive and immersive experiences.

The system is also compatible with other similar RGB-D devices, such as Intel RealSense depth cameras.[1] This compatibility enhances the system's versatility, allowing it to be deployed with a variety of depth-sensing hardware beyond the Kinect originally used in development.

The software developed for the experiment was built using the **Unity development platform**.[2] Unity is a cross-platform game engine and development environment used for creating interactive 2D, 3D, AR and VR applications. It includes a physics engine, rendering capabilities, scripting in C#.

Various libraries and tools were employed to develop the system, necessary for creating the shared mixed reality environment, as shown in Fig. 2. Specifically, the following tools and Software Development Kits (SDKs) were utilized:

- XR Interaction Toolkit and Mixed Reality Toolkit, frameworks provided by Unity and Microsoft, respectively, to facilitate the creation of XR or VR environments.
- Netcode for GameObject (NGO), a framework used to manage the sharing mechanisms within the developed application.

[1] https://www.intelrealsense.com/.
[2] https://unity.com/.

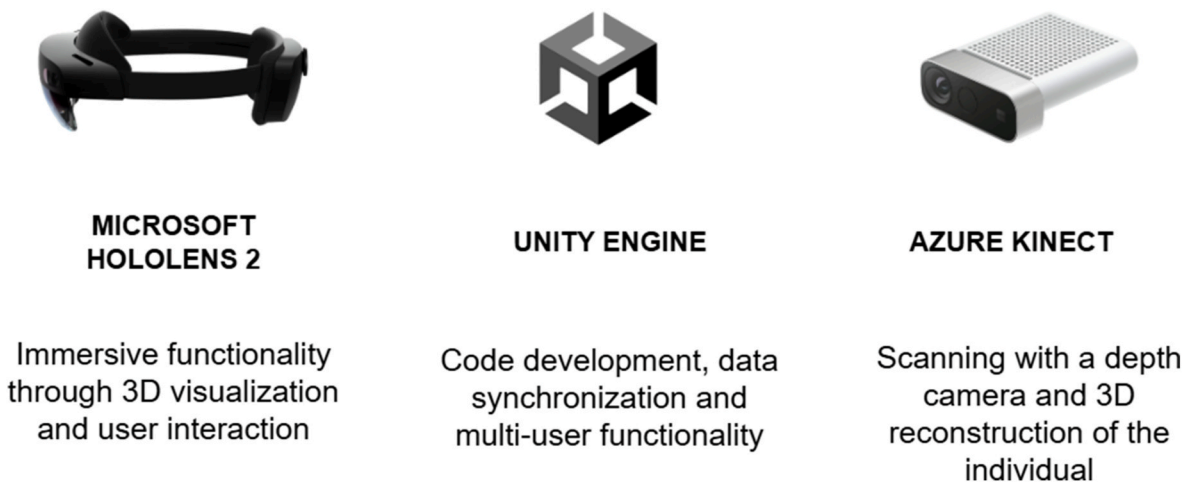


Fig. 1. Principal components of the Mixed-Reality Telepresence Systems.

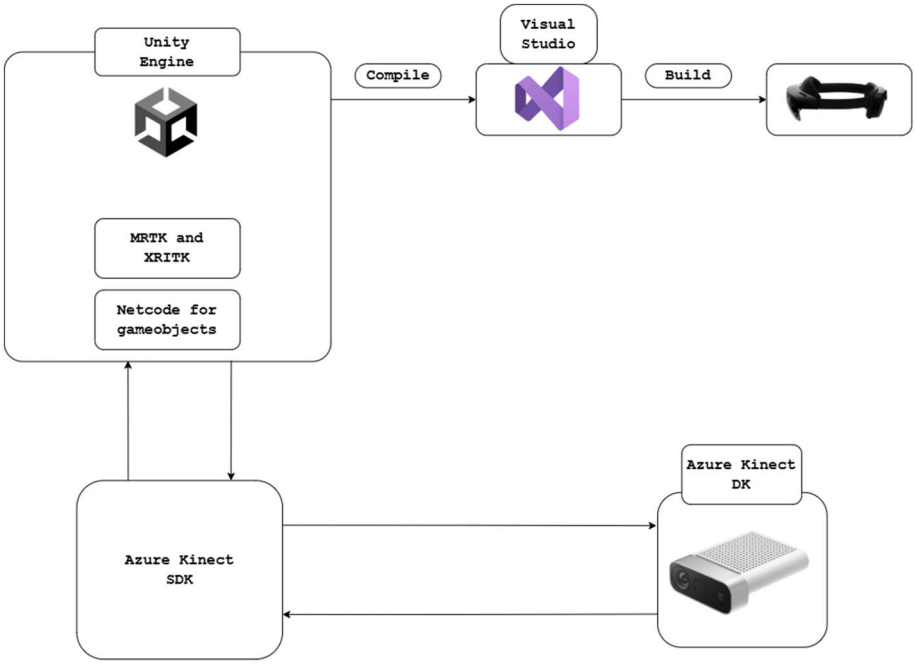


Fig. 2. Architectures and main components of the Mixed-Reality Telepresence Systems, with real-time acquisition and 3D Reconstruction.

- Azure Kinect DK, a device developed by Microsoft, used to perform real-time scans of physical objects, enabling their sharing with various users on the network.
- Azure Kinect SDK, the Software Development Kit used to access the functionalities of the Azure Kinect and implement them within Unity.

Once the software is developed using Unity and a build is made for the Universal Windows Platform (UWP), the application can then be imported onto the HoloLens 2 via Visual Studio.

4.1.2. Real-time 3D data acquisition and streaming

Our telepresence application allows for real-time scanning of an individual, enabling the 3D model to be transmitted to various clients on the network. It is important to note that, in the developed software, only the network host, equipped with a Kinect, is able to perform this scan.

Scanning with the Azure Kinect results in a point cloud, which, to be manipulated in the virtual environment within Unity, must be converted into a Mesh. A Mesh is a component provided by Unity that consists of various properties (vertices, UVs, colors, indices, and triangles) used to create solid objects.

To optimize performance and prevent client disconnections, two different transmission modes were tested: *performance mode* and *quality mode*, as shown in Fig. 3. Both modes operate at a reduced resolution, specifically with a depth resolution of 320x288. Quality mode (Fig. 3a) captures data at a lower rate of 5 frames per second. This mode prioritizes a higher level of detail in the rendered model. In contrast, in performance mode (Fig. 3b), half of the vertices are discarded to reduce the size of the Mesh, and the frame rate is set to 10 frames per second. This lower capture rate helps to decrease the filling time of the Reliable Queue, thereby reducing the risk of congestion and improving system responsiveness. Furthermore, the materials used to render the point cloud differ between the two modes. As detailed in the following chapter, the shader employed in performance mode renders larger triangles compared to the shader used in quality mode. This adjustment enhances the visualization of the scanned model in quality mode,



Fig. 3. Difference between real-time scanning in the two transmission modes tested for the experiment, highlighting variations in vertex resolution and rendering detail.

offering a clearer and more detailed representation, which is particularly beneficial when high visual fidelity is required.

4.2. Study design

This study employed a combined between-subject and within-subject experimental design to measure perceptions of the immersive call (experimental condition). We included a video call in the design as a control condition, which serves as a well-established baseline for measuring improvements or advances in communication technologies. Following the Media Richness Theory ((Daft & Lengel, 1986)), this design enables a systematic evaluation of whether the added sensory and spatial cues of real-time immersive interactions, where the communication partner is perceived as a real, co-present individual rather than an avatar, enhance social presence, engagement, and user experience, potentially reducing the cognitive load by supporting more natural multimodal communication (Steed et al., 2016, pp. 67–76). Moreover, our design seeks to leverage comparative judgment (Marangunić & Granić, 2015) when forming user perceptions about the technology to help reduce recall bias and strengthen causal interpretation by isolating the effect of the medium. Importantly, each participant experiences both the video and immersive calls, thus serving as their own control, which may reduce intercept variability and strengthen statistical power (section 4.2.2).

4.2.1. Setting and sampling

The experiment took place in the Cognitive Robotics and Autonomous Systems Laboratory at the School of Computing, University of Kent, UK. 50 participants aged 18-65 (mean = 38.6, SD = 12.2), involving 11 full-time students and 39 staff, were drawn from a university environment primarily for practical and methodological reasons.

University students and staff represent a diverse group of everyday technology users, spanning social, professional and medical contexts due to their routine use of IT tools for work, study and communication. Recruiting from this population allowed for controlled, reliable access to participants familiar with digital communication technologies, facilitating timely data collection and ensuring participants could competently interact with both video and immersive communication systems. To assemble an appropriate exploratory sample with potential relevance to the wider population, we employed purposive sampling to capture diversity in age, sex, educational attainment, occupation, professional roles and digital proficiency. This heterogeneity ensures that user perceptions are not biased toward a single demographic, allows demographic factors to be analysed as potential moderators, and supports tentative generalization to digitally literate working-age adults. Snowball sampling was used to recruit participants who were already acquainted (friends, family, colleagues), encouraging more natural interactions during the study. The demographic profile of the final sample is summarized in Table 1. All demographic data were collected in ranges to prevent identification.

Before the experiment, participants were required to provide informed consent. They were assigned unique and paired participant IDs to ensure anonymity. They were also informed of their right to withdraw at any point until the end of the experiment, after which data

Table 1
Participant demographics.

Demographic variable	Category	n
Sex	Male	29
	Female	21
Age (years)	18–24	9
	25–34	14
	35–54	18
	55–65	9
	Mean (SD)	38.6 (12.2)
Education	GCSE or equivalent	8
	College/Sixth Form	4
	Undergraduate Degree	13
	Master's Degree	14
	PhD	11
Role/Occupation	Professional services	32
	Academic staff (all levels)	7
	Students	11

withdrawal would not be possible due to anonymization.

4.2.2. Protocol

The experiment involved pairs of acquainted participants, situated in separate rooms and engaging in real time communication through our immersive system. Before the experiment, participants completed a questionnaire gathering demographic information and prior experience with technology (Questionnaire 1, see Appendix 1). System startup (video, immersive), headsets, cameras, and network connections were pre-configured by the experimenters. Participants were only responsible for donning the headset, adjusting the visor for comfort, performing any necessary eye calibration and using basic interface controls, such as volume or hand-tracking gestures. This procedure aimed to optimize their visual experience and comfort during the interaction and minimize learning effects associated with system novelty. It also sought to reduce potential biases due to familiarity differences, as our university-based participants had extensive experience with video conferencing, but were generally unfamiliar with immersive technologies. After initial setup, participants were offered a beverage and left to interact independently and comfortably, without experimenter intervention. No audio or video recordings were collected.

The participants engaged in two sequential, remote, asymmetrical communication interactions. In the first interaction session, Participant 1 was randomly assigned to the control condition and Participant 2 to the experimental condition. In the control condition, Participant 1 communicated using a tablet device and was simultaneously captured using a Kinect camera, which reconstructed their volumetric representation in real time (Fig. 4, left). This 3D reconstruction was transmitted to Participant 2 in the experimental condition, who viewed Participant 1 as a life-sized, spatially anchored 3D model through a Microsoft HoloLens augmented reality headset (Fig. 4, right). Participant 2's image was captured by a second Kinect camera and transmitted to Participant 1 as a standard 2D video displayed on the tablet screen. This asymmetrical setup allowed us to obtain between-subject evaluations of immersive calls, which were collected immediately after the interaction using a post-condition questionnaire (Questionnaire 2, see Appendix 2). Participants then swapped roles, so each could experience both conditions, and engaged in another interaction session. After completing both sessions, participants filled in a final questionnaire capturing comparative, within-subject perceptions of the two communication modalities across the entire sample (Questionnaire 3, see Appendix 3), which helps reduce variability from individual differences and increases the sensitivity of the statistical tests to detect meaningful effects.

To mitigate mental fatigue (Dallaway et al., 2022) and cybersickness (E. . and I. S. B, 2020), each interaction session was limited to a maximum of 10 min. Order effects related to learning and fatigue were further controlled by counterbalancing the sequence of exposure to conditions across participant pairs. While we cannot completely rule out

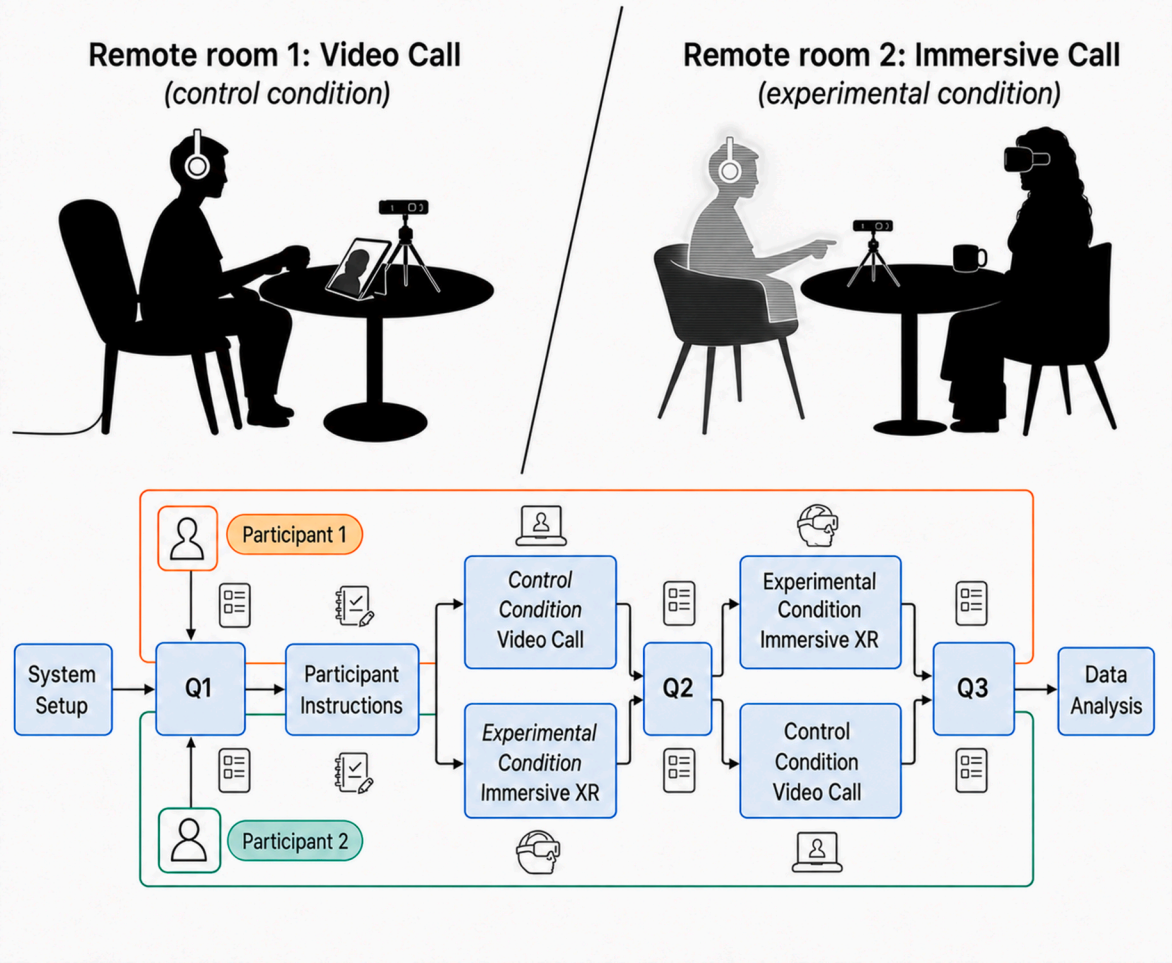


Fig. 4. The experimental setup. Participant in remote room 1 (*control condition*) communicates via a traditional video call with the participant in remote room 2 (*experimental condition*). The latter sees the participant in room 1 as a life-size 3D model through the augmented reality headset (a Microsoft HoloLens 2). Two Kinect Azure cameras are used, one to capture RGB and depth data for reconstructing the volumetric representation and the other to transmit video data to the tablet. The participants engage in a natural, non-scripted conversation while consuming a beverage to create a more naturalistic interaction. They swap roles so that both experience the technology under similar conditions and to eliminate confounding variables.

subtle perceptual influences, counterbalancing the sequence of conditions across participants helps mitigate systematic bias, ensuring that any first-exposure effects are not consistently confounded with a particular condition.

We deliberately asked participants to engage in unconstrained, unrecorded and spontaneous conversation while sharing a beverage (coffee, tea, water) to maintain ecological validity. Although our study aims to explore the relative advantage of immersive communication across social, professional and medical contexts, we avoided operationalizing them explicitly, as this would require extensive time and task-specific training and could introduce confounds related to task complexity (e.g., cognitive load, stress, unfamiliar procedures). Moreover, given that we recruited lay users of digital communication with prior acquaintance, scripted or artificial scenarios would have been poorly aligned with their profile and risked forcing unnatural interactions. Instead, our experimental design aimed to create a relaxed and relatable environment in which they could naturally talk and observe realistic gestures and interaction patterns under controlled conditions. We anticipated that fundamental interaction dynamics, such as real-time communication, gaze, gesture, turn-taking, non-verbal cues and call fidelity, would emerge organically and provide generalizable insights that are informative for multiple real-world domains.

4.2.3. Ethics statement

This study was approved by the Central Research Ethics Advisory Group (CREAG) of the University of Kent (Approval number: CREAG052-04-24) on 24 May 2024. All participants provided informed consent before participating in the study, in accordance with the Declaration of Helsinki.

4.3. Questionnaires and data collection

Three questionnaires were administered to participants at different stages of the experiment, as described in section 4.2.2 and the flowchart in Fig. 4. These served different purposes. Questionnaire 1 (see Appendix 1) was administered before the interaction sessions to collect demographics and prior experience with technology. Demographic data from questionnaire 1 are summarized in Table 1 and support the analysis and interpretation of later results. Explicit definitions of augmented reality, virtual reality and technological advancements were provided to ensure a consistent understanding of these terms. These items helped characterize the sample, their self-assessment and prior exposure to technology, to inform interpretation of usability, experiential and adoption outcomes. Questionnaire 2 (see Appendix 2) was administered immediately after the first interaction, in which participants experienced only one communication medium. This questionnaire aimed to capture between-subject perceptions of the technology experienced during that phase. As described earlier, including the video call in the interaction was intended to provide participants with a direct reference for comparison, instead of relying solely on their general familiarity with video calls. Depending on the assigned condition, participants answered the same set of items, with wording adapted to refer either to the video call (control condition) or immersive call (experimental condition) to ensure that the measured constructs were directly comparable across conditions. Specific items from this questionnaire were used in different analyses reported in the Results section and are explicitly indicated when each analysis is described to clarify which items contributed to which outcome measures. Questionnaire 3 (see Appendix 3) was administered after participants swapped roles and completed both interaction conditions to capture within-subject comparative perceptions across video and immersive calls. These items were intended to provide a robust complement to the between-subject data collected in Questionnaire 2 once participants experienced both technologies. Items in questionnaires 1-3 were either adapted from formal, validated questionnaires or designed *ex novo* for this study. The rationale for each item and its intended purpose are explained in the Results section when describing the corresponding analyses. Where items were adapted from validated questionnaires, their psychometric properties are established in the original sources. Instead, items designed ex-novo were created to capture exploratory, context-specific perceptions and experiential judgements relevant to immersive communication interactions, emphasizing face validity rather than formal scale validation.

5. Findings

In this section, we report quantitative and qualitative findings by thematic clusters. The analysis was conducted in Python and R.

5.1. Cluster A: perceptions and value judgements of immersive communication

Participant perceptions were measured with eight items on a 7-point Likert scale, where 1 indicated strong disagreement and 7 indicated strong agreement. These items are listed in Table 2 and correspond to questions 1 to 8 in Questionnaire 2 (see Appendix 2) administered post-condition.

The items are the same for both conditions, with wording adapted to the specific call, hence indicated by […] in each item, i.e., […] refers to either 'immersive' for the participant in the experimental condition or 'video' for the participant in the control condition. The items were adapted from existing questionnaires, which provide validated constructs for behavioural and experiential perceptions and overall gain. Adaptations were made to achieve the specific aims of this study when not covered in a standard form. This adaptation ensures face validity for our experimental scenario, while retaining comparability to established scales. Items listed in Table 2 were selected because they directly assess participants' experiential enjoyment and functional advantage of immersive calls relative to familiar methods, allowing us to address our research questions. Questions/items 1, 5, and 7 capture subjective experiential qualities (RQ-A1), whereas questions 2, 3, 4, 6, and 8 reflect the participant's perceived ability to act successfully in the system (RQ-A2), which is predictive of behaviour. We note that usability in Table 2 is

Table 2
Questionnaire 2 items mapped to validated constructs and existing questionnaires (see also Appendix 2).

Questionnaire items	Mapped constructs	Adapted from
Q1. I enjoyed the […] call	User enjoyment Intrinsic motivation	Technology Acceptance Model (TAM)User Experience Questionnaire (UEQ)
Q2. The […] was easy to use	Perceived ease of use Usability	TAM System Usability Scale (SUS) UEQ
Q3. I was able to use the […] without help	Independence Self-efficacy	Computer Self-Efficacy Scales (CSES) SUS (reverse-coded), UEQ
Q4. I felt confident in my ability to make the […] call	Confidence Self-efficacy	TAM, CSES, SUS
Q5. The experience of sharing a beverage with my friend during the […] call was natural	Social presence Naturalness	Social Presence Questionnaire (SPQ) ITU-T P.85 standard
Q6. The […] call provides me with useful options and choices	Perceived usefulness	TAM, UEQ
Q7. The […] call made me feel connected to the other person	Social presence Connectedness	SPQ Networked Minds Social Presence Inventory (NMSPI)
Q8. The […] was fit for purpose (i.e., capable of performing the social interaction task)	Fit for use	UEQ ISO 9241-210 standard

defined in line with ISO 9241-11 as the extent to which a system can be used to achieve specific goals with effectiveness, efficiency and satisfaction. We operationalized it through items adapted from the System Usability Scale (SUS), which captures participants' perceived ease of use and overall interaction experience, including confidence, independence and self-efficacy. In this study, usability therefore refers specifically to participants' experience during the interaction task itself, but not their ability to set up the immersive call. As described in section 4.2.2, all system hardware, software and network connections were pre-configured to isolate usability perceptions of the interaction and reduce potential biases from differences in familiarity, cognitive load or lack of exposure to immersive hardware. Internal consistency was examined for the questionnaire items. As the items map onto distinct constructs (Table 2) rather than a single latent scale, reliability was assessed at the subscale level. The resulting Cronbach's α values were: user experience/technology acceptance (Q1, Q6, Q8), $\alpha = 0.70$ (video) and 0.66 (immersive); usability (Q2–Q4), $\alpha = 0.86$ (video) and 0.81 (immersive); and social presence/connectedness (Q5, Q7), $\alpha = 0.62$ (video) and 0.72 (immersive). These values should be read with caution as the subscales are computed on a small number of items per construct and the resulting reliability estimates may be less stable than those obtained with larger numbers of items.

The following analyses are based on 25 data points and reflect the between-subject outcomes evaluated on the same set of Questionnaire 2 items using both descriptive and inferential statistics. For each participant, a composite score was calculated as the mean across items Q1-Q8 on the 1–7 Likert scale. Their combined effect is presented in the boxplots in Fig. 5. An independent Welch's *t*-test was conducted to compare overall evaluations between the conditions. The results indicated that, when participants were exposed to only one technology at a time (video call or immersive call), they rated the video call ($M = 5.93$, $SD = 0.58$) more favourably than the immersive call ($M = 5.12$, $SD = 0.98$) overall, with the difference reaching statistical significance ($p = 0.0015$, $t(48) = 3.43$). Item-level analyses further revealed that ratings for the video call were higher for Q2 ($p = 0.002$), Q3 ($p < 0.001$), Q4 ($p < 0.001$), and Q8 ($p = 0.004$), but not significant on the remaining items. This indicated that participants perceived higher ease of use, independence in operation, confidence and overall fit for purpose for video calls.

This outcome is unsurprising, as technologies at the ideation stage often exhibit greater variability in user uptake and usability perceptions than mature systems (Schwabe et al., 2021). In fact, our results revealed a high variability between the intercepts for the immersive call ($SD = 0.98$). Due to the observed variability, we used responses from Questionnaire 3, administered after participants experienced both technologies, to corroborate our findings with within-subject insights.

From this questionnaire, we used a set of binary-choice items corresponding to questions 1 and 2 (see Appendix 3), which aimed to provide a direct comparison between video and immersive calls. The results draw on the entire sample (50 data points), providing a robust complement to the between-subject analysis, capturing individual differences and reducing potential bias from small and partial sampling. The items were designed ex-novo for this study, based on common components of user experience in immersive environments (Bareišyte et al., 2024; Cheiran et al., 2025).

Items used from Questionnaire 3 (Appendix 3):

1. Which type of call did you find more: enjoyable; easier to use; natural; convenient; fit for use; aesthetically appealing.
2. Which type of call makes you feel more: comfortable; confident to use; connected to others; offering useful options and choices; engaged or immersed.

To quantify participants' preferences for one technology after their exposure to both, a diverging bar was generated (Fig. 6), which illustrates the preference difference between immersive and video calls (i.e., difference in the number of participants) across all dimensions measured by the items. To assess whether these differences were statistically meaningful, chi-square and binomial tests were performed against the equal-probability null that participants are equally likely to prefer either type of call. This analysis determines which technology participants consistently preferred, controlling for the baseline expectation of no preference and is used to address research questions RQ-A1 and RQ-A2.

The results confirmed that participants consistently rated immersive calls higher for enjoyment and immersion, whereas video calls were rated higher for ease of use, confidence, convenience, and task suitability. No significant differences were found for aesthetic appeal, naturalness and offering useful options. Combining the between-subject and within-subject evaluations comprehensively confirmed that once participants had a baseline for comparison, they perceived immersive calls as more engaging and enjoyable, highlighting an experiential advantage (RQ-A2). Video calls continued to outperform in usability, confidence and functional adequacy, commonly associated with familiar technologies. Given that video calls represent a mature and widely adopted communication standard, the absence of significant differences

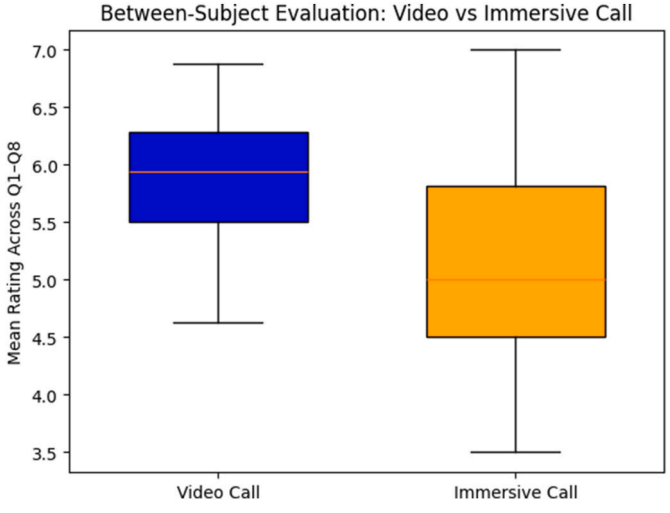


Fig. 5. Participant scores for each call type, showing the combined effect of all items and the standard deviation. Video calls received generally higher ratings, with a median of around 6 and an IQR (interquartile range) between about 5 and 7. Instead, immersive calls ranked lower, with a median of around 5, but with greater variability of the intercepts.

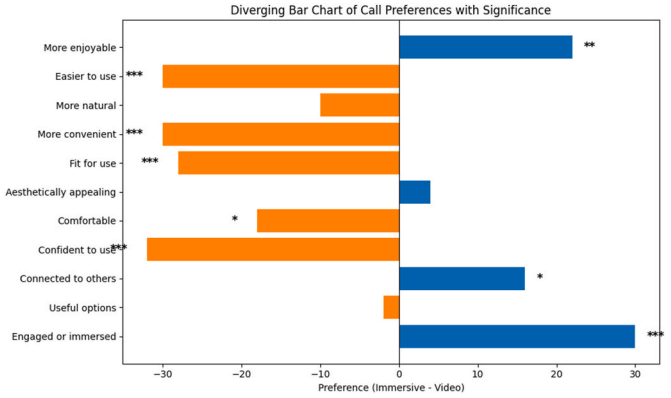


Fig. 6. The diverging bar illustrates the magnitude and direction of preference, along with its statistical significance. The magnitude represents the number of participants who preferred the type of call over the other. Rightwards direction indicates preference for the immersive call, whereas leftwards direction for the video call. Significance is indicated as: *** highly significant ($p < 0.001$), ** significant ($p < 0.01$), and * moderately significant ($p < 0.05$).

between the technologies across several dimensions suggests a degree of performance convergence, even at this early stage of technological development (RQ-A1).

Finally, we corroborated quantitative ratings with a qualitative analysis derived from participants' written responses. Across the administered questionnaires, participants were invited to briefly justify their preferences for each technology when rating the items listed previously (Appendices 2 and 3). These responses were analysed using a word cloud method (Fig. 7) to identify the dominant vocabulary associated with immersive and video calls.

In support of the quantitative outcomes on overall perceived value (RQ-A1) and experiential differences between the technologies (RQ-A2), this analysis also confirmed that video calls were generally associated with ease of use, reliability, confidence, and comfort, whereas immersive calls were characterized as natural, immersive, and fostering a stronger sense of connection. The qualitative vocabulary distributions therefore reinforce the experiential and usability distinctions observed in the statistical analysis.

In addition, participants provided extended written feedback at the conclusion of the study (Questionnaire 3, Appendix 3). This richer dataset was analysed using thematic analysis to address RQ-A3. The thematic map in Fig. 8 summarises the systematic coding of participants' responses to identify recurring patterns of perceived potential and limitations of immersive technology.

Positive responses consistently highlighted the technology's immersive quality, sense of realism and capacity to foster social connectedness, as clear experiential advantages over conventional video calls. At the same time, participants identified practical limitations associated with the current level of technology readiness, such as audio-visual quality, comfort and usability challenges, particularly related to headset wear, motion sickness and restricted mobility. A few participants questioned whether the immersive system currently offers sufficient added value compared with existing communication tools, while most affirmed that it holds substantial promise once it reaches adequate maturity (RQ-A3).

Collectively, the qualitative analyses deepen the interpretation of RQ-A3 and contextualize the quantitative results for RQ-A1 and RQ-A2, demonstrating that the perceived trade-off between the experiential value gain of immersive calls and their current limitations is primarily technological rather than conceptual.

5.2. Cluster B: psychological constructs affecting adoption and use of immersive communication

To achieve our research aims in cluster B, we first investigate context-sensitive indicators of user preference formation (sections 5.2.1 and 5.2.2 The contextual domain (social, professional, medical) is treated as a situational factor, in line with the Diffusion of Innovation Theory (Rogers et al., 2014). As argued previously, these contexts were purposefully selected as they represent a range of real-world scenarios in which communication technologies are commonly deployed. We operationalize adoption preference across these contexts using behavioural intention (i.e., willingness to use) as a proximal indicator and acceptability as a normative judgement, consistent with the TAM and UTAUT models. These measures capture expressed preference for everyday use, collected in a scenario simulating casual social interaction. This approach provides a controlled and realistic setting for capturing participants' authentic behaviours and true perceptions of the technology's value, rather than the idiosyncrasies of the task design. The factors influencing those preferences are subsequently analysed via selected variables, such as prior experience with technology, general attitudes towards technology and self-perception about technology, drawn as predictors from TAM constructs and our adapted questionnaires (section 5.2.3).

5.2.1. Willingness to use immersive calls

We designed two items ex-novo to measure willingness to use on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). These items map onto behavioural intention constructs of the TAM and UTAUT frameworks. They were administered post-condition and correspond to questions 9 and 10 of Questionnaire 2 (see Appendix 2). Each participant completed the items for the technology to which they were assigned, either video calls (control) or immersive calls (experimental), therefore providing between-subject measures.

Video calls (affirmative) for participants in the control condition:

1. I use and will continue to use video calls to communicate with friends and family.
2. I use and will continue to use video calls in my workplace for virtual meetings.

Immersive calls (conditional), for participants in the control condition:

1. In the future, if the technology were perfect and readily available, I would use immersive calls to communicate with friends and family.
2. In the future, if the technology were perfect and readily available, I would use immersive calls in my workplace for virtual meetings.

We conducted two-way ANOVA and Ordinary Least Squares (OLS) regression analyses. The goal of this evaluation was not to compare the two technologies directly. Rather, video calls served as a baseline condition to measure effects in the experimental condition. Our primary objective was to assess how participants' responses approached or exceeded the outcomes associated with the baseline, despite immersive technology being an emerging and experimental medium.

The two-way ANOVA tested the main effects of call type (video, immersive) and context (social, professional), as well as their interaction. Call type and context were treated as independent variables, whereas the single dependent variable was the mean score of the two willingness items (one observation per participant). The results



Fig. 7. Word cloud analysis illustrating the dominant vocabulary used by participants to characterize each technology. This analysis reinforces the quantitative findings reflecting higher ratings for immersive calling in terms of immersion and realism, whereas usability, comfort and confidence scored higher for video calls.

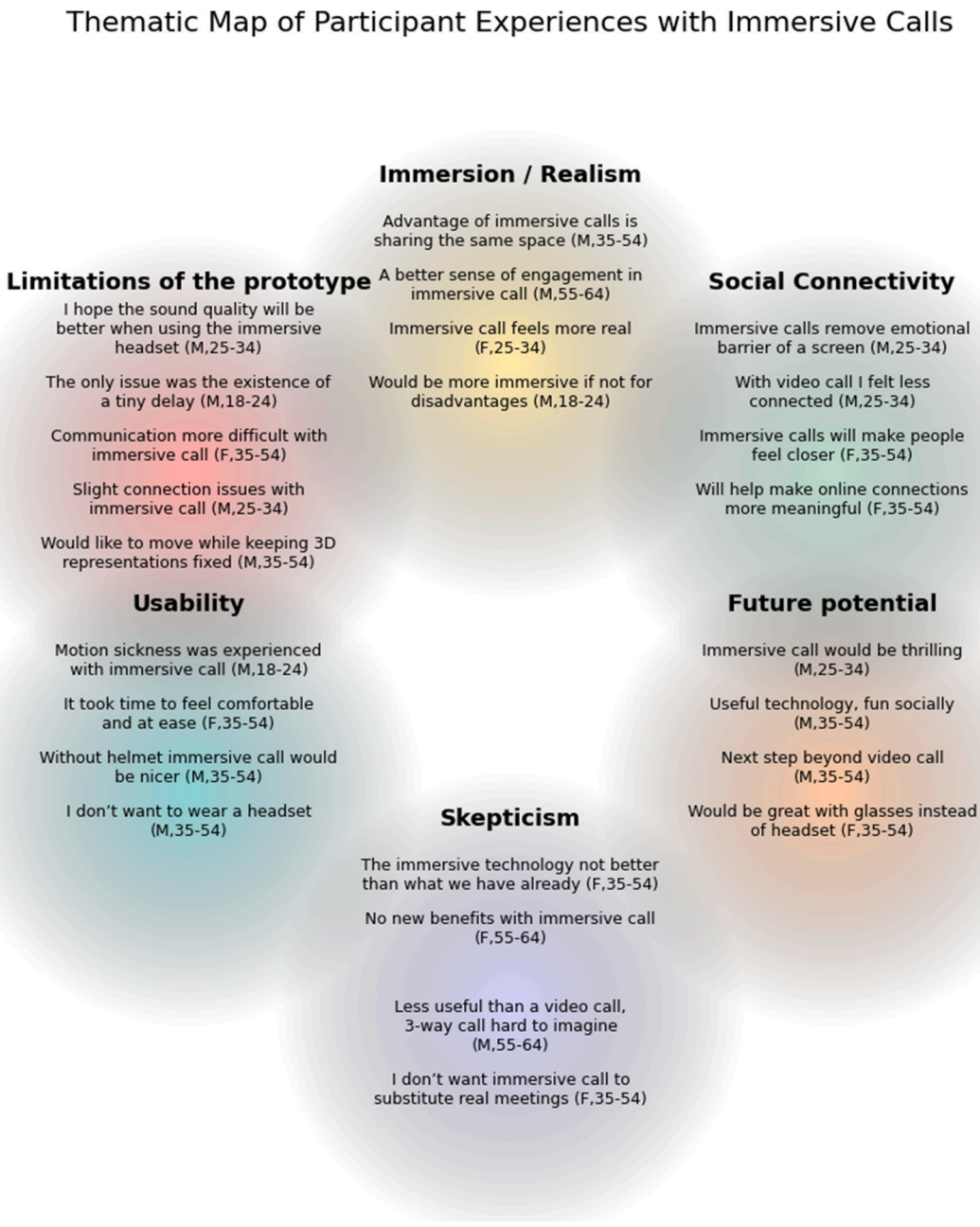


Fig. 8. Thematic map of participant experiences with immersive calls. The map visualises the key themes emerging from participant feedback at the conclusion of the study. Quotes inside the theme bubbles highlight representative participant experiences. We note that the limitations of the prototype capture technical constraints and issues specific to our designed system, whereas the other themes reflect general attitudes about the technology. In particular, social connectivity summarises feedback about how the system influenced feelings of closeness and meaningful interaction. However, it is reasonable to assume that the prototype's limitations may have had a cascading effect on the other themes, both directly (e.g., usability) and indirectly (e.g., scepticism).

indicated a marginal trend toward higher ratings for video calls, which approached but did not reach standard significance ($F(1,93) = 3.58$, $p = 0.061$). No significant effect in context ($F(1,93) = 0.31$, $p = 0.581$) or interaction ($F(1,93) = 0.51$, $p = 0.477$) emerged. The OLS regression test also confirmed that video calls scored only negligibly higher (+0.47, p-value = 0.061) than immersive calls across social and professional contexts. Descriptive statistics confirmed that ratings of both calls across the items were comparable and above average, with the difference being non-significant (video calls – Q1: $M = 6.48$; Q2: $M = 6.20$; immersive calls – Q1: $M = 5.64$; Q2: $M = 5.68$). Taken together, the near parity of evaluations between video and immersive communication suggests that participants' future adoption intentions for immersive communication are both realistic and substantiated.

At the conclusion of the study, we asked participants to express their willingness to use each technology after having experienced both. The following binary item was administered to the entire sample (corresponding to question 3 in Questionnaire 3, see Appendix 3), providing within-subject data on participants' willingness to use immersive versus video calls across contexts.

Item: Assuming that either technology is perfect and readily available, which type of call are you willing to use more in the following contexts?

1. Social contexts (e.g., interact with friends and family);
2. *Professional contexts* (e.g., virtual meetings and networking);
3. *Healthcare contexts* (e.g., doctor-patient remote appointments).

We conducted pairwise cross-tabulations and applied chi-square goodness-of-fit tests to determine whether the observed distributions differed from the null hypothesis of no preference. (Table 3). When assessed independently (between-subject, two-way ANOVA), immersive communication did not exhibit a willingness gap relative to the established baseline (video call). The results in Table 3, however, indicate that this parity did not persist under direct comparison. When participants evaluated the calls comparatively (within-subject), their preferences shifted primarily in favour of immersive calls. Specifically, the preference difference was statistically significant in social and healthcare contexts (i.e., higher preference for immersive calls), whereas professional use showed no meaningful difference. In professional contexts, both video and immersive calls were considered equally suitable, which is a noteworthy result given that video calls are already deeply embedded in workplace practices. Participants also provided explanations for their choices (question 3 in Questionnaire 3, see Appendix 3). Among the listed reasons, they recognised a clear promise of immersive communication for social contexts, stating that immersive, lifelike, three-dimensional interaction could enhance experiences such as virtual gatherings, gaming, or long-distance communications. They noted that the technology's ability to recreate a stronger sense of co-presence may fundamentally reshape how people connect remotely. A particularly interesting finding was the greater preference for immersive calls in healthcare, where participants cited benefits such as spatial awareness and hands-free 3D interaction, which they believed could support telemedicine, surgical guidance and medical training.

5.2.2. Acceptability of immersive calls: social, professional, healthcare

We designed a composite item ex-novo to measure participants' acceptability on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The item was administered post-condition and corresponds to question 11 in Questionnaire 2 (see Appendix 2). Each participant provided ratings for the three contexts for the technology to which they were assigned, yielding between-subject evaluations.

Video call (affirmative) for participants in the control condition:

It is acceptable to use video calls in many areas of life:

1. *Social contexts* (e.g., interact with friends and family);
2. *Professional contexts*(e.g., virtual meetings and networking);
3. *Healthcare contexts*(e.g., doctor-patient remote appointments).

Immersive call (conditional) for participants in the experimental condition

In the future, if the technology were perfect and readily available, it is acceptable for immersive calls to become applicable in many areas of life:

1. *Social contexts* (e.g., interact with friends and family);
2. *Professional contexts* (e.g., virtual meetings and networking);
3. *Healthcare contexts* (e.g., doctor-patient remote appointments).

We performed a two-way mixed-design (split-plot) ANOVA on participants' acceptability ratings. The dependent variable was the acceptability rating (1–7 Likert scale) for each context, log-transformed to approximate a normal distribution of residuals for ANOVA assumptions. Call type (video, immersive) was treated as a between-subjects factor, since each participant experienced only one technology, whereas context (social, professional, healthcare) was treated as a within-subjects factor, because each participant rated all three contexts. The ANOVA thus tested whether perceived acceptability differed by call type, by context and for the interaction between the two. The analysis revealed no significant main effect of call type ($F(1, 48) = 0.045$, $p = 0.833$, $np2 = 0.0009$), i.e., no significant difference in overall acceptability ratings between the calls. This indicated no statistically significant difference in overall acceptability between immersive and video calls. However, acceptability varied significantly by context ($F(2, 96) = 11.21$, $p < 0.001$, $np2 = 0.189$). The effect of context remained significant even after adjusting for the Greenhouse-Geisser correction ($p_GG_corr = 0.000243$) due to sphericity violation ($W = 0.661$, $p < 0.001$). Nevertheless, the interaction effect (call type x context) was not significant ($F(2, 96) = 0.536$, $p = 0.587$, $np2 = 0.011$), which indicated that the effect of context on acceptability occurred for both video calls and immersive calls. To explain the significant main effect of context, we conducted a post-hoc Tukey's Honestly Significant Difference (HSD) test, corroborated with descriptive statistics. The pairwise comparisons indicated that participants' ratings for social and professional contexts did not differ significantly from each other. Both contexts were rated highly. For social contexts, mean ratings were video: $M = 6.28$, $SD = 0.74$ and immersive: $M = 6.40$, $SD = 0.69$, while for professional contexts, video: $M = 6.40$, $SD = 0.71$ and immersive: $M = 6.16$, $SD = 0.85$. In contrast, healthcare was rated significantly less acceptable than both social and professional contexts, regardless of the type of call, with mean ratings, video: $M = 5.56$, $SD = 1.23$ and immersive: $M = 5.64$, $SD = 1.32$. These findings suggest that immersive calls are perceived as acceptable and comparably viable to mature systems for social and professional communication. However, both established and emerging technologies appear to lack sufficient acceptability in healthcare contexts, highlighting a persistent caution and slower adoption of digitalisation of health-related interactions.

Table 3
Chi-square test results for different use cases.

Use case	XR	Video	Chi-square Result
Social	34	16	$\chi^2(1, N = 50) = 6.48, p = 0.011$
Professional	27	23	$\chi^2(1, N = 50) = 0.32, p = 0.572$
Healthcare	34	16	$\chi^2(1, N = 50) = 6.48, p = 0.011$

5.2.3. Adoption moderators

After assessing participants' willingness to use immersive calls and their general acceptability, we sought to synthesize the underlying factors shaping those preferences and interpret them within the selected social, professional and healthcare domains. The following three-point polarity item was administered in Questionnaire 3, corresponding to question 4 (see Appendix 3).

Item: Do you consider upgrading from video calls to immersive calls a) positive; b) negative; c) neutral. The collected within-subject evaluations indicated a generally favourable response: 39 participants (78%) responded positively, 7 (14%) negatively and 4 (8%) were neutral. We examined whether this adoption tendency could be explained or predicted by participants' previous responses about themselves and technology in general. We conducted an ordinal logistic regression including the three-point polarity item (Negative = 0, Neutral = 1, Positive = 2) as the dependent variable. Seven items from Questionnaire 1 (see Appendix 1, questions 1 with 4 sub-items, and questions 3 - 5) were used as predictors: previous experience with AR (Yes/No), previous experience with VR (Yes/No), confidence in the ability to use technology, asking for help when using technology, perceived usefulness of technology, excitement about technological innovations and feeling connected through technology. The latter two items reflect attitudinal or intrinsic motivators (extensions of TAM). The model was fitted using an Ordered Model with a logit link function, which estimated thresholds automatically as Negative → Neutral (0/1): 7.245, SE = 5.239, p = 0.167 and Neutral → Positive (1/2): 0.017, SE = 0.476, p = 0.971. The model converged successfully after 49 iterations with 50 observations, producing an Akaike Information Criterion (AIC) of 59.81 and a Bayesian Information Criterion (BIC) of 77.02, indicating a reasonable fit given the sample size. The overall model was statistically significant (Likelihood Ratio Test, $\chi 2 = 25.299$, df = 7, p = 0.0007), indicating that the set of predictors collectively explains the participants' choice better than a null model. The model demonstrated a moderate explanatory power (McFadden pseudo-$R2 = 0.547$). The results indicated that prior

experience with VR ($\beta = 3.95$, p = 0.015) and excitement about technological innovations ($\beta = 1.30$, p = 0.031) were the strongest predictors of adopting immersive calls. Instead, AR experience did not reach conventional significance ($\beta = -3.48$, p = 0.065). Other predictors, including perceived self-confidence in using technology, usefulness, connectedness and asking for help, were not statistically significant for adoption in this sample. These findings are consistent with the extended TAM model. A behavioural segmentation analysis (Fig. 9) further illustrates the effect of familiarity with AR/VR on participants' willingness to adopt immersive calls, complementing the statistical results and partially supporting hypothesis RH-B1. These results align with prior evidence that emerging technologies are evaluated in the context of use-case relevance and familiarity with related tools (Nilashi & Abumalloh, 2024, pp. 1–29).

Beyond the predictors identified with the statistical analysis, the broader mixed-methods analysis was used to synthesize the contextual factors shaping these preferences across social, professional, and healthcare settings. The combined outcomes in cluster B reveal that adoption patterns are, indeed, context-sensitive. In social contexts, adoption is driven primarily by experiential and hedonic benefits, such as lifelike interaction, the sense of shared space, engagement and fun. The participants' behavioural intention and acceptability are high, and moderators, like prior experience with emerging technology, self-perception or general positive attitudes towards technological innovations, are more likely to boost adoption. In professional contexts, adoption is shaped by instrumental and practical considerations, such as stability, efficiency, and compatibility with existing solutions. Behavioural intention is moderate, but acceptability is high, with dominant factors affecting adoption being typically workflow integration. In healthcare contexts, adoption depends on trust, fidelity and regulatory compliance, in addition to usability and professional risk aversion. The behavioural intention is conditional, due to higher perceived risk, and acceptability is lower than for other contexts due to technical and regulatory constraints. These patterns suggest that extended TAM measures must be interpreted with situational moderators in mind, rather than assuming uniform effects across domains.

6. Discussion

This study investigated user perceptions of immersive communication as a potential successor to existing media. The study is the first to implement a fully functional, real-time interactive system in which users experience life-sized, spatially anchored immersive communication mirroring the dynamics of existing digital communication media. This design maximizes ecological validity by closely approximating how immersive interactions are expected to function in the future. We explored the innovation's perceived advantage and early adoption trends across contexts where the technology is likely to operate, such as social, professional and medical domains. These were examined using a mixed methods approach, which provides an additional layer of value by capturing experiential, behavioural, and psychological dimensions of user engagement with the technology.

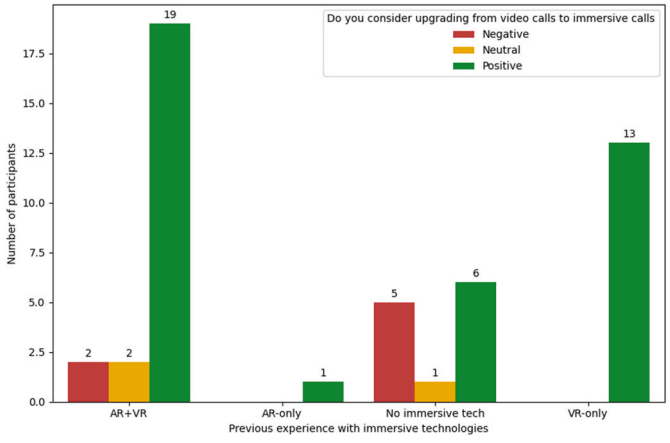


Fig. 9. Behavioural segmentation analysis indicating how participants' previous experience with AR/VR technologies mediated their willingness to displace video calls for immersive calls. This descriptive analysis complements the outcomes of ordinal logistic regression, which identified VR experience as the most significant predictor of adoption of immersive calls.

Our results indicated that, overall, participants recognised immersive calls as offering a distinctly richer and more engaging experience than video calls, particularly in social interactions. The sense of shared space and realism was frequently described as a major advantage, producing high levels of immersion and enjoyment. This aligns with other work showing that virtually immersive environments can enhance co-presence and social interaction through spatial communication and embodied representation, to a level comparable to face-to-face interaction (Streuber et al., 2026). It is worth noting that the sample included a balanced distribution of participants with and without prior experience using AR/VR technologies, indicating that the reported effects are unlikely due to novelty alone, but potentially reflect the genuine experiential value of immersive communication. Nevertheless, concerns were raised about usability and social practicality. Challenges, such as headset dependency and comfort, motion sickness, and occasional audio-visual desynchronisation, were noted, which remain well-documented issues in immersive systems to date (Shen et al., 2023). This is to be expected at the current level of technology readiness, making such limitations a key barrier to adoption rather than the technology's conceptual feasibility, consistent with the early-stage dynamics described in Diffusion of Innovations (Rogers et al., 2014).

In this study, we purposefully asked participants to communicate naturally and comfortably while sharing a beverage remotely, rather than simulating task-specific or role-based interactions across the chosen contexts (social, professional, and medical). This was intended to avoid introducing task-related covariates and cognitive load that could confound user perceptions, and to allow pairs of friends, colleagues, or family members to interact naturally. Naturalistic interaction paradigms, rather than artificial or scripted interactions, have frequently been recommended in user studies to capture authentic social dynamics and user behaviour (Subramanian et al., 2023). Our focus was therefore on preserving ecological validity and observing how realistic reconstruction and social presence emerged under authentic conversational conditions, rather than simulating artificial relationships between acquainted participants, like patient-doctor interactions. In line with the recommendations, this design approach likely accentuated the social affordances of the immersive call by reducing the cognitive burden or pressure associated with the experimental setting. While social presence was not formally measured in this study, it emerged naturally during participants' interactions and is illustrated through qualitative responses in the thematic map, particularly in the themes of social connectivity and immersion/realism. By contrast, in professional contexts, where participants are accustomed to mature and highly effective video platforms, this design revealed cautious interest. On one hand, participants recognised the added value of immersive calls over video calls for collaborative and spatial tasks; on the other hand, they also indicated that this novelty is currently insufficient to outweigh practical considerations, such as the stability and integrated features of established video platforms. Similarly, in healthcare, the approach allowed participants to express conditional acceptance of immersive communication, emphasizing the importance of fidelity, reliability, and regulatory compliance. Clinical adoption of emerging technologies does not only depend on perceived benefits but also on trust, usability and compatibility with existing workflows (Olaiya et al., 2026). Nevertheless, participants perceived the greatest overall gain of immersive communication in clinical contexts, recognizing it as a more promising alternative to in-person consultations than other technologies under

ideal conditions, despite current acceptability being lower than for social or professional uses. This reflects a theoretical prediction of (Marangunić & Granić, 2015) that anticipated usefulness in critical contexts, such as healthcare, can sustain positive adoption intention even when current systems are not yet mature (Ibrahim et al., 2025).

Our mixed-method analysis revealed that participants attributed high usability and confidence to video calls, but associated immersive calls with better realism, engagement and enjoyment. Importantly, once participants experienced both technologies, their overall willingness to adopt immersive calls approached parity with video calls, which is an important insight considering that the technologies are at opposite ends of technological readiness. This also likely suggests that early user exposure can meaningfully shift perceptions of feasibility and value, aligning with the principles of Diffusion of Innovation theory (Rogers et al., 2014), according to which relative advantage becomes apparent once users experience the innovation's distinct affordances. Nevertheless, the results must be interpreted as exploratory. The study was based on a modest sample size (N = 50), combining both between- and within-subject designs, and participants were drawn from a demographic aged 18-64 based at a University environment, though spanning across different employment statuses, digital literacy, and educational attainment. This may not fully represent the broader population, limiting generalisability. However, the participants reflect characteristics typical of early adopters who engage frequently with many forms of technology. This makes the sample suitable for exploratory investigation and for capturing informed user perceptions of the innovation at an early stage of conception (Cho et al., 2023). In this study, participants experienced a prototype rather than a commercially mature system. This also means that any observed results may have been influenced by temporary technical imperfections rather than the innovation's full potential.

Our findings also raise important directions for future design and implementation. For example, for use at work, our participants expected the integration of immersive communication with existing workflows and features familiar to professional users, such as screen sharing, file collaboration, or meeting management tools. They noted that adoption would be more likely if system stability were improved, latency reduced, and the interface made more intuitive, for instance, by removing reliance on headsets. Participants proposed that a possible transition could first be realised through a hybrid system, where video and immersive modes can co-exist flexibly. Immersive features could be initially embedded for tasks that benefit most from spatial awareness, co-presence and collaborative visualization, while maintaining the convenience of video calls for routine interactions. In healthcare, they stressed that future designs must prioritize high fidelity in both audio and visual channels, ensure strict compliance with privacy and security regulations and support hands-free or minimally obtrusive interactions to reduce cognitive and physical burden on both patients and clinicians, as well as ensuring they do not disrupt existing clinical practices. More broadly, the results suggest that immersive communication may initially follow a hybrid trajectory, with participants indicating a willingness to use immersive calls alongside existing media rather than entirely displacing them. This pattern is consistent with the Media Synchronicity Theory (Dennis & ROBERT, 2009, pp. 261–287), which proposes that different communication media are suitable depending on task requirements and with recent research on hybrid communication ecosystems that highlight how new and established technology (ICT)-mediated interactions can coexist and complement one another in organizational and social contexts (Torro et al., 2022).

Future work should first focus on replicating the study with larger and more diverse samples, along with improved software quality, to determine whether the observed perceptual patterns hold in longitudinal and ecologically valid contexts. It is important to note that our findings emerged from task-free conversational interactions and do not capture the constraints of specific domains, as structured domain-specific interactions were not explicitly simulated or evaluated in this study. As such, the results provide baseline insights into immersive communication, with further studies needed to extend these findings to specific contextual workflows. Further to our experimental design, we acknowledge that the participant in the control condition could see the partner wearing a headset, which does not reflect natural video-based communication. However, given our participants' extensive prior experience with video calls, it is unlikely that this single encounter meaningfully affected their perception of the video interaction. Future work could explore alternative configurations to further isolate the perceptual effects of immersive hardware within the technology's current hardware constraints. Newer versions of the prototype could also include hybrid interaction modalities, like integrated video and immersive features. Further insights into the inclusivity and equitable accessibility of both hardware and software could be gained by investigating populations that are less familiar with emerging technologies, such as older adults or individuals with low digital confidence and literacy. It is clear that, as with any technology, long-term widespread adoption will depend on achieving both technical maturity and social trust. Lightweight, affordable hardware, like emerging lightweight smart glasses to replace headsets, combined with transparent and user-friendly interfaces, will be essential for mainstream acceptance. Immersive communication, when fully realised, has the potential to transform not only how people connect remotely, but also how they experience presence, empathy, and collaboration across distance.

CRediT authorship contribution statement

Pietro Ruiu: Writing – review & editing, Writing – original draft, Software, Methodology, Funding acquisition, Conceptualization. **Ioanna Giorgi:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Filippo Casu:** Writing – review & editing, Software, Methodology. **Lorenzo Mascia:** Software, Data curation. **Chloe Bates:** Data curation. **Enrico Grosso:** Supervision, Funding acquisition, Conceptualization. **Giovanni Luca Masala:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Ethics declaration

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place. This study was approved by the Central Research Ethics Advisory Group University of Kent Canterbury. (Approval No. CREAG052-04-24)

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Enrico Grosso reports financial support was provided by Italian Ministry of University and Research (MUR). Enrico Grosso reports financial support was provided by European Commission. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been developed within the framework of the project e. INS- Ecosystem of Innovation for Next Generation Sardinia (cod. ECS 00000038) funded by the Italian Ministry for Research and Education (MUR) under the National Recovery and Resilience Plan (NRRP) - MISSION 4 COMPONENT 2, “From research to business” INVESTMENT 1.5, “Creation and strengthening of Ecosystems of innovation” and

construction of “Territorial R&D Leaders”. - CUP: J83C21000320007.

Moreover, it has received financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union —NextGenerationEU— Project Title “METATwin— Metaverse & Human Digital Twin: digital identity, Biometrics and Privacy in the future virtual worlds”, – CUP: J53D23015030001- Grant Assignment Decree No. 0001382 adopted on 01/09/2023 by the Italian Ministry of University and Research (MUR).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2026.101280.

Data availability

The data that has been used is confidential.

References

Al-Momani, A., & Ramayah, T. (2025). The utaut model in understanding ehr adoption: A systematic review. *Intelligence-Driven Circular Economy: Regeneration Towards Sustainability and Social Responsibility, 2*, 179–194.

Alexiadis, D. S., Zarpalas, D., & Daras, P. (2013). Real-time, realistic full-body 3d reconstruction and texture mapping from multiple kinects. In *IVMSP 2013* (pp. 1–4). IEEE.

Alsamhi, S. H., Nashwan, F., & Shvetsov, A. V. (2025). Transforming digital interaction: Integrating immersive holographic communication and metaverse for enhanced immersive experiences. *Computers in Human Behavior Reports, 18*, Article 100605.

Álvarez, I. M., Manero, B., Romero-Hernández, A., Cárdenas, M., & Masó, I. (2024). Virtual reality platform for teacher training on classroom climate management: Evaluating user acceptance. *Virtual Reality, 28*, 78.

Balakrishnan, P., & Guo, H.-J. (2024). Hololens 2 technical evaluation as mixed reality guide. In *International conference on human-computer interaction* (pp. 145–165). Springer.

Balani, M. S., & Tümler, J. (2021). Usability and user experience of interactions on vr-pc, hololens 2, vr cardboard and ar smartphone in a biomedical application. In *International conference on human-computer interaction* (pp. 275–287). Springer.

Barabas, J., & Bove, V. M. (2013). Visual perception and holographic displays. In *Journal of physics: Conference series, 415*. IOP Publishing, Article 012056.

Bareišyte, L., Slatman, S., Austin, J., Rosema, M., van Sintemaartensdijk, I., Watson, S., & Bode, C. (2024). Questionnaires for evaluating virtual reality: A˙systematic scoping review. *Computers in Human Behavior Reports, 16*, Article 100505.

Bleakley, A., Rough, D., Edwards, J., Doyle, P., Dumbleton, O., Clark, L., Rintel, S., Wade, V., & Cowan, B. R. (2022). Bridging social distance during social distancing: Exploring social talk and remote collegiality in video conferencing. *Human-Computer Interaction, 37*, 404–432.

Cheiran, J. F. P., Bandeira, D. R., & Pimenta, M. S. (2025). Measuring the key components of the user experience in immersive virtual reality environments. *Frontiers in Virtual Reality, 6*, Article 1585614.

Cho, J., DeStefano, T., Kim, H., Kim, I., & Paik, J. H. (2023). What's driving the diffusion of next-generation digital technologies? *Technovation, 119*, Article 102477.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*, 554–571.

Dallaway, N., Lucas, S. J. E., & Ring, C. (2022). Cognitive tasks elicit mental fatigue and impair subsequent physical task endurance: Effects of task duration and type. *Psychophysiology, 59*, Article e14126.

Dennis, A. R., & Robert, M. F. (2009). *Media synchronicity and media choice: Choosing media for performance*. Media Choice, Routledge.

E. . I. S. B. (2020). *Department of business, the safety of domestic virtual reality systems: A literature review*.

Fadzli, F. E., Ismail, A. W., & Abd Karim Ishigaki, S. (2023). A systematic literature review: Real-time 3d reconstruction method for telepresence system. *PLoS One, 18*, Article e0287155.

Fairchild, A. J., Campion, S. P., García, A. S., Wolff, R., Fernando, T., & Roberts, D. J. (2016). A mixed reality telepresence system for collaborative space operation. *IEEE Transactions on Circuits and Systems for Video Technology, 27*, 814–827.

Fu, Y., Hu, Y., & Sundstedt, V. (2022). A systematic literature review of virtual, augmented, and mixed reality game applications in healthcare. *ACM Transactions on Computing for Healthcare (HEALTH), 3*, 1–27.

Gunkel, S. N., Hindriks, R., Assal, K. M. E., Stokking, H. M., Dijkstra-Soudarissanane, S., Haar, F.t., & Niamut, O. (2021). Vrcomm: An end-to-end web system for real-time photorealistic social vr communication. In *Proceedings of the 12th ACM multimedia systems conference* (pp. 65–79).

Haleem, A., Javaid, M., Singh, R. P., Suman, R., & Rab, S. (2022). Holography and its applications for industry 4.0: An overview. *Internet of Things and Cyber-Physical Systems, 2*, 42–48.

Ibrahim, F., Münscher, J.-C., Daseking, M., & Telle, N.-T. (2025). The technology acceptance model and adopter type analysis in the context of artificial intelligence. *Frontiers in Artificial Intelligence, 7*, Article 1496518.

Ishigaki, S. A. K., & Ismail, A. W. (2022). Real-time 3d reconstruction for mixed reality telepresence using multiple depth sensors. In *International conference on advanced communication and intelligent systems* (pp. 67–80). Springer.

Kazakou, G., Koutromanos, G., & Christogiannis, C. (2025). Examining teachers' acceptance of augmented reality smart glasses through the immersive technologies acceptance model. *Education and Information Technologies, 30*, 26329–26372.

Khirodkar, R., Bagautdinov, T., Martinez, J., Zhaoen, S., James, A., Selednik, P., Anderson, S., & Saito, S. (2024). Sapiens: Foundation for human vision models. In *European conference on computer vision* (pp. 206–228). Springer.

Kopplin, C. S. (2022). Chatbots in the workplace: A technology acceptance study applying uses and gratifications in coworking spaces. *Journal of Organizational Computing and Electronic Commerce, 32*, 232–257.

La Cava, S. M., Orrù, G., Drahansky, M., Marcialis, G. L., & Roli, F. (2023). 3d face reconstruction: The road to forensics. *ACM Computing Surveys, 56*, 1–38.

Langa, S. F., Montagud, M., Cernigliaro, G., & Rivera, D. R. (2022). Multiparty holomeetings: Toward a new era of low-cost volumetric holographic meetings in virtual reality. *IEEE Access, 10*, 81856–81876.

Lawrence, J., Overbeck, R., Prives, T., Fortes, T., Roth, N., & Newman, B. (2024). Project starline: A high-fidelity telepresence system. In *ACM SIGGRAPH 2024 emerging technologies* (pp. 1–2).

Lee, C. S. (2022). Analyzing zoombombing as a new communication tool of cyberhate in the covid-19 era. *Online Information Review, 46*, 147–163.

Lemay, D. J., Doleck, T., & Bazelais, P. (2019). Context and technology use: Opportunities and challenges of the situated perspective in technology acceptance research. *British Journal of Educational Technology, 50*, 2450–2465.

Lo, S., Fowers, S., Darko, K., Spina, T., Graham, C., Britto, A., Rose, A., Tittsworth, D., McIntyre, A., O'Dowd, C., et al. (2023). Participatory development of a 3d telemedicine system during covid: The future of remote consultations. *Journal of Plastic, Reconstructive & Aesthetic Surgery, 87*, 479–490.

Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., & Sheikh, Y. (2019). Neural volumes: Learning dynamic renderable volumes from images. *ACM Transactions on Graphics, 38*, 1–14.

Maas, M. J., & Hughes, J. M. (2020). Virtual, augmented and mixed reality in k–12 education: A review of the literature. *Technology, Pedagogy and Education, 29*, 231–249.

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society, 14*, 81–95.

Mbatha, B. (2024). Diffusion of innovations: How adoption of new technology spreads in society. In *Information, knowledge, and technology for teaching and research in Africa: Human machine interaction and user interfaces* (pp. 1–18). Springer.

Nilashi, M., & Abumalloh, R. A. (2024). *i-tam: A model for immersive technology acceptance, education and information technologies*.

Olaiya, O., Chan, G., & Orji, R. (2026). Application of immersive technologies and machine learning in diagnosis and treatment: A systematic review. *International Journal of Human-Computer Interaction*, 1–28.

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., et al. (2016). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 741–754).

Park, S., Bokijonov, S., & Choi, Y. (2021). Review of microsoft hololens applications over the past five years. *Applied Sciences, 11*, 7259.

Partarakis, N., & Zabulis, X. (2024). A review of immersive technologies, knowledge representation, and ai for human-centered digital experiences. *Electronics, 13*, 269.

Razavi, S., Farrokhnia, N., & Davoody, N. (2022). Nurses' experience of using video consultation in a digital care setting and its impact on their workflow and communication. *PLoS One, 17*, Article e0264876.

Rhee, T., Thompson, S., Medeiros, D., Dos Anjos, R., & Chalmers, A. (2020). Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE Transactions on Visualization and Computer Graphics, 26*, 1923–1933.

Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of innovations. In *An integrated approach to communication theory and research* (pp. 432–448). Routledge.

Rouidi, M., Hamdoune, A., Choujtani, K., Chati, A., et al. (2022). Tam-utaut and the acceptance of remote healthcare technologies by healthcare professionals: A systematic review. *Informatics in Medicine Unlocked, 32*, Article 101008.

Ruiu, P., Cadoni, M. I., Lagorio, A., Nixon, S., Casu, F., Farina, M., Fadda, M., Trunfio, G. A., Tistarelli, M., & Grosso, E. (2025). *Uniss-mdf: A multidimensional face dataset for assessing face analysis on the move, computer vision and image understanding*, Article 104384.

Ruiu, P., Lagorio, A., Cadoni, M., & Grosso, E. (2023). Enhancing eid card mobile-based authentication through 3d facial reconstruction. *Journal of Information Security and Applications, 77*, Article 103577.

Ruiu, P., Mascia, L., & Grosso, E. (2024). Saliency-guided point cloud compression for 3d live reconstruction. *Multimodal Technologies and Interaction, 8*, 36.

Ruiu, P., Nitti, M., Pilloni, V., Cadoni, M., Grosso, E., & Fadda, M. (2024). Metaverse & human digital twin: Digital identity, biometrics, and privacy in the future virtual worlds. *Multimodal Technologies and Interaction, 8*, 48.

Sai, S., Garg, A., & Chamola, V. (2024). *Navigating the metaverse: A comprehensive analysis of consumer electronics prospects and challenges, ACM transactions on internet technology*.

Saito, S., Simon, T., Saragih, J., & Joo, H. (2020). Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 84–93).

Schwabe, O., Bilge, P., Hoessler, A., Tunc, T., Gaspar, D., Price, N., Sharir, L., Pasher, E., Erkoyuncu, J., De Almeida, N. M., et al. (2021). A maturity model for rapid diffusion of innovation in high value manufacturing. *Procedia CIRP, 96*, 195–200.

Seaborn, K., Edey, J., Dolinar, G., Whitfield, M., Gardner, P., Branje, C., & Fels, D. I. (2016). Accessible play in everyday spaces: Mixed reality gaming for adult powered chair users. *ACM Transactions on Computer-Human Interaction, 23*, 1–28.

Shen, X., Gao, J., Li, M., Zhou, C., Hu, S., He, M., & Zhuang, W. (2023). Toward immersive communications in 6g. *Frontiers of Computer Science, 4*, Article 1068478.

Shi, M., Aberman, K., Aristidou, A., Komura, T., Lischinski, D., Cohen-Or, D., & Chen, B. (2020). Motionet: 3d human motion reconstruction from monocular video with skeleton consistency. *ACM Transactions on Graphics, 40*, 1–15.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London and New York: Wiley.

Shvetsov, A. V., & Alsamhi, S. H. (2024). When holographic communication meets metaverse: Applications, challenges and future trends. *IEEE Access*, (12), 197488–197515.

Song, Y., Qiu, X., & Liu, J. (2025). The impact of artificial intelligence adoption on organizational decision-making: An empirical study based on the technology acceptance model in business management. *Systems, 13*, 683.

Steed, A., Pan, Y., Zisch, F., & Steptoe, W. (2016). The impact of a self-avatar on cognitive load in immersive virtual reality. *IEEE virtual reality (VR)*. IEEE, 2016.

Stieger, S., Lewetz, D., & Willinger, U. (2023). Face-to-face more important than digital communication for mental health during the pandemic. *Scientific Reports, 13*, 8022.

Streuber, S., Rogula, S., Quirós-Ramírez, M. A., & Pruessner, J. (2026). Remote collaboration in virtual reality induces physiological synchrony comparable to face-to-face interaction. *Scientific Reports, 16*, 3721.

Subramanian, S., De Moor, K., Fiedler, M., Koniuch, K., & Janowski, L. (2023). Towards enhancing ecological validity in user studies: A systematic review of guidelines and implications for qoe research. *Quality and User Experience, 8*, 6.

Suh, A., & Prophet, J. (2018). The state of immersive technology research: A literature analysis. *Computers in Human Behavior, 86*, 77–90.

Torro, O., Pirkkalainen, H., & Li, H. (2022). Media synchronicity in organizational social exchange. *Information Technology & People People, 35*, 162–180. Author et al.: Preprint submitted to Elsevier Page 24 of 24.

Wang, E., Wang, L., Xu, Q., & Lin, Y. (2024). Three-dimensional human body reconstruction based on volumetric rendering learning. In *Proceedings of the 2024 3rd Asia conference on algorithms, computing and machine learning, CACML '24* (pp. 437–442). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3654823.3654909.

Workforce, Training and Education Knowledge Management Team. (2025). NHS England, evidence brief: Virtual, augmented, and extended reality. *Evidence Brief Summarizing Research and Reports on Workforce-Related Applications of VR, AR, and XR Technologies*. http://www.ewin.nhs.uk/tools_and_resources/evidence-briefs.

Xi, N., Chen, J., Gama, F., Korkeila, H., & Hamari, J. (2024). Acceptance of the metaverse: A laboratory experiment on augmented and virtual reality shopping. *Internet Research, 34*, 82–117.

Xie, R., Huang, K., Cho, I.-Y., Yang, S., Chen, W., Bao, H., Zheng, W., Li, R., & Huo, Y. (2024). Ren human: Learning relightable neural implicit surfaces for animatable human rendering. *ACM Transactions on Graphics, 43*.

Yang, Y., Feng, Q., Lai, Y.-K., & Li, K. (2024). Real-time 3d human reconstruction and rendering system from a single rgb camera. In *SIGGRAPH Asia 2024 technical communications*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3681758.3697993.