International Journal of Sustainable Applied Science and Engineering (IJSASE)

Contents lists available at IJSASE
International Journal of Sustainable Applied Science and Engineering
journal homepage: IJSASE
Volume 3, No. 1, 2026

Design and Implementation of a Cloud-Based Multi-User Virtual Reality Conferencing System with Natural Human–Computer Interaction

Ehsan Akhlaghi [a], Arman Kavoosi Ghafi [b]

[a] *Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran,*
[b] *Department of Computer Engineering,Bo.C. Islamic Azad University, Borujerd, Iran.*

ARTICLE INFO

Received: 2026/05/22

Revised: 2026/06/15

Accept: 2026/07/22

Keywords:

Virtual Reality, Multi-user Collaboration, Virtual Conferencing, Natural Interaction, Cloud Networking, Unity3D, Human–Computer Interaction, Avatar Synchronization.

ABSTRACT

Virtual reality (VR) has emerged as a promising technology for enhancing remote collaboration by providing immersive three-dimensional environments that improve users' sense of presence and social interaction. However, conventional video conferencing platforms remain limited in supporting natural communication, real-time avatar synchronization, and intuitive interaction among geographically distributed participants. This study presents the design and implementation of a cloud-based multi-user virtual reality conferencing system that integrates immersive visualization, real-time networking, natural hand tracking, and avatar-based communication into a unified collaborative platform. The proposed system was developed using the Unity game engine and a cloud networking architecture to enable simultaneous participation of multiple users within a shared virtual environment. Natural interaction was achieved through head-mounted displays (HMDs), controller-based hand tracking, synchronized avatar movements, and real-time voice communication. A complete system architecture covering avatar generation, networking infrastructure, interaction management, and collaborative virtual environments was designed and implemented. The proposed framework was experimentally evaluated through usability studies involving eighteen participants and compared with conventional video conferencing platforms. Quantitative statistical analysis demonstrated significantly higher user satisfaction, perceived presence, usability, and collaborative interaction within the proposed VR conferencing environment. The obtained results indicate that immersive visualization combined with natural interaction substantially improves communication quality and collaborative engagement compared with traditional two-dimensional conferencing systems. The proposed framework provides a practical foundation for next-generation immersive collaboration environments and can be extended to education, healthcare, industrial training, remote teamwork, and intelligent collaborative metaverse applications.

[a] Corresponding author email address: e.akhlaghi@alumni.sbu.ac.ir (Ehsan Akhlaghi)
DOI: https://doi.org/10.22034/ijsase.v3i1.216
Available online 2026/07/22

3060-6349/BGSA Ltd.

1. Introduction

In recent years, remote collaboration technologies have become an essential component of communication across education, industry, healthcare, engineering, and scientific research. Video conferencing platforms have significantly reduced geographical barriers by enabling users in different locations to communicate without the need for physical travel. Despite their widespread adoption, conventional video conferencing systems remain constrained by two-dimensional interfaces, limited interaction capabilities, restricted non-verbal communication, and a relatively low sense of presence, all of which can reduce user engagement and the realism of collaborative experiences (Steed & Oliveira, 2010; Slater & Sanchez-Vives, 2016; Karl et al., 2021; Correia et al., 2020).

Virtual Reality (VR) has emerged as one of the most influential technologies for creating immersive digital environments that support natural and effective human interaction. Unlike traditional multimedia systems, VR enables users to experience computer-generated environments through first-person interaction, stereoscopic visualization, head tracking, and natural body movements. These characteristics substantially enhance immersion and presence, allowing users to perceive themselves as physically located within a shared virtual environment rather than merely observing it through a conventional display (Slater, 2009; Bowman & McMahan, 2007; Sherman & Craig, 2019). Consequently, VR has found widespread applications in education, engineering, healthcare, military training, collaborative design, industrial simulation, and remote communication (Radianti et al., 2020; Berg & Vance, 2017).

The rapid development of the Metaverse concept and immersive collaborative technologies has further accelerated research on multi-user virtual environments. Modern VR conferencing systems integrate three-dimensional avatars, spatial audio, hand tracking, and real-time networking to simulate face-to-face communication more naturally than conventional video conferencing platforms. By allowing participants to communicate using head movements, hand gestures, body language, and shared virtual spaces, these systems improve social presence, communication efficiency, and collaborative performance while providing richer interpersonal interaction (Dionisio et al., 2013; Bailey & Bailenson, 2015; Speicher et al., 2019).

Although several commercial platforms, including Spatial, Mozilla Hubs, FrameVR, and MeetinVR, provide practical solutions for remote meetings, important technical and interaction challenges remain unresolved. Many existing systems still depend on handheld controllers, provide limited support for natural interaction, or fail to reproduce realistic body movements and expressive avatar behaviors. In addition, issues related to network synchronization, interaction latency, avatar realism, and intuitive user interaction continue to represent active research challenges in collaborative virtual environments (Freina & Ott, 2015; Benford et al., 2001; Churchill & Snowdon, 1998; Radianti et al., 2020).

Natural interaction has become one of the fundamental requirements for improving the effectiveness of immersive collaborative systems. Vision-based hand-tracking technologies, such as Leap Motion, eliminate the need for conventional controllers and enable users to manipulate virtual objects and communicate through intuitive gestures. Integrating accurate hand tracking with articulated humanoid avatars significantly improves realism, reduces cognitive workload, and enhances users' perception of presence during collaborative activities. Previous studies have consistently shown that natural interaction contributes substantially to user immersion, communication quality, and collaborative efficiency in virtual environments (Weichert et al., 2013; Han et al., 2020; Hatzipanayioti et al., 2019).

Motivated by these challenges, this study presents the design and implementation of a multi-user virtual reality conferencing system that combines immersive virtual environments, humanoid avatars, controller-free hand interaction, and real-time network synchronization within a unified collaborative platform. The proposed framework enables geographically distributed users to participate in virtual meetings using head-mounted displays while interacting through head movements, natural hand gestures, synchronized avatars, and real-time voice communication. Compared with conventional video conferencing systems, the proposed approach is designed to enhance users' sense of presence, improve communication quality, and provide a more realistic and engaging collaborative experience.

The proposed system was implemented using a complete virtual reality infrastructure consisting of a VR headset, a Leap Motion hand-tracking sensor, articulated humanoid avatars, and a real-time networking framework developed in Unity. Its effectiveness was evaluated through controlled user experiments employing standardized presence and usability questionnaires.

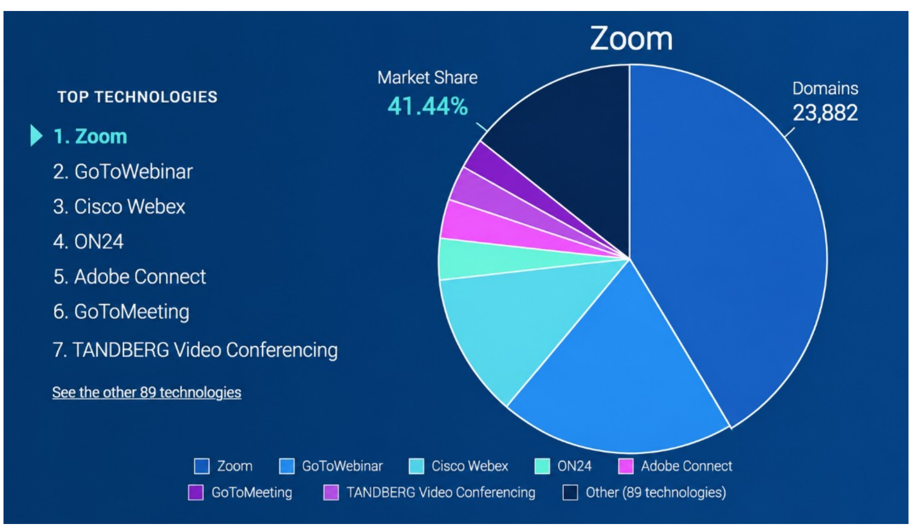

Figure 1. Comparison Between Conventional Video Conferencing and Immersive VR Conferencing Environments

The experimental results demonstrated that the proposed VR conferencing system significantly improved users' perceived presence while achieving positive usability outcomes compared with a conventional video conferencing environment. These findings highlight the potential of immersive virtual reality technologies to serve as an effective next-generation platform for remote collaboration and distributed communication.

2. Related Work

The rapid evolution of virtual reality (VR) technologies has significantly transformed remote collaboration and virtual communication. During the past decade, both academic researchers and industry have introduced numerous immersive collaboration platforms designed to enhance users' sense of presence, communication quality, and collaborative efficiency. Compared with conventional video conferencing systems, immersive virtual environments enable users to communicate through three-dimensional avatars, spatial audio, gesture-based interaction, and shared virtual workspaces, thereby providing a communication experience that more closely resembles face-to-face meetings (Slater & Sanchez-Vives, 2016; Radianti et al., 2020; Sherman & Craig, 2019).

Early research on collaborative virtual environments established the theoretical foundations for immersive remote collaboration. Benford et al. (2001) and Churchill and Snowdon (1998) demonstrated that shared three-dimensional environments can improve communication, teamwork, and information exchange among geographically distributed users. These pioneering studies highlighted the importance of immersion, avatar representation, and real-time interaction in supporting collaborative activities and inspired the development of modern VR conferencing platforms.

Several commercial VR conferencing platforms have recently emerged, including Spatial, Mozilla Hubs, MeetinVR, FrameVR, and ENGAGE. These systems support different levels of immersion, avatar representation, collaborative interaction, and cross-platform accessibility. Nevertheless, many of these platforms continue to rely on handheld controllers, provide limited support for natural hand interaction, or require expensive hardware configurations, thereby restricting their practical adoption in educational, industrial, and professional environments (Freina & Ott, 2015; Speicher et al., 2019). Furthermore, studies on immersive telepresence have shown that realistic avatar synchronization and low-latency networking are essential for achieving effective remote collaboration (Tan et al., 2017).

Natural interaction has become one of the principal research directions in immersive collaborative systems. Vision-based hand-tracking technologies enable users to interact directly with virtual objects through intuitive hand gestures without relying on conventional controllers. Previous investigations have demonstrated that accurate hand tracking significantly improves immersion, reduces cognitive workload, and enhances collaborative performance by enabling more realistic non-verbal communication (Weichert et al., 2013; Hatzipanayioti et al., 2019). Likewise, Bailey and Bailenson (2015) emphasized that immersive collaborative environments facilitate richer interpersonal communication by increasing users' perception of co-presence and social interaction.

Recent studies have also focused on improving human–computer interaction within collaborative virtual environments. Intelligent interaction techniques based on cognitive-affective architectures and natural language interfaces have demonstrated considerable potential for enhancing communication quality, adaptive interaction, and user engagement. Such approaches can complement immersive VR environments by dynamically adapting communication according to users' cognitive states and contextual information, thereby improving collaboration efficiency in shared virtual spaces (Shaykholeslami et al., 2025a).

Another important research direction concerns user engagement and behavioral analytics in immersive digital environments. Recent investigations have demonstrated that data-driven approaches can analyze users' behavioral patterns, interaction frequency, participation levels, and communication dynamics to provide a comprehensive understanding of collaborative behavior in computer-mediated environments. These analytical frameworks enable researchers to identify factors affecting user satisfaction, interaction quality, and collaborative efficiency while supporting the development of more adaptive and user-centered virtual systems. Such findings are particularly relevant to virtual reality conferencing applications, where sustained attention, active participation, and natural interaction are essential for successful collaboration (Shaykholeslami et al., 2025b).

Beyond communication quality, immersive virtual reality has also demonstrated considerable value in engineering, product development, and industrial collaboration. Berg and Vance (2017) reported that VR technologies significantly improve collaborative engineering workflows by enabling geographically distributed experts to interact with complex three-dimensional models within shared virtual workspaces. These findings further highlight the growing importance of immersive collaborative technologies across a wide range of professional applications.

Although considerable progress has been achieved in collaborative VR systems, several important limitations remain. Existing solutions frequently lack realistic full-body interaction, efficient controller-free hand tracking, scalable networking architectures, adaptive avatar synchronization, and comprehensive experimental evaluation involving real users. Moreover, relatively few studies simultaneously investigate technical performance, users' sense of presence, and system usability within a fully implemented multi-user VR conferencing framework. These limitations motivate the development of more comprehensive immersive conferencing systems capable of providing realistic communication while maintaining high usability, interaction quality, and reliable collaborative performance.

Table 1. Comparison of Commercial Virtual Reality Conferencing Platforms

Platform	Multi-user	Avatar	Hand Tracking	Spatial Audio	Cross-platform	Primary Application
Spatial	✔	✔	Partial	✔	✔	Business collaboration
Mozilla Hubs	✔	✔	✖	✔	✔	Education & Meetings
MeetinVR	✔	✔	Partial	✔	✔	Enterprise collaboration
FrameVR	✔	Basic	✖	✔	Browser-based	Education
ENGAGE	✔	✔	Partial	✔	✔	Training & Education
Proposed System	✔	Full-body Humanoid	✔ (Leap Motion)	✔	VR Platform	Immersive conferencing

Unlike existing commercial platforms, the proposed system integrates controller-free hand tracking using Leap Motion with humanoid avatars and real-time synchronization to provide a more natural and immersive conferencing experience. This combination distinguishes the proposed framework from many currently available commercial VR conferencing solutions while maintaining compatibility with collaborative virtual environments.

3. Proposed System Architecture

The proposed virtual reality conferencing system was designed based on a layered architecture to support immersive, real-time, and collaborative communication among geographically distributed users. The architecture integrates sensing devices, middleware, a three-dimensional simulation engine, and a cloud-based networking infrastructure to create a shared virtual environment in which users can communicate naturally through voice, head movements, and hand gestures. By separating the system into independent functional modules, the proposed framework improves scalability, maintainability, and compatibility with different hardware and software platforms.

Each participant is equipped with a virtual reality headset and a Leap Motion sensor. The headset continuously captures the user's head position and orientation, while the Leap Motion sensor tracks hand and finger movements using computer vision techniques. These data are processed through their corresponding middleware components before being transferred to the Unity game engine, where they are mapped onto articulated three-dimensional humanoid avatars. At the same time, a cloud networking infrastructure synchronizes the interaction data among all connected users, allowing each participant to observe the movements, gestures, and voice communications of others in real time. This integrated architecture provides a realistic and immersive collaborative environment while reducing the limitations of conventional video conferencing systems.

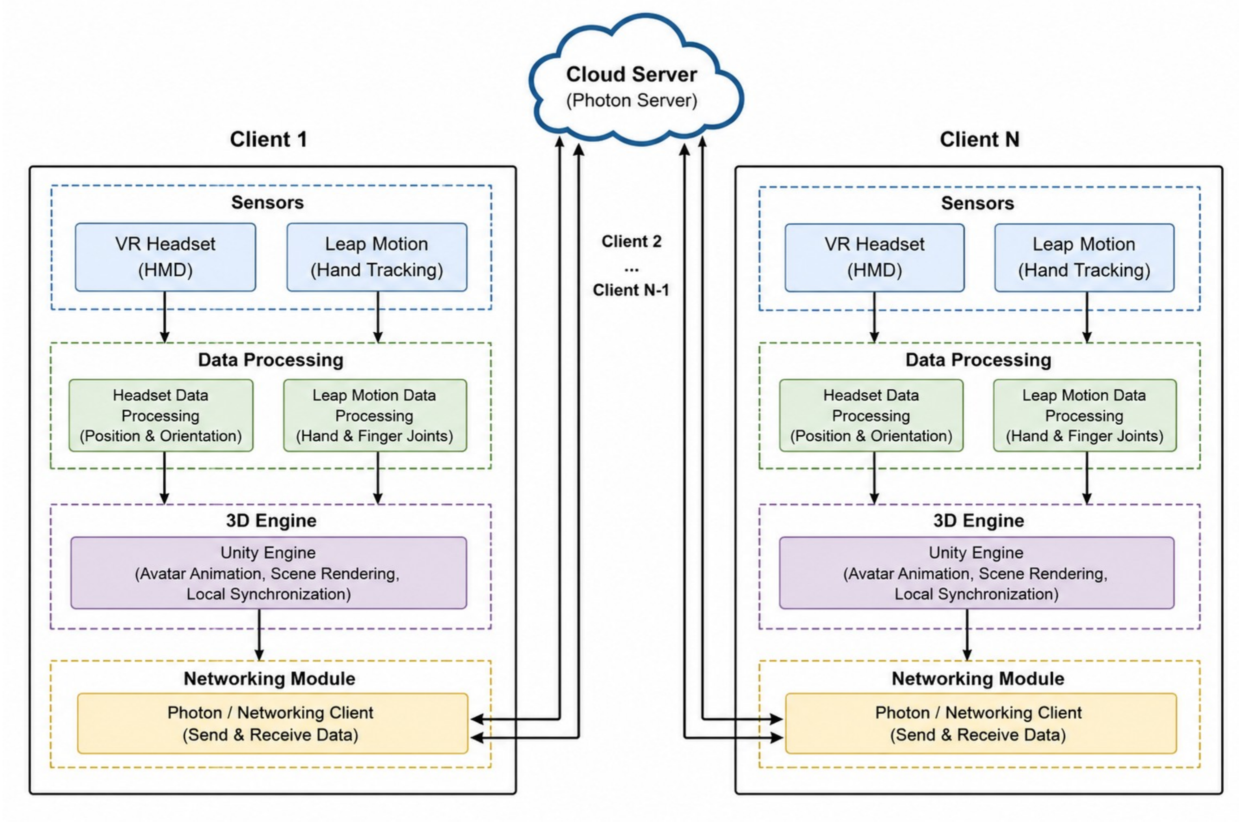

Figure 2. Overall Architecture of the Proposed Virtual Reality Conferencing System

The overall workflow of the proposed framework begins with collecting users' physical interactions through the sensing devices. The captured data are subsequently interpreted by the middleware layer and transferred to the Unity engine for avatar animation and scene rendering. Finally, the synchronized information is transmitted through the cloud infrastructure to all connected clients, ensuring that every participant experiences a consistent and interactive virtual conferencing environment. The detailed operational workflow of the proposed system is presented in the following subsection.

Following the overall system architecture, the internal processing pipeline of the proposed framework is illustrated in Figure 3. Once the sensing devices capture users' physical actions, all input streams—including head movements, hand gestures, and voice data—are transmitted to the Unity engine through their corresponding middleware components. Within the engine, a dedicated data acquisition module integrates these heterogeneous inputs before applying them to the virtual environment.

The Unity engine performs several essential tasks simultaneously. First, it imports and manages three-dimensional avatar models generated during the modeling stage. Second, it maps the received tracking data onto the corresponding skeletal joints of the local humanoid avatar, enabling real-time animation based on the user's physical movements. Third, the engine continuously exchanges synchronization data with the cloud networking infrastructure to receive the updated states of remote participants while transmitting the current user's interaction data. This bidirectional communication ensures that all users observe consistent avatar movements and interactions throughout the shared virtual environment.

Unlike conventional video conferencing systems that primarily exchange compressed video streams, the proposed framework synchronizes lightweight interaction data, including avatar joint coordinates, head orientation, user identification, and voice communication. Consequently, the communication overhead is reduced while maintaining a highly immersive collaborative experience. The modular organization of the processing pipeline also facilitates the integration of additional interaction modalities, intelligent user behavior analysis, and adaptive communication mechanisms in future implementations.

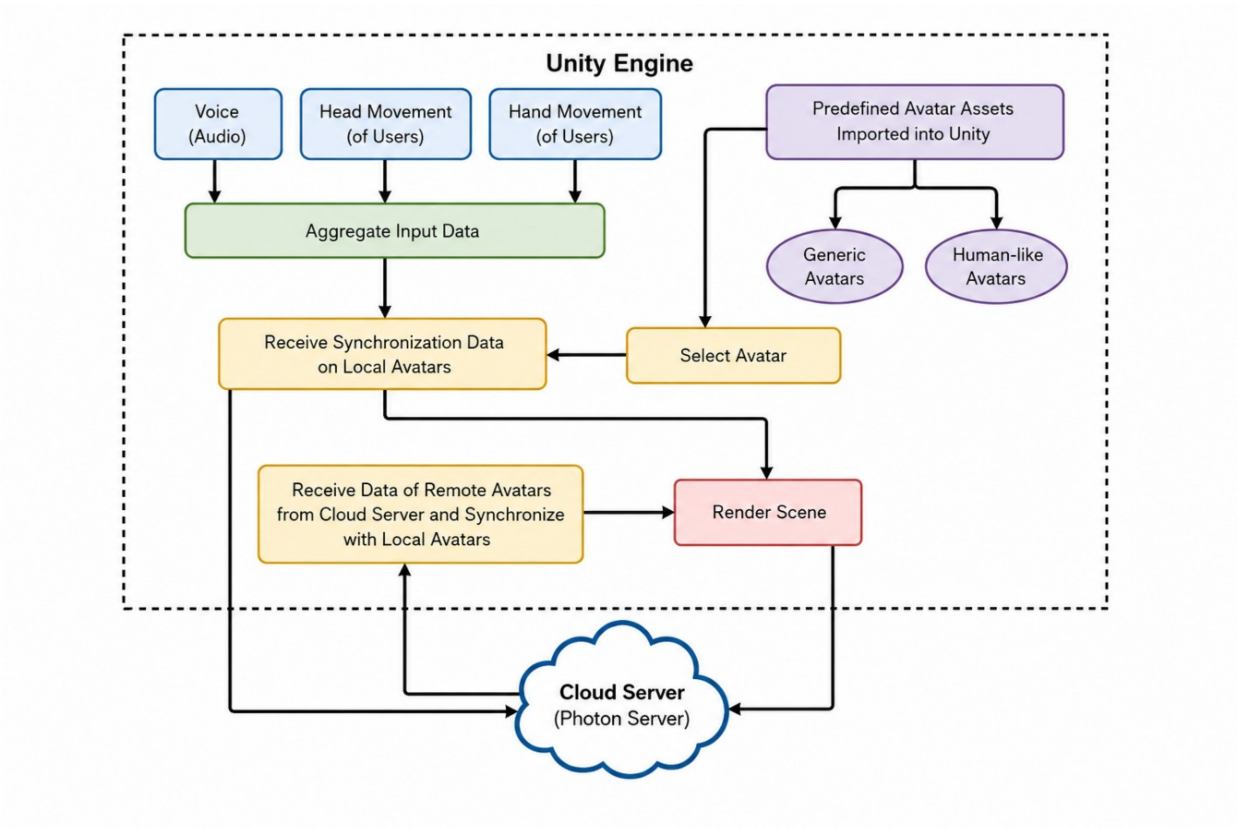

Figure 3. Main Components and Data Flow of the Proposed Virtual Reality Conferencing System

The proposed architecture demonstrates that immersive virtual conferencing requires the seamless integration of sensing technologies, middleware services, real-time rendering, avatar animation, and cloud networking. Such integration enables geographically distributed users to communicate naturally within a synchronized three-dimensional virtual environment while providing the flexibility to incorporate emerging technologies, including intelligent human–computer interaction, cognitive-affective computing, and user engagement modeling (Shaykholeslami et al., 2025a).

3.1 Overall System Workflow

The overall development process of the proposed virtual reality conferencing system is illustrated in Figure 4. Unlike conventional video conferencing platforms, which are primarily limited to two-dimensional audio and video communication, the proposed framework aims to create an immersive collaborative environment in which participants interact naturally through three-dimensional avatars, real-time voice communication, and controller-free hand tracking. Achieving such an experience requires a structured development workflow encompassing avatar creation, virtual environment design, interaction modeling, networking, synchronization, and system integration.

The implementation process begins with designing personalized three-dimensional avatars that represent participants inside the virtual environment. These avatars are subsequently rigged to enable skeletal animation and natural body movements. Next, the collaborative virtual environment is developed using the Unity game engine, where avatars, interaction mechanisms, and networking components are integrated into a unified simulation platform. Natural interaction capabilities, including head tracking, hand tracking, and voice communication, are then incorporated to reproduce users' physical actions in real time.

Following the implementation of the local interaction mechanisms, cloud-based networking services are integrated to synchronize users' positions, gestures, avatar states, and voice streams among all connected participants. Finally, the complete system undergoes functional testing and experimental evaluation to assess usability, user presence, and collaborative performance. This structured workflow provides a systematic framework for developing immersive multi-user conferencing applications while ensuring scalability, maintainability, and real-time synchronization.

Structured development workflow consisting of several sequential stages. The process begins with the creation of personalized three-dimensional avatars that represent users inside the virtual environment. To enable realistic animation, the avatars are subsequently rigged by constructing a hierarchical skeletal structure that supports head movements, body motions, and natural hand interactions.

In parallel with avatar preparation, the three-dimensional conference environment is designed using the Unity game engine. Both the virtual environment and the prepared avatars are then imported into Unity, where the interactive simulation is developed. At this stage, the required software components are assigned to the virtual objects, enabling the system to acquire tracking information from the VR headset and the Leap Motion sensor and map these data onto the corresponding avatar joints in real time.

After implementing local interaction mechanisms, networking components are integrated into the system. Upon entering the virtual environment, each client establishes a connection with the cloud server, which is responsible for synchronizing user information among all participants. Once synchronization is completed, an individual avatar is instantiated for every connected user and automatically positioned within the appropriate virtual meeting room. This workflow ensures that all participants experience a consistent and synchronized collaborative environment while maintaining real-time communication and natural interaction.

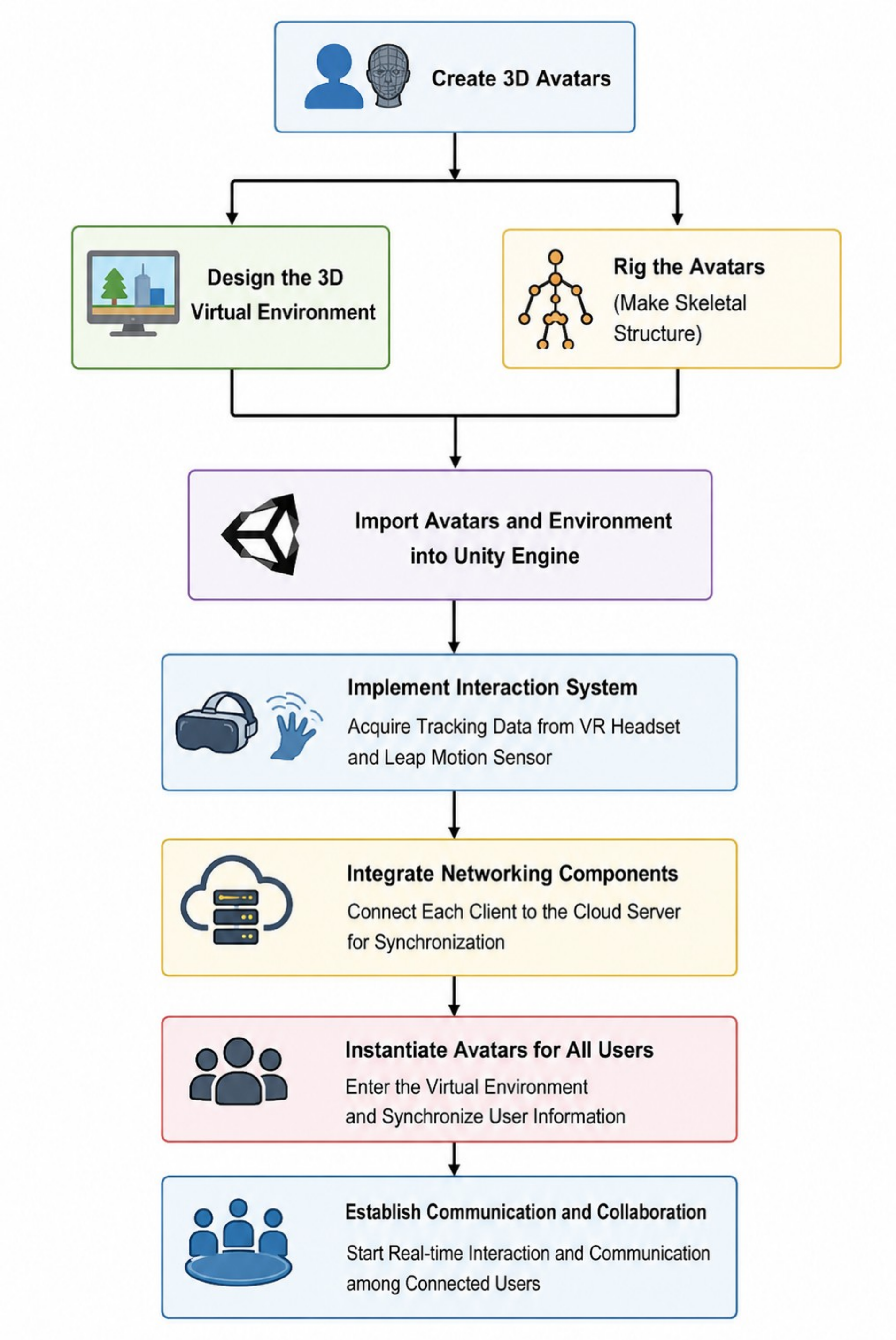

Figure 4. Overall Development Workflow of the Proposed Virtual Reality Conferencing System

The sequential development process presented in Figure 4 demonstrates that the proposed framework integrates avatar generation, environment modeling, interaction processing, cloud networking, and avatar instantiation into a unified implementation pipeline. Such an organized workflow simplifies system development, facilitates scalability, and enables future extensions through the integration of additional interaction modalities and intelligent collaboration mechanisms (Jerald, 2015; Radianti et al., 2020; (Shaykholeslami et al., 2025b).

3.2 Avatar Modeling and Skeletal Animation

The realism of collaborative virtual environments largely depends on the quality and expressiveness of user avatars. In immersive conferencing applications, avatars serve as the primary representation of participants and play a crucial role in conveying non-verbal communication, including body posture, head orientation, and hand gestures. Therefore, the proposed system employs articulated humanoid avatars capable of reproducing users' physical movements with a high degree of realism (Jerald, 2015; Slater & Sanchez-Vives, 2016).

The avatar development process begins with creating a complete three-dimensional human model using the MakeHuman platform. This software enables the generation of customizable human characters with realistic body proportions and anatomical structures. Various physical attributes, including gender, body shape, facial characteristics, and clothing, can be adjusted before exporting the model to the game engine. The generated avatars provide a suitable basis for representing users within the collaborative virtual environment while maintaining compatibility with standard animation pipelines (MakeHuman Team, 2023).

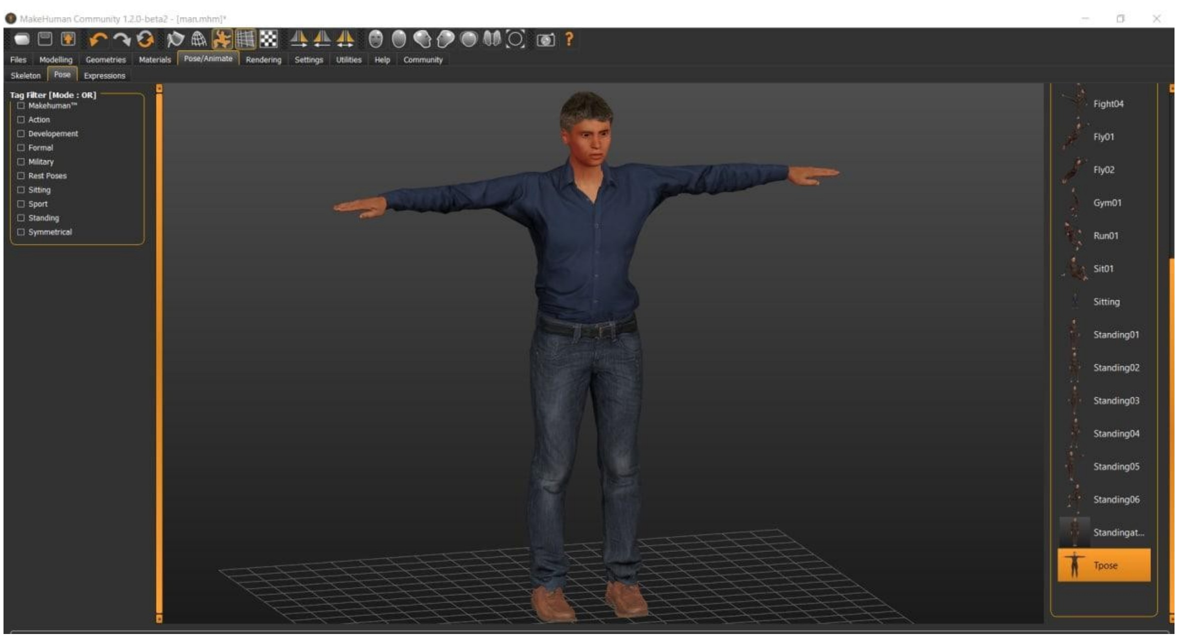

Figure 5. Creation of Three-Dimensional Human Avatars Using MakeHuman

Although the generated avatar provides a realistic visual appearance, it cannot perform body movements until an internal skeletal structure is assigned. Therefore, a rigging process is applied to construct an articulated skeleton that defines the hierarchical relationships among body joints. This skeletal model enables the animation system to reproduce human movements by transferring tracking data from the user's physical body to the virtual avatar. Rigging represents one of the most critical stages in avatar development because it directly influences animation quality, movement accuracy, and interaction realism (Han et al., 2020).

3.3 Virtual Environment Development

The virtual environment constitutes the core component of the proposed conferencing system because it provides the shared collaborative space in which geographically distributed users interact. Unlike conventional conferencing applications that rely on two-dimensional graphical interfaces, the proposed system employs a fully immersive three-dimensional environment designed to simulate a realistic meeting space. The environment was developed using the Unity game engine due to its advanced rendering capabilities, physics engine, animation framework, and networking support, making it suitable for real-time collaborative virtual reality applications (Jerald, 2015; Unity Technologies, 2023).

The virtual conference room was designed to provide sufficient space for multiple participants while maintaining realistic spatial perception and user comfort. Three-dimensional objects, lighting conditions, textures, and environmental elements were carefully configured to enhance visual realism without imposing excessive computational overhead. Since prolonged exposure to virtual environments may influence user comfort and performance, the scene was optimized to maintain stable rendering performance and minimize latency during real-time interaction (Radianti et al., 2020).

The developed environment also incorporates dedicated interaction areas where participants can communicate naturally through voice, body orientation, and hand gestures.

Figure 6. Developed Three-Dimensional Virtual Conferencing Environment

Instead of simply displaying static avatars, the system dynamically updates users' positions and orientations according to the synchronized tracking data received from the cloud server. Consequently, each participant experiences a coherent and continuously updated virtual meeting space.

The Unity engine continuously renders the virtual scene based on synchronized user interactions. Whenever a participant moves, rotates the head, or performs a hand gesture, the corresponding avatar is immediately updated within the environment. Real-time rendering ensures visual consistency among all connected users while preserving the immersive characteristics of the conferencing experience. Furthermore, Unity's modular architecture enables additional collaborative functionalities, such as shared virtual objects, presentation screens, interactive whiteboards, and intelligent virtual assistants, to be incorporated without modifying the overall system architecture (Slater & Sanchez-Vives, 2016; Speicher et al., 2019).

The immersive characteristics of the developed environment provide significant advantages over conventional video conferencing systems. Users perceive themselves as physically present inside the shared virtual space rather than communicating through isolated video windows. This enhanced spatial awareness contributes to improved social presence, more natural communication, and greater collaboration efficiency; (Shaykholeslami et al., 2025b).

3.4 Network Synchronization and Cloud Communication

Real-time synchronization is one of the fundamental requirements of collaborative virtual reality systems. Unlike single-user virtual environments, multi-user conferencing platforms must continuously exchange interaction data among geographically distributed participants while maintaining low communication latency. To achieve this objective, the proposed system adopts a cloud-based networking architecture that synchronizes users' movements, voice communication, and avatar states in real time. This synchronization mechanism enables all participants to observe a consistent virtual environment regardless of their physical locations.

The communication process begins immediately after a participant joins the virtual conference. The client establishes a connection with the cloud server and receives the information required to enter the shared virtual session. Once authentication is completed, the server creates or joins the corresponding virtual room and starts exchanging interaction data among connected users. Instead of transmitting high-bandwidth video streams, the proposed framework exchanges lightweight state information, including avatar position, head orientation, hand movements, and voice packets. This approach considerably reduces network traffic while preserving the immersive characteristics of the conferencing experience (Photon Engine, 2024).

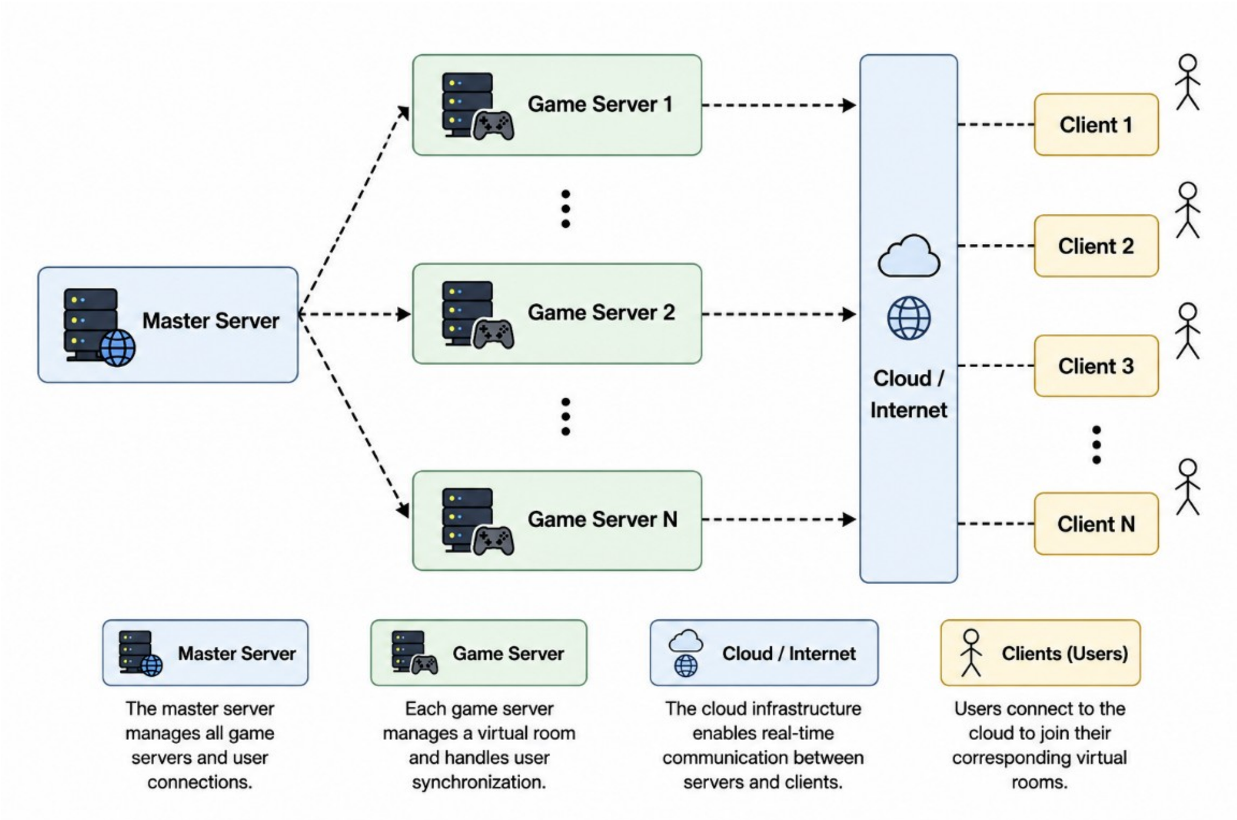

Figure 7. Real-Time User Synchronization Through the Cloud Infrastructure

After establishing the network connection, each participant is represented by an independent humanoid avatar inside the virtual environment. Whenever a new user enters the session, the corresponding avatar is instantiated automatically and synchronized with all connected clients. The cloud server continuously updates users' interaction states and distributes the latest information to every participant, ensuring that all users observe identical avatar behaviors throughout the conference.

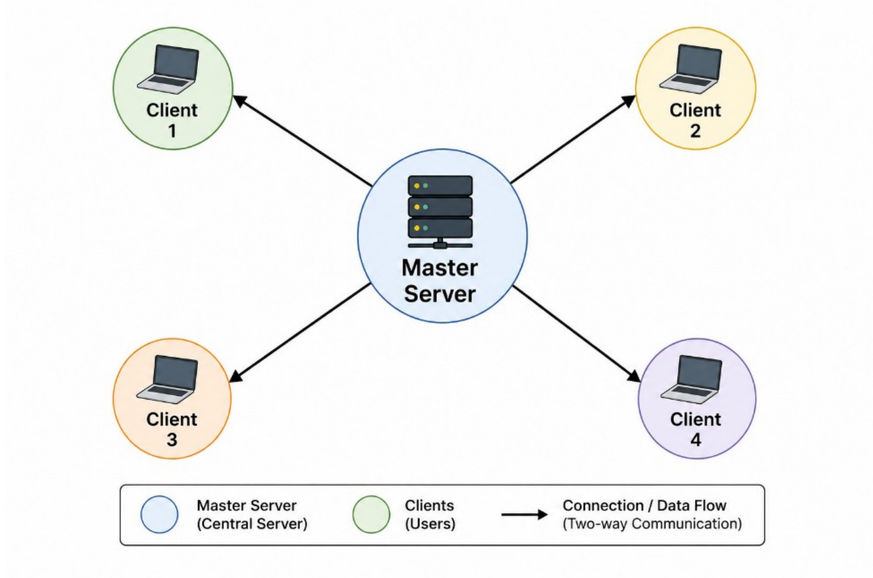

Figure 8. Multi-User Virtual Conference Room

The proposed architecture supports multiple virtual meeting rooms, allowing different conferencing sessions to operate independently without interfering with one another. Participants belonging to the same room exchange interaction data exclusively with users in that session, thereby improving scalability and reducing unnecessary network communication. This room-based organization also simplifies session management and enables multiple collaborative meetings to be conducted simultaneously.

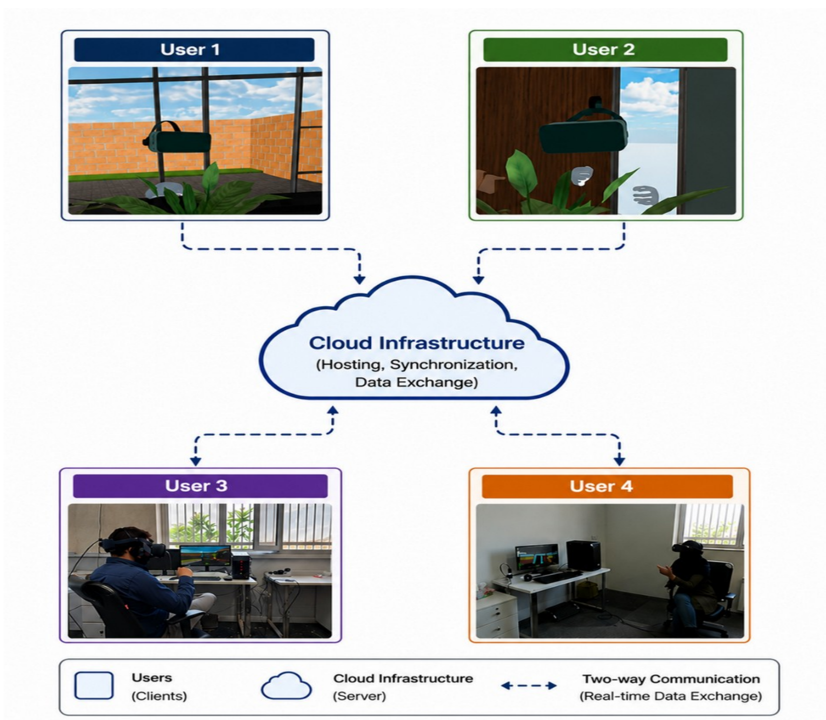

Figure 9. Cloud-Based Networking Architecture

The cloud infrastructure serves as the communication backbone of the proposed conferencing system. It manages user authentication, session establishment, room allocation, state synchronization, and message distribution among connected clients. By separating networking services from the rendering engine, the proposed framework improves modularity and allows future migration to alternative cloud platforms without requiring substantial modifications to the application architecture. Such a design also facilitates future extensions, including larger conferencing sessions, distributed collaborative workspaces, and intelligent communication services.

3.5 Natural Hand Tracking and Avatar Interaction

Natural hand interaction is one of the key features that distinguishes immersive virtual reality conferencing systems from conventional video conferencing platforms. Traditional virtual environments often rely on handheld controllers to manipulate virtual objects and perform user interactions. Although controllers provide acceptable tracking accuracy, they reduce the naturalness of communication because users must continuously hold external devices. To overcome this limitation, the proposed system employs a vision-based hand-tracking solution that allows participants to interact directly with the virtual environment using their own hands.

The Leap Motion sensor was selected as the primary hand-tracking device because it provides high-resolution tracking of both hands and individual finger joints without requiring wearable equipment. During runtime, the sensor continuously captures the position and orientation of each hand and estimates the motion of multiple finger joints. These tracking data are subsequently transferred to the Unity engine, where they are mapped onto the skeletal structure of the humanoid avatar. This process enables the avatar to reproduce users' natural hand movements in real time while maintaining synchronization with other participants (Weichert et al., 2013).

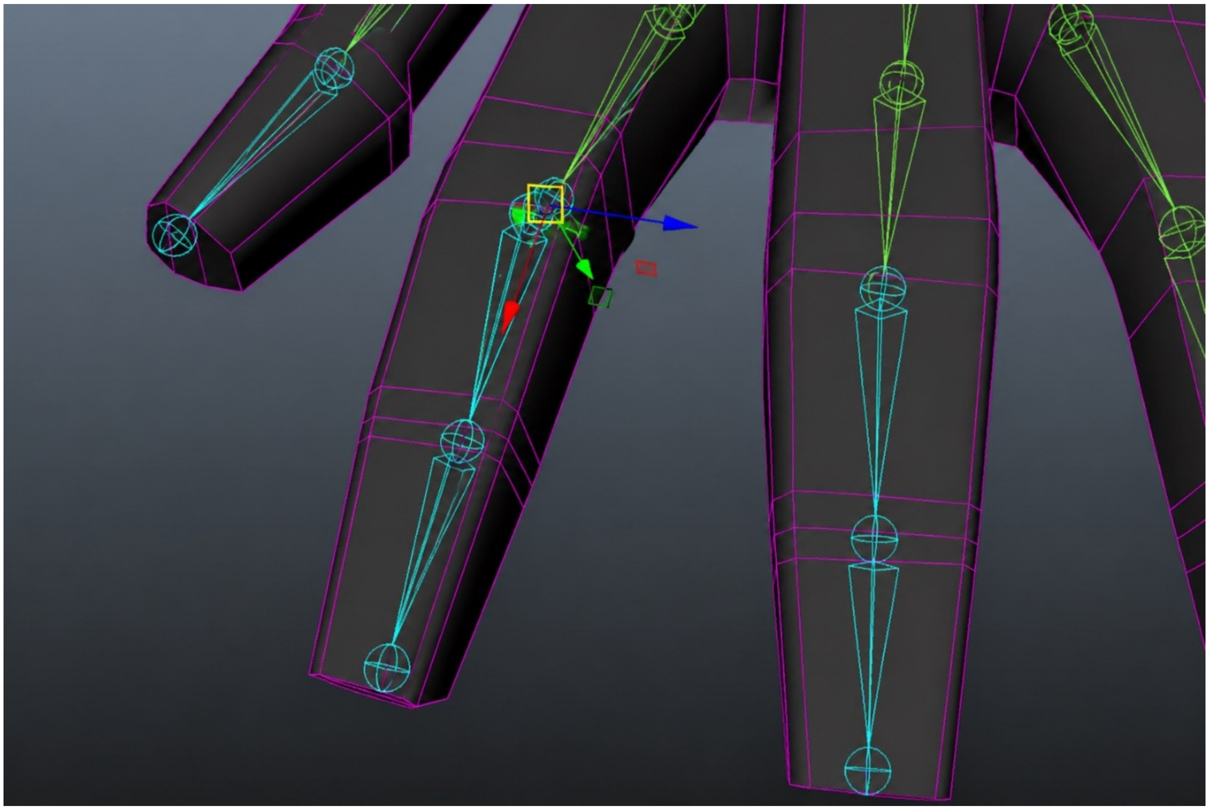

Figure 10. Skeletal Model of the Human Hand Used for Real-Time Tracking

The skeletal representation of the hand consists of multiple articulated joints connected through a hierarchical bone structure. Each finger is modeled independently, allowing the proposed system to detect complex gestures such as pointing, grabbing, and object manipulation. Instead of transmitting image frames, only joint coordinates and rotational information are exchanged, considerably reducing communication overhead while preserving interaction accuracy.

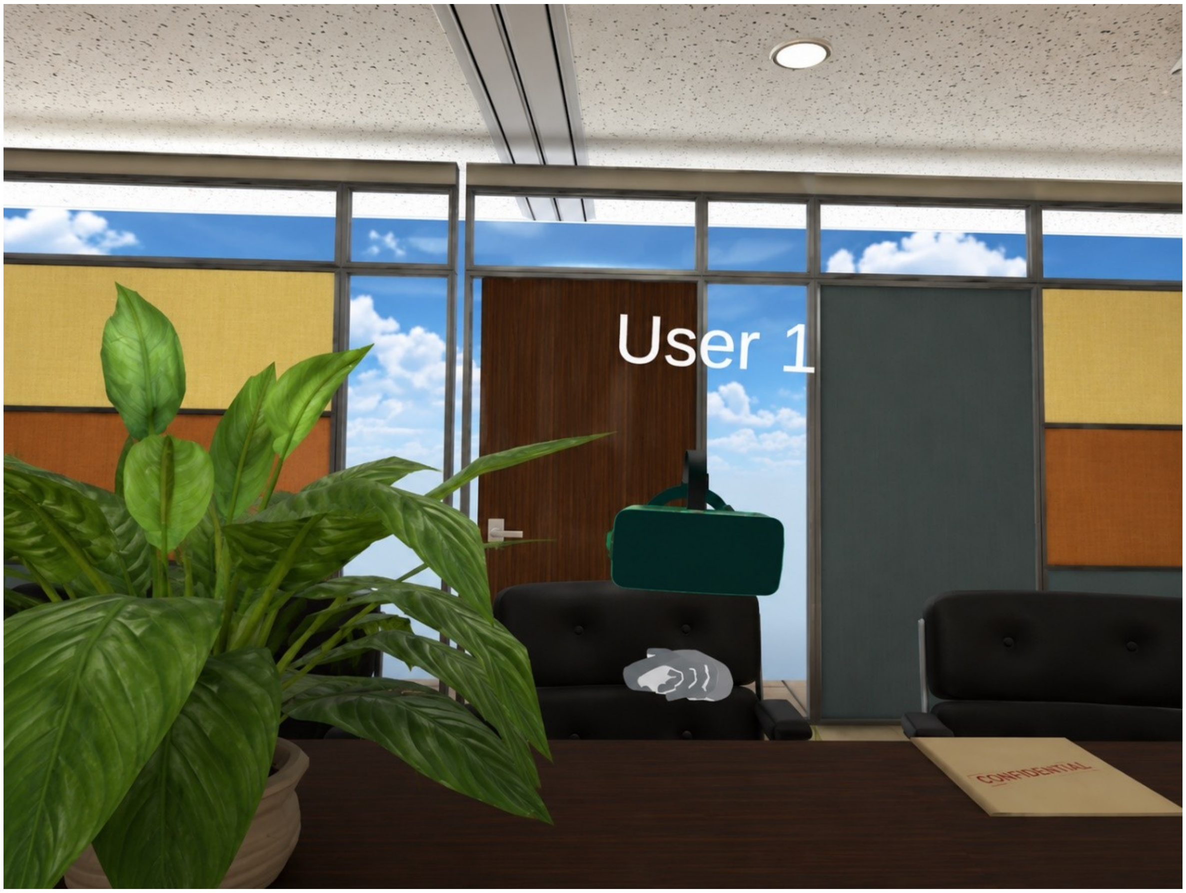

Figure 11. Real-Time Mapping of Hand Gestures to the Virtual Avatar

After the hand joints are reconstructed inside the Unity engine, inverse kinematics algorithms are applied to update the corresponding avatar bones. Consequently, every physical movement performed by the participant is immediately reflected by the virtual avatar. This mapping process significantly improves the realism of interpersonal communication because gestures become an integral part of user interaction rather than predefined animations.

Figure 12. Continuous Gesture Synchronization During User Interaction

The proposed interaction mechanism supports continuous synchronization of users' hand movements across the network. Whenever a participant performs a gesture, the updated skeletal information is transmitted to the cloud infrastructure and distributed to all connected clients. Remote users therefore observe identical gestures with minimal latency, resulting in a more immersive collaborative experience.

Figure 13. Bidirectional Hand Interaction Between Participants

In addition to gesture synchronization, the proposed framework enables users to interact naturally with virtual objects inside the conferencing environment. Interactive actions such as pointing, selecting, grabbing, and object manipulation can therefore be performed without requiring dedicated handheld controllers, making communication more intuitive and closer to real face-to-face meetings.

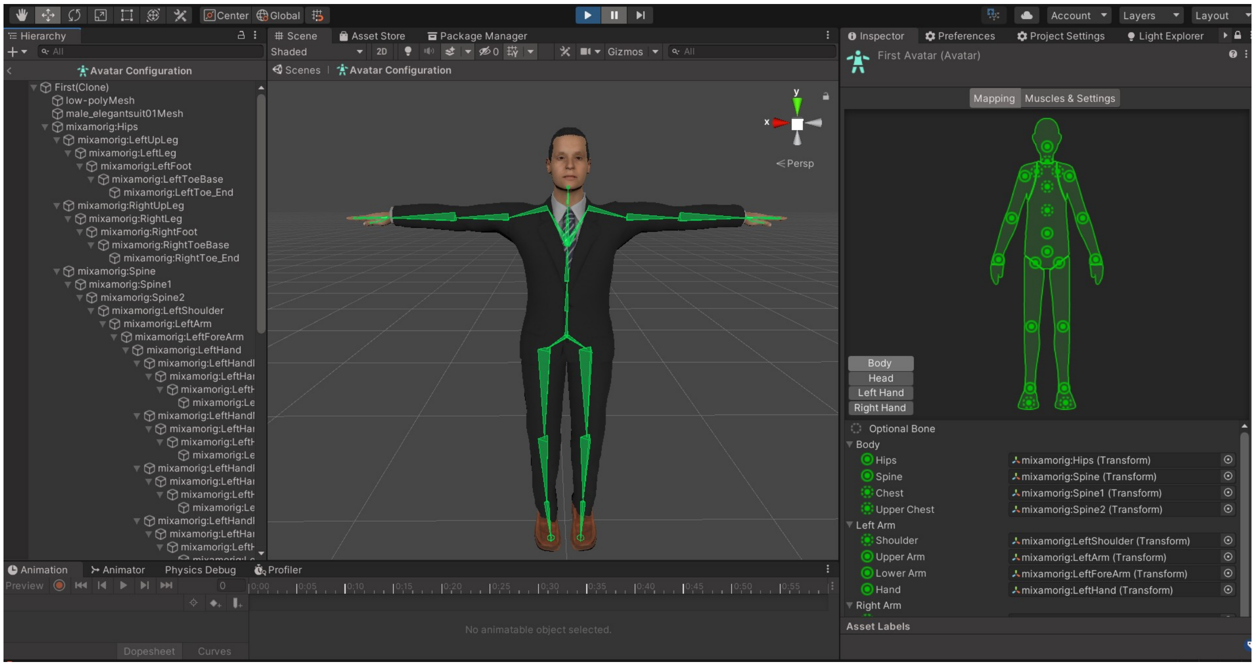

Figure 14. Final Humanoid Avatar Integrated with Natural Hand Tracking

The complete avatar interaction framework integrates real-time hand tracking, skeletal animation, cloud synchronization, and immersive rendering into a unified communication system. This integration enables participants to communicate using body language, hand gestures, and spatial movements in addition to voice communication. Consequently, the proposed conferencing platform provides a substantially richer collaborative experience than conventional two-dimensional video conferencing systems.

4. Experimental Setup

To evaluate the practical implementation of the proposed virtual reality conferencing system, a controlled experimental setup was prepared using two independent workstations connected through the cloud-based networking infrastructure. Each workstation was equipped with a compatible virtual reality headset, a Leap Motion hand-tracking sensor, audio communication devices, and the software components required to execute the developed Unity application. The experimental configuration enabled participants located in separate physical spaces to enter the same virtual conference room and interact with one another in real time. The evaluation environment was designed to examine whether the implemented architecture could successfully integrate virtual reality rendering, avatar animation, natural hand tracking, voice communication, and cloud synchronization within a unified multi-user system. During each experimental session, the users wore the head-mounted displays and entered the shared virtual environment through separate client applications. Their head movements, hand gestures, voice data, and avatar states were continuously exchanged through the network infrastructure.

4.1 Participants and Experimental Design

The study followed a quasi-experimental design in which participants experienced both the proposed virtual reality conferencing system and a conventional video conferencing environment. This design made it possible to compare the two communication settings under similar conditions. The evaluation focused on three principal dimensions: the users' sense of presence, the technical performance of the implemented system, and its perceived usability.

The participants were introduced to the equipment and received brief instructions before the experiments began. Each participant first completed the assigned conferencing scenario and subsequently answered the evaluation questionnaires. The collected responses were then used to assess the effectiveness and practical usability of the proposed system.

Table 2. Variables Considered in the Experimental Evaluation

Variable	Description	Measurement Method	Data Type
Technical Performance	Adequacy and reliability of the technical components of the proposed VR conferencing system	Questionnaire	Binary (Yes/No)
Sense of Presence	Degree to which users feel immersed and physically present within the virtual environment	Questionnaire	5-point Likert Scale (1–5)
Usability	Ease of learning, operating, and interacting with the proposed conferencing system	Questionnaire	5-point Likert Scale (1–5)

The independent variable was the type of conferencing environment, consisting of the conventional video conferencing condition and the proposed virtual reality conferencing condition. The dependent variables included the perceived sense of presence, technical acceptability, and system usability. Maintaining identical communication tasks across the two conditions reduced the influence of unrelated experimental factors and supported a more reliable comparison between the systems.

4.2 Hardware and Software Configuration

To ensure stable execution of the proposed virtual reality conferencing platform, the experiments were conducted using two identical workstations with sufficient computational resources for real-time rendering, networking, and interaction processing. Each workstation was equipped with a virtual reality headset, a Leap Motion sensor for hand tracking, headphones with an integrated microphone for voice communication, and a dedicated graphics processing unit to support immersive rendering. The Unity game engine served as the primary software platform for implementing the virtual environment, avatar animation, and interaction logic.

The client applications communicated through the cloud-based networking infrastructure, allowing multiple users to participate simultaneously in the same virtual conference room. Throughout the experiments, all hardware and software configurations remained unchanged to ensure consistency across different evaluation sessions and to eliminate external factors that could influence system performance.

Table 3. Hardware and Software Configuration Used in the Experiments

Workstation	CPU	GPU	RAM
Workstation 1	Intel Core i5 @ 4.2 GHz	NVIDIA GeForce RTX 2060	8 GB
Workstation 2	Intel Core i7 @ 4.7 GHz	NVIDIA GeForce GTX 1070	8 GB

The selected hardware configuration satisfied the computational requirements of immersive virtual reality applications while maintaining stable rendering performance and low interaction latency. The software environment integrated the Unity game engine with networking libraries and hand-tracking middleware, providing a unified platform for implementing collaborative virtual conferencing. Maintaining identical configurations for all experimental sessions improved the reliability and reproducibility of the obtained results.

4.3 Experimental Procedure

The evaluation procedure consisted of several sequential stages designed to assess both the technical functionality and the user experience of the proposed system. Initially, participants received a brief introduction to the experimental objectives and the operation of the virtual reality equipment. After becoming familiar with the controls and interaction mechanisms, each participant entered the shared virtual conference room through an individual client application.

During the experiment, participants performed predefined collaborative tasks, including verbal communication, head movement, hand gesture interaction, and avatar-based navigation within the virtual environment. All interaction data were synchronized through the cloud infrastructure in real time. Upon completion of the conferencing session, participants completed the evaluation questionnaire measuring technical performance, perceived presence, and system usability.

The collected responses were subsequently analyzed using descriptive statistics and inferential statistical methods to determine the effectiveness of the proposed conferencing system compared with the conventional communication environment.

4.3 Experimental Procedure

The experimental evaluation was conducted in a controlled laboratory environment to assess the effectiveness of the proposed virtual reality conferencing system. Before the beginning of each session, participants received a brief introduction to the objectives of the experiment and the operation of the virtual reality equipment. They were then given sufficient time to become familiar with the head-mounted display, hand-tracking sensor, and interaction mechanisms provided by the developed system.

Following the familiarization phase, participants entered the shared virtual conference room through separate client applications connected to the cloud infrastructure. During each session, users performed a series of collaborative tasks, including verbal communication, navigation within the virtual environment, object observation, and natural hand-based interaction. Throughout the experiment, users' positions, head orientations, hand movements, and avatar states were synchronized continuously among all connected clients.

After completing the conferencing session, participants completed a structured questionnaire designed to evaluate the proposed system from three different perspectives: technical performance, perceived presence, and system usability. The collected responses were subsequently analyzed to determine the overall effectiveness of the proposed conferencing platform and to compare its performance with that of conventional video conferencing systems.

5. Results and Discussion

The collected experimental data were analyzed using descriptive and inferential statistical techniques. The primary objective of the analysis was to determine whether the proposed virtual reality conferencing system could improve users' sense of presence and interaction quality while maintaining acceptable technical performance and usability. The following subsections present the statistical findings obtained from the experimental evaluation.

Table 4. Descriptive Statistics of the Experimental Variables

Group	Sample Size (n)	Mean	Standard Deviation (SD)
Conventional Video Conferencing	18	2.90	0.95
Virtual Reality Conferencing	18	4.61	1.25

Table 4 summarizes the descriptive statistics of the three evaluation variables. Overall, the participants reported positive perceptions regarding the proposed conferencing system. The technical performance of the platform received favorable evaluations, indicating stable execution throughout the experimental sessions. Similarly, the sense of presence achieved relatively high scores, suggesting that the immersive virtual environment successfully enhanced users' perception of being physically present within the shared workspace. The usability scores also demonstrated that participants considered the proposed system easy to learn and operate.

5.1 Statistical Comparison Between the Two Conferencing Environments

To determine whether the observed difference between the conventional video conferencing system and the proposed virtual reality conferencing platform was statistically significant, inferential statistical analysis was performed using a paired-samples *t*-test. This statistical procedure was selected because the same participants experienced both conferencing environments under comparable experimental conditions.

The null hypothesis assumed that no significant difference existed between the two conferencing environments, whereas the alternative hypothesis stated that the proposed virtual reality conferencing system would provide a superior collaborative experience. The analysis was conducted using a significance level of **0.05**, and all statistical calculations were performed using standard statistical software.

Table 5. Paired-Samples *t*-Test Comparing Conventional Video Conferencing and the Proposed Virtual Reality Conferencing System

Comparison Group	Mean Difference	Standard Deviation	*t*-Value	Degrees of Freedom (df)	Significance (p-value)
Conventional Video Conferencing vs. Virtual Reality Conferencing	-1.70	1.41	-5.12	17	0.000

The paired-samples *t*-test confirmed that the difference between the conventional video conferencing environment and the proposed virtual reality conferencing system was statistically significant ($t = -5.12$, $df = 17$, $p < 0.001$). Since the significance level was considerably lower than the accepted threshold of 0.05, the null hypothesis was rejected, indicating that the proposed system provided a significantly better conferencing experience than the conventional platform.

The mean difference of -1.70 demonstrates a substantial improvement in users' evaluations after experiencing the virtual reality conferencing environment. These findings suggest that integrating immersive visualization, avatar-based interaction, natural hand tracking, and cloud synchronization considerably enhances users' perception of communication quality and collaborative engagement. Although individual responses exhibited moderate variation, the overall trend consistently favored the proposed virtual reality system.

The statistical results are consistent with previous studies reporting that immersive virtual reality environments improve users' sense of presence, interaction quality, and collaborative performance compared with traditional video conferencing systems (Radianti et al., 2020). Furthermore, the integration of real-time avatar synchronization and natural interaction mechanisms contributed to creating a more realistic communication environment, allowing participants to interact more naturally than in conventional two-dimensional conferencing platforms (Slater & Sanchez-Vives, 2016).

5.2 Discussion

The experimental results demonstrate that the proposed virtual reality conferencing system successfully addresses several limitations of conventional video conferencing platforms. By combining immersive visualization, real-time networking, avatar-based communication, and controller-free hand interaction, the developed framework significantly improved users' collaborative experience. Participants consistently reported higher levels of presence, usability, and overall satisfaction when interacting within the proposed virtual environment.

The obtained findings indicate that natural interaction plays a crucial role in improving remote collaboration. Unlike conventional conferencing systems, where communication is primarily restricted to video streams and voice transmission, the proposed framework enables participants to communicate through body posture, head orientation, and hand gestures. These additional interaction channels contribute to more effective non-verbal communication and strengthen the feeling of co-presence among geographically distributed users.

The results also demonstrate the feasibility of integrating commercially available virtual reality hardware with cloud-based networking technologies to develop scalable collaborative environments. The proposed architecture provides sufficient flexibility for future integration of intelligent virtual assistants, shared virtual workspaces, and artificial intelligence–based interaction support, making it a promising platform for next-generation immersive communication systems.

The findings of this study also demonstrate that integrating cloud-based networking with immersive virtual reality technologies provides an effective solution for geographically distributed collaboration. Unlike conventional conferencing systems, where interaction is primarily limited to voice and video transmission, the proposed framework enables participants to communicate using spatial movements, natural gestures, and avatar-based interaction. These capabilities contribute to a more engaging collaborative experience and improve the quality of interpersonal communication.

Despite these promising results, several limitations should be acknowledged. The experimental evaluation involved a relatively limited number of participants and was conducted under controlled laboratory conditions. Future studies should investigate the scalability of the proposed framework with larger user groups, more diverse collaborative scenarios, and heterogeneous network conditions. Furthermore, integrating artificial intelligence techniques for adaptive interaction, gesture recognition, and intelligent virtual assistants could further enhance the functionality of immersive conferencing systems.

6. Conclusion

The present study introduced a cloud-based virtual reality conferencing framework that integrates immersive visualization, real-time networking, avatar animation, and natural hand tracking into a unified collaborative environment. The proposed system was designed to overcome several limitations of conventional video conferencing platforms by providing users with a more realistic and interactive communication experience.

Experimental evaluation demonstrated that participants consistently preferred the proposed virtual reality conferencing system over the conventional conferencing environment. Statistical analysis confirmed significant improvements in users' perceived presence, interaction quality, and overall usability. These findings indicate that immersive virtual reality technologies have considerable potential for supporting next-generation remote collaboration, education, professional meetings, and distributed teamwork.

Future research will focus on improving the scalability of the networking architecture, incorporating artificial intelligence for adaptive user interaction, supporting larger collaborative environments, and evaluating the proposed framework under real-world deployment conditions.

References

[1] Bailey, J. O., & Bailenson, J. N. (2015). Virtual reality and collaboration. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), The Oxford Handbook of Affective Computing. Oxford University Press.

[2] Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. Communications of the ACM, 44(7), 79–85. https://doi.org/10.1145/379300.379322

[3] Berg, L. P., & Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: A survey. Virtual Reality, 21(1), 1–17. https://doi.org/10.1007/s10055-016-0293-9

[4] Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer, 40*(7), 36–43. https://doi.org/10.1109/MC.2007.257

[5] Churchill, E. F., & Snowdon, D. (1998). Collaborative virtual environments: An introductory review of issues and systems. Virtual Reality, 3(1), 3–15. https://doi.org/10.1007/BF01409793

[6] Correia, A. P., Liu, C., & Xu, F. (2020). Evaluating videoconferencing systems for the quality of the educational experience. Distance Education, 41(4), 429–452. https://doi.org/10.1080/01587919.2020.1821607

[7] Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. In *Proceedings of the 11th International Scientific Conference eLearning and Software for Education (eLSE)* (Vol. 1, pp. 133–141). https://doi.org/10.12753/2066-026X-15-020

[8] Hatzipanayioti, A., Pavlidou, A., Dixken, M., Bülthoff, H. H., Meilinger, T., Bues, M., & Mohler, B. J. (2019). Collaborative problem solving in local and remote virtual reality environments. Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1734–1739. https://doi.org/10.1109/VR.2019.8797758

[9] Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Association for Computing Machinery and Morgan & Claypool. https://doi.org/10.1145/2792790

[10] Karl, K. A., Peluchette, J. V., & Aghakhani, N. (2021). Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly. Small Group Research, 53(3), 343–365. https://doi.org/10.1177/10464964211015286

[11] Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. Journal of Computer-Mediated Communication, 3(2). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

[12] Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

[13] Shaykholeslami, P., Atashafrouz, M., & Kavoosi Ghafi, A. (2025). Modeling User Engagement in Environments: A Data-Driven Approach. International Journal of Sustainable Applied Science and Engineering. https://doi.org/10.22034/ijsase.v2i3.188

[14] Shaykholeslami, P., Kavoosi Ghafi, A., & Atashafrouz, M. (2025). Text-Based Simulation of Human–Computer Interaction Using Cognitive-Affective Architectures and Natural Language Interfaces. International Journal of Sustainable Applied Science and Engineering. https://doi.org/10.22034/ijsase.v2i3.189

[15] Sherman, W. R., & Craig, A. B. (2019). Understanding Virtual Reality: Interface, Application, and Design (2nd ed.). Morgan Kaufmann.

[16] Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

[17] Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, Article 74. https://doi.org/10.3389/frobt.2016.00074

[18] Speicher, M., Hall, B. D., & Nebeling, M. (2019). What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Paper 537, pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300767

[19] Tan, Z., Hu, Y., & Xu, K. (2017). Virtual reality-based immersive telepresence system for remote conversation and collaboration. In Next Generation Computer Animation Techniques (pp. 234–247).

[20] Weichert, F., Bachmann, D., Rudak, B., & Fisseler, D. (2013). Analysis of the accuracy and robustness of the Leap Motion Controller. *Sensors*, *13*(5), 6380–6393. https://doi.org/10.3390/s130506380