Immersive Volumetric Media Streaming - A Subjective Study

Alan Teixeira *a265560@dac.unicamp.br*
Rafael Clerici *r273034@dac.unicamp.br*
Vanessa Testoni *vtestoni@unicamp.br*
Christian Rothenberg *chesteve@unicamp.br*

Departamento de Engenharia de Computação e Automação (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (UNICAMP)

Abstract

Volumetric media streaming for Sixth Mobile Communications Generation (6G) networks requires rigorous Quality of Experience (QoE) assessment beyond traditional objective metrics. This research conducts subjective evaluation of volumetric media in Virtual Reality (VR) under varying network conditions. Results reveal acceptable quality thresholds at 2-3% packet loss and 2-3ms latency—stricter than Fifth Mobile Communications Generation (5G) targets. These findings establish perceptual benchmarks to inform 6G network design and advance volumetric media standardization.

Keyword — Immersive, Volumetric Media, VR, QoE.

1. Introduction

Throughout the years, the transition from basic mobile phones and early connected laptops to sophisticated smartphones and connected devices figure as the main catalysts of mobile data surge. Multimedia streaming services of increasingly definition, such as UHD (Ultra high definition) 4K videos, specially after the explosion of social media, are well fitted by the 5G. Nevertheless, this generation still lacks the specific characteristics for a fully connected world, which could enable even more transformative solutions and accommodate a plethora of intelligent devices, services, advanced interactive/immersive technologies, and applications, for instance, Multisensorial Extended Reality (MXR) [1] in the Metaverse [2]. In this context, the upcoming 6G expects to provide improvements at the physical and network layers to accommodate emerging use cases [3] (e.g. precision agriculture, remote healthcare, intelligent environments, industrial automation, autonomous vehicles, space connectivity, innovative infrastructure, unmanned aerial vehicles (UAV), and Holographic-type Communications (HTC) [4]).

The evolution of QoE across mobile network generations demonstrates a fundamental shift from reactive measurement to proactive design integration. Looking toward 6G networks, QoE is envisioned to transcend traditional boundaries through cyberphysical experiences encompassing holographic communication, advanced Augmented Reality (AR)/VR, and tactile internet applications [5]. Consequently, subjective QoE assessment of immersive volumetric media in VR becomes relevant to establish perceptual quality benchmarks to help advance in the standardization process of this new type of media.

2. Background

2.1. HTC & Immersive Media

HTC, also referred as holographic telepresence [6] represent an advanced form of telecommunication that has the potential to span a wide range of industries, promising to transform various aspects of our lives and work. The Figure 1 specify examples further illustrate this transformative potential [7], including holographic telepresence for more engaging remote meetings, advancements in healthcare through holographic imaging and consultations, innovative retail experiences with virtual product demonstrations, enhanced education and training through interactive 3D learning environments, and captivating sports and gaming experiences.

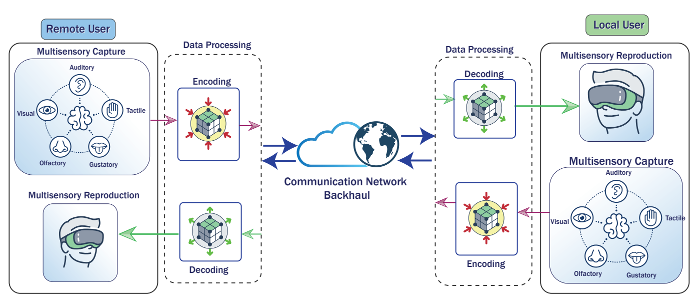


Figure 1: Holographic-type Communications concept.

The transmission of holographic content places extraordinary demands on network infrastructure, far exceeding the capabilities required for conventional communication methods. For instance, a high-resolution hologram of a single physical object, captured from multiple viewpoints, can necessitate data rates reaching several terabits per second (Tbps), coupled with a latency of less than one millisecond to achieve a truly immersive experience. These figures highlight the bandwidth challenge that HTC presents, often surpassing the capabilities of even advanced 5G networks in many deployment scenarios, as depicted in Figure 2.

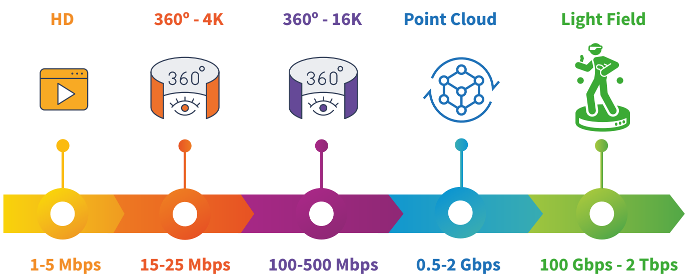


Figure 2: Throughput demands for each type of media, from standard to volumetric.

The immersiveness - the phenomenon of feeling inserted in a place other than one is physically located - is the core concept of VR. Six Degrees of Freedom (6DoF) refers to the complete set of independent movements a rigid body can make in three-dimensional space. It represents the total number of ways an object can move through and orient itself in 3D space. That is, the translational aspect of Three Degrees of Freedom (3DoF) in the X-axis, Y-axis and Z-axis plus the rotational aspect that consists in rotation around X-axis (Roll), the rotation around Y-axis (Pitch) and rotation around Z-axis (Yaw).

2.2. *Subjective Quality Assessment*

Subjective quality assessment for volumetric video involves conducting user studies to obtain ground truth quality scores. Since volumetric video offers 6DoF interaction, traditional 2D video assessment methods need adaptation. The main challenge is that 3D content occludes itself, making it impossible to view the entire model at once. Two primary inspection methods address this:

- Non-interactive (Passive): Pre-rendered videos from fixed camera trajectories shown to all participants. This method ensures reproducibility and consistent viewing experience and is simpler to implement, though less realistic.
- Interactive: Users freely manipulate viewpoints in real-time. This method better simulates real-world consumption and captures user preferences, but introduces inter-viewer variability. Often requires more participants or unlimited interaction time to compensate for variability.

3. Methodology

To carry out the QoE of immersive virtual reality, a local testbed for volumetric media streaming was designed and implemented at LCA's laboratory.

3.1. *Testbed Setup*

To support the delivery and processing requirements of volumetric streaming in VR Head Mounted Display (HMD), the testbed incorporates a multi-tier edge computing model in which network slices are mapped to far-edge and near-edge computing tiers, as depicted in Figure 3.

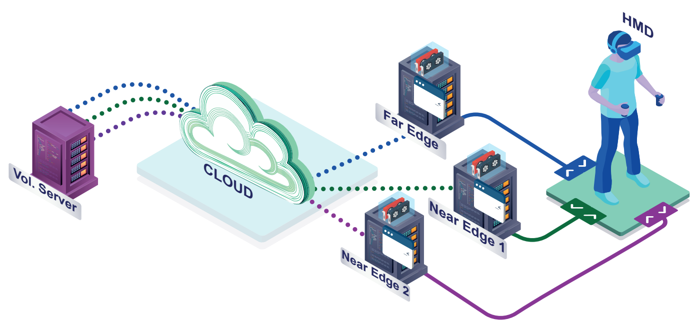


Figure 3: Setup for experimental testing.

Slice 1 is mapped to far-edge computing, which is generally located at the network periphery (e.g., cell sites) offering ultra-low latency capabilities, particularly suitable for real-time viewpoint rendering offloading. In contrast, both *Slice 2* and *Slice 3* operate within the near-edge computing domain, situated at network aggregation points or regional data centers where higher latency, but greater computational capacity, are anticipated than the far edge. Within this tier, *Slice 2* represents a moderately impaired condition where services, such as content transcoding/caching, can still be performed reliably by maintaining a satisfactory user experience. *Slice 3*, also on the near edge, reflects further degraded conditions with increased latency and packet loss to represent scenarios where network congestion or resource contention limits performance.

The transport protocol of choice for this testbed is Quick UDP Internet Connections (QUIC). QUIC is a transport layer protocol operating over User Datagram Protocol (UDP). Its implementation resides in user space rather than the operating system kernel. As QUIC runs on top of UDP rather than replacing Transmission Control Protocol (TCP) directly, and it's the underlying protocol for Hypertext Transfer Protocol Third Version (HTTP/3).

3.2. HMD and Immersive Application

The HMD device employed was the *Meta Quest 3* model, which features a field of view of 110 degrees, a video refresh rate of 120 Hz, a resolution of 2064x2208 pixels per lens, a *Qualcomm Snapdragon XR2 Gen 2* chipset, an *Adreno 740* GPU, 8 GB of RAM and a *Horizon OS* operating system.

Unity[1] is a cross-platform, real-time development environment engineered for the creation of interactive 2D, 3D, VR, and AR content. At its core is a component-based architecture where entities, or `GameObjects`, are imbued with functionality through the attachment of various components that dictate their behavior, appearance, and physical properties. The platform's primary scripting language is C#, which is used to implement complex logic and manipulate objects within the scene graph.

Google's Draco[2] is an open-source library created by the Chrome Media team to enhance the storage and transmission of 3D graphics. This testbed applied Draco with Quantization Parameter (QP) of 11.

3.3. Subjective Study

This study conducted a subjective study where the participants provided an Absolute Category Rating (ACR) Mean Opinion Score (MOS) score following the ITU-T P.910 recommendation. The Table 1 illustrate the network disruptive metrics deployed for each Slice. The two first experiments deployed latency whereas the third and fourth presented packet loss percentages.

Table 1: Applied Metrics for each experiment.

	1st Exp.	2nd Exp.	3rd Exp.	4th Exp.
*Slice*1	0 ms	0 ms	0%	0%
*Slice*2	2 ms	5 ms	2%	4%
*Slice*3	3 ms	6 ms	3%	5%

[1]Unity Engine - https://unity.com/products/unity-engine
[2]https://github.com/google/draco

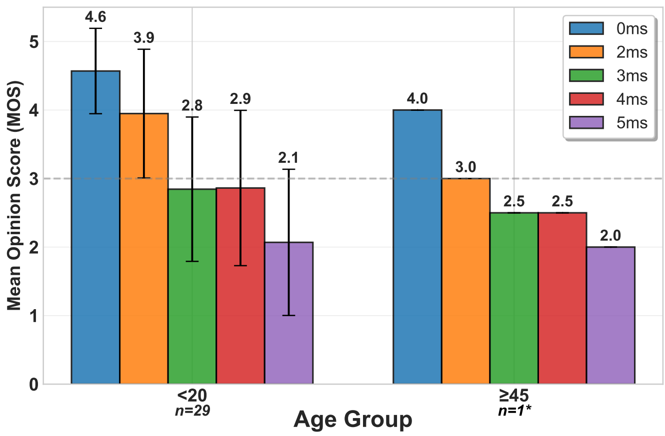
4. Impacts of Research

This research on subjective QoE assessment for volumetric media streaming holds potential across multiple societal domains. Establishing perceptual quality benchmarks enables holographic healthcare consultations that democratize access to specialized medical expertise and immersive educational environments for communities. As society transitions toward increasingly digital interactions, rigorous quality standards ensure these transformative 6G technologies deliver meaningful experiences that genuinely enhance human communication rather than merely adding complexity.

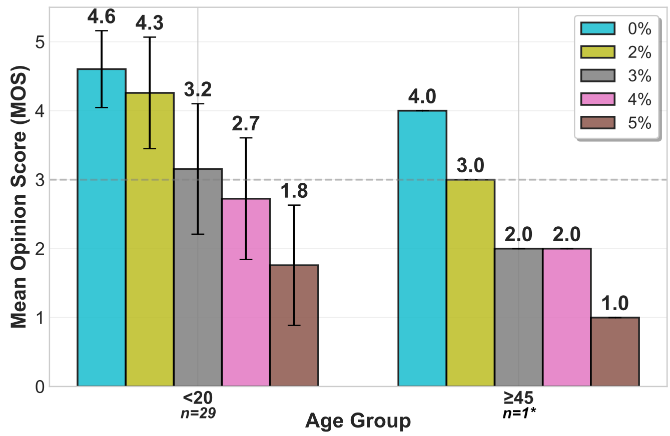
5. Results and Discussion

Figure4a illustrates the influence of network latency on MOS across age groups. Younger participants (n=29) reported a MOS of 4.6 under baseline conditions (0ms added latency), which decreased to 3.9 at 2ms latency and further degraded to approximately 2.8-2.9 for 3-4ms latency, before reaching 2.1 at 5ms. The acceptable quality threshold (MOS=3.0) was crossed between 2ms and 3ms latency addition. Older participants (n=1) demonstrated heightened sensitivity, with scores declining from 4.0 (0ms) to 2.0 (5ms), remaining consistently below the younger cohort. Notably, the marginal difference between 3ms and 4ms latency in the younger group suggests a potential perceptual saturation effect. These findings underscore the critical importance of ultra-low latency for maintaining acceptable QoE in immersive volumetric applications, with degradation occurring at latency levels below typical 5G network performance targets.

Figure 4b presents the impact of packet loss on MOS across age groups. For participants under 20 years old (n=29), MOS degraded progressively from 4.6 (excellent quality) under ideal conditions to 1.8 (poor quality) at 5% packet loss. The quality threshold of acceptable experience (MOS=3.0) was crossed between 2% and 3% packet loss, where scores dropped from 4.3 to 3.2. Participants aged 45 and above (n=1) exhibited a similar degradation pattern but with consistently lower scores across all conditions, ranging from 4.0 (0% loss) to 1.0 (5% loss), suggesting increased sensitivity to network impairments. The results demonstrate that even modest packet loss rates (2-3%) significantly impact perceived quality in volumetric media streaming, with impairments becoming unacceptable beyond 3% packet loss for both age groups.



(a) Latency



(b) Packet Loss

Figure 4: MOS Values for collected age groups.

6. Conclusions

This research establishes critical QoE benchmarks for volumetric media streaming in VR environments through comprehensive subjective assessment. The findings demonstrate that network impairments significantly degrade user experience, with acceptable quality thresholds crossed at 2-3% packet loss and 2-3ms additional latency—requirements stricter than current 5G performance targets. These empirical results provide essential perceptual quality baselines that inform the development of 6G network architectures and advance standardization efforts for immersive volumetric applications. The observed age-related sensitivity variations highlight the importance of considering diverse user populations in future QoE modeling. Future work will extend these findings through objective metric development and cross-layer optimization strategies to ensure reliable delivery of multi-sensory immersive experiences across heterogeneous network conditions.

Acknowledgments

References

[1] F. Ferrise *et al.*, "Multisensory experiences in extended reality," *IEEE Computer Graphics and Applications*, vol. 44, no. 4, pp. 11–13, 2024.

[2] P. Enenche *et al.*, "On the road to the metaverse: Point cloud video streaming: Perspectives and enablers," *ICT Express*, vol. 11, no. 1, pp. 93–104, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405959524001395

[3] C.-X. Wang *et al.*, "On the road to 6g: Visions, requirements, key technologies, and testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

[4] I. F. Akyildiz *et al.*, "Holographic-type communication: A new challenge for the next decade," *ITU Journal on Future and Evolving Technologies*, vol. 3, no. 2, pp. 421–442, 2022.

[5] S. Dang *et al.*, "What should 6g be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 1 2020.

[6] J. van der Hooft *et al.*, "A tutorial on immersive video delivery: From omnidirectional video to holography," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1336–1375, 2023.

[7] A. Clemm *et al.*, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.