



QoE Benchmarking of Resilient 6G Teleoperation: VR Digital Twin vs. Video Streaming

Kevin-Ismet Šabanović, Hendrik Schippers, Karsten Heimann and Christian Wietfeld

Chair of Communication Networks, TU Dortmund University, 44227 Dortmund, Germany

Email: {Kevin.Sabanovic, Hendrik.Schippers, Karsten.Heimann, Christian.Wietfeld}@tu-dortmund.de

Abstract—Teleoperation of robots is a key enabler for operations in complex and unpredictable environments such as rescue robotics or urban automated driving. Next-generation 6G networks target immersive teleoperation as a core service, using Virtual Reality (VR) interfaces that offer more intuitive control and a comprehensive perception of the environment through Digital Twins (DTs), which provide a virtual representation of the remote system. We present a testbed that combines two industrial robotic arms with a virtualized air hockey game, enabling interactive real-time control under different visual feedback modes. Four distinct modes were studied: high-quality 6G VR Video and 6G VR DT, and their degraded counterparts, impaired VR Video and impaired VR DT, under emulated wireless impairments. Players control the robots via hand gestures in VR, ensuring consistent interaction while isolating the impact of visual feedback. Our results show that the VR DT is significantly more robust to packet loss than video streaming. While the impaired VR DT mode remained close to optimal performance, impaired VR Video breaks down due to severe artifacts and interruptions, leading to a breakdown in playability and player confidence. In fact, players using the digital twin feedback achieved up to 70% better offensive and defensive gameplay performance compared to the degraded video mode. Overall, transmitting only state data using semantic communication instead of an entire video stream reduced bandwidth requirements drastically while maintaining or even improving controllability. This work demonstrates how immersive teleoperation over digital twins can enhance both Quality of Experience (QoE) and reliability in future 6G networks and provides a platform for systematically studying human-robot interaction under adverse network conditions.

Index Terms—Quality of Experience, Virtual Reality, Teleoperation, Gamification, Immersive Media, 6G, Digital Twin, Video Streaming, Semantic Communication

I. INTRODUCTION

The demand for remotely operated or autonomous robotic systems is steadily increasing, not only in industrial settings but also in everyday environments. Robots are no longer limited to fixed production lines but are now used in agriculture, healthcare, logistics, and domestic assistance. This expansion into less controlled environments increases the variety of possible use cases, but also the complexity of systems. While full automation is desirable, it is not always realistic.

In unpredictable scenarios such as rescue robotics, search-and-rescue missions, or emergency situations, human intervention must always remain possible. A robot may encounter an unknown obstacle, a victim in need of delicate handling, or a sudden change in the environment. In these cases, having a human teleoperator as a fallback can mean the difference between mission success and failure.

Teleoperation faces several challenges in such scenarios. Reliable and low-latency communication is required to ensure safe and precise control. At the same time, operators depend

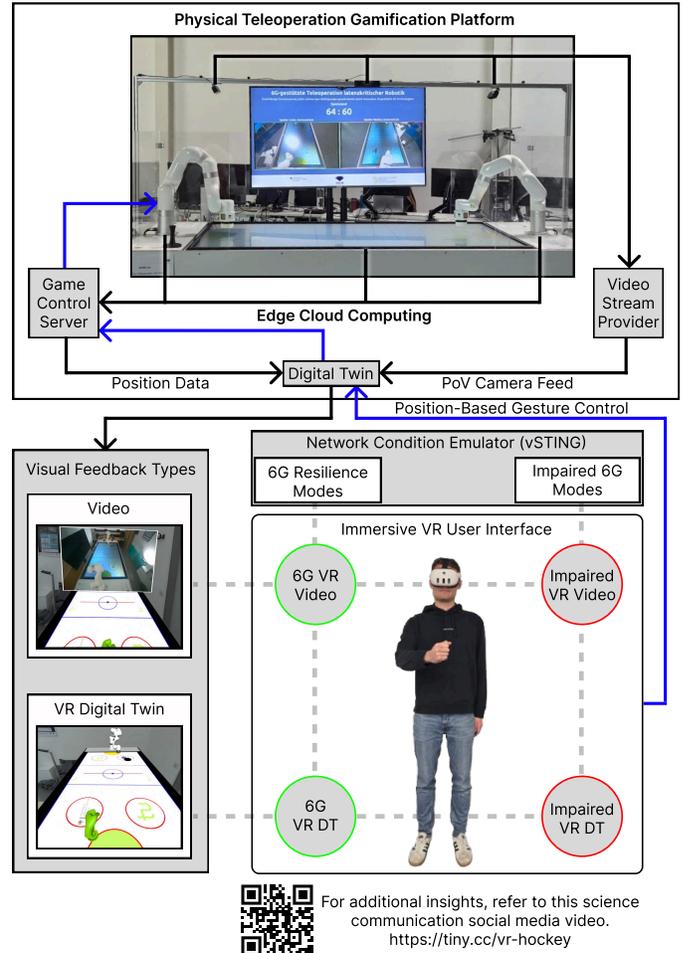


Fig. 1. System architecture of the teleoperation testbed with robotic arms, an edge-hosted digital twin, a VR headset and an emulated wireless channel.

on effective feedback, either through video streaming or immersive digital twins. In addition, teleoperation performance must be preserved under adverse network conditions such as increased latency, packet loss, jitter, and limited bandwidth.

From a 6G perspective, immersive teleoperation is expected to become a native service enabled by extended ultra-reliable low-latency communication (xURLLC), semantic communication, and edge/cloud integration. 6G mechanisms such as SEAMLESS multi-link solutions [1] and RISE proactive video encoding [2] already address parts of these challenges by improving connectivity and video robustness. However, despite these advances, adverse network conditions and their impact

on teleoperation performance may still occur in practical deployments. To investigate these remaining challenges, we built a real-time critical, virtual teleoperation system inspired by air hockey, enabling a direct comparison of different control schemes under variable network conditions. The system supports two immersive visual feedback modalities rendered in VR: conventional video streaming and semantic digital twin feedback. While additional interaction modes such as line-of-sight or tablet-based control are available, this work focuses on immersive VR teleoperation. We experimentally compare video-based and semantic digital twin-based teleoperation under adverse network conditions using vSTING [3], which enables real-time emulation of wireless channel effects, and analyze their impact on controllability, operator experience, and bandwidth requirements. Our results show that transmitting semantic information instead of raw video streams can significantly reduce throughput demands while maintaining effective control.

The remainder of the paper is structured as follows. Related work is discussed in Sec. II, followed by the system architecture in Sec. III, methodological aspects in Sec. IV, and the main findings in Sec. V.

II. RELATED WORK

Research on robot teleoperation spans a wide range of domains, including industrial, medical and collaborative robotics. A central theme across these works is balancing intuitive control, reliability, and communication quality. Early studies established the value of immersive interaction for improving teleoperation, while recent work increasingly addresses degraded network conditions. Luu et al. [4] investigated immersive teleoperation under varying impairments, including packet loss and latency, and showed that immersive setups are more resilient than traditional control schemes. Similarly, Cheng et al. [5] introduced an immersive teleoperation framework with active stereo visual feedback, showing that coupling head movements with robot-mounted cameras enhances spatial awareness and precision during fine-grained tasks. Both works highlight the importance of robust, perceptually aligned visual feedback, which directly relates to our finding that feedback stability under impaired networks critically determines playability. In contrast, our study provides a direct comparison of conventional video streaming and semantic digital twin feedback under identical network impairments, combining subjective QoE ratings with objective gameplay metrics in a controlled benchmark. This explicit contrast and quantification set our contribution apart from prior work. Compared to related immersive teleoperation systems ([4], [5]), our work isolates the visual feedback modality and evaluates DT vs. video under the same induced loss, providing a reproducible QoE and performance benchmark rather than a single-system demonstration.

Research explored gesture-based interaction and collaborative robot programming. Barattini et al. [6] proposed a gesture set for industrial robots, highlighting the advantages of natural hand-based control in human–robot interaction. Calandra et al. [7] proposed a digital twin- and extended reality-based telepresence framework for collaborative robot programming, showing how immersive XR interaction and

real-time 3D reconstruction enable remote cooperation. Their study analyzed bandwidth and latency constraints, emphasizing that next-generation 6G connectivity will be critical for high-throughput telepresence. Zheng et al. [8] examined multi-modal feedback in cognitive human–robot interfaces, showing that visual and semantic channels can be combined to reduce teleoperation workload. Additionally, research in collaborative robotics emphasizes safety and usability. El Zaatari et al. [9] reviewed cobot programming methods, stressing the integration of human input and safety mechanisms. Rosen et al. [10] evaluated mixed-reality headsets for communicating robot arm motion intent, highlighting both potential and hardware limitations. Together, these studies form a foundation for our work, which explicitly contrasts video-based and semantic digital twin teleoperation in a real-time, gamified environment.

III. ROBOTIC ARM CONFIGURATION AND AIR HOCKEY SIMULATION PLATFORM

This section introduces the developed platform, which enables real-time, network-dependent teleoperation and explains the complete system architecture.

A. Physical Demonstrator

The setup consists of two robotic arms with six degrees of freedom. These arms are commonly used in industrial scenarios and are well-suited due to their acceleration and movement accuracy. Each robot provides a networked control interface that enables access to sensory data and supports external movement commands. A large horizontal display is placed between the two arms to represent the playing field (cf. Fig. 2). This screen supports a range of two-dimensional test environments and also enables the discussed air hockey inspired use case. The gamified design makes it easy for participants to understand the task quickly, while still stressing timing, precision, and reaction speed in a reproducible way. A puck moves across the field and must be blocked or pushed into the opponent’s goal. The puck position is shared over the network as latency-critical state information and thus acts as a sensitive

TABLE I
CONFIGURATION PARAMETERS

<i>Parameter</i>	<i>Value</i>
Video Stream Encoding and Processing	
Framework	GStreamer with NVENC
Resolution	1920 x 1080 @ 60 Hz
Videocodec	H.264 (xfast, zerolatency)
Digital Twin	
Frameworks	ROS2, WebXR
DT Update Rate	60 Hz
Hardware Components	
Robotic Arms	2 × 6-DOF industrial arms
Playing Field	86” horizontal screen
Cameras	2 × PoV cameras
VR-Device	Standalone VR Headset (pass-through enabled)

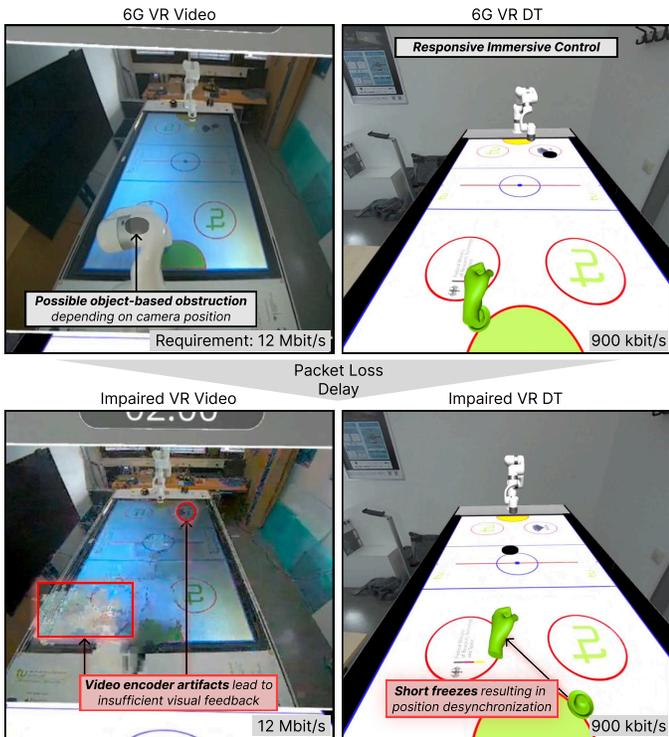


Fig. 2. Visual feedback comparison between *impaired VR Video* (left) and *6G VR Video* (right). Packet loss without stabilization leads to visible compression artifacts, while 6G countermeasures (RISE) preserve quality.

benchmark for communication quality. Above the field, a beam holds cameras, one above each arm, serving as point-of-view cameras, as shown in the top left of Fig. 2. Network impairments are emulated using vSTING [3], which enables reproducible latency, jitter, and bursty packet loss. This allows the evaluation of teleoperation performance under identical, repeatable network stress conditions. Fig. 1 shows how all components, including the robotic arms, user devices, the edge-hosted digital twin, and vSTING, are interconnected. The detailed configuration parameters of video streaming, digital twin rendering, and hardware components are summarized in Tab. I. The air hockey scenario is chosen as benchmark because it requires fast reactions, precise timing, and continuous bidirectional control, while still being safe, repeatable and accessible to participants.

B. Data Transmission

All data exchange is implemented using ROS2. Robot joint states, puck positions, and control commands are published as topics at fixed update rates. This ensures synchronized operation of robotic arms, digital twin, and VR headset. ROS2 also distributes the semantic state data used to render the digital twin and to log experiment events. Using ROS2’s built-in QoS profiles, we can enforce reliable or best-effort delivery depending on data type, which is critical when comparing digital twin versus video feedback.

C. Digital Twin and Control

The digital twin is implemented in WebXR [11], creating a virtual air hockey table with player pieces (paddles), the

puck, and even the opposing robot arm. The player’s hand is tracked above the table and is mapped to the local robot arm, providing a direct position-based control. This gesture-driven mapping ensures identical control across all conditions, isolating the effect of visual feedback. While the platform also supports alternative input schemes, such as velocity-based joystick control, we deliberately restrict this study to position-based control via VR hand tracking.

Two visual feedback modalities are supported (Fig. 2):

- **Virtual Reality Video (*VR Video*):** A video feed from the PoV cameras, encoded and transmitted via WebRTC, is displayed on a virtual screen inside the VR headset. This mirrors traditional remote operation via camera views but remains bandwidth intensive and fragile under loss.
- **Virtual Reality Digital Twin (*VR DT*):** The VR headset shows a fully virtual scene rendered in WebXR, including the air hockey table, the puck, both paddles, and the opposing robot arm. This semantic representation is continuously updated from ROS2 state messages, making it lightweight, robust, and inherently synchronized with the physical setup.

IV. EXPERIMENTAL DESIGN AND MEASUREMENTS

This section outlines the methodology applied to evaluate our air hockey teleoperation platform as an VR- and 6G-oriented test bed. The goal is to analyze how different communication setups and network conditions affect task performance and player experience. For consistency, we use a VR headset across all modes. Control is performed exclusively via the gesture-based interface described in Sec. III-C, ensuring that the evaluation focuses purely on the visual feedback required for successful teleoperation. The opposing robotic arm is controlled via a simple algorithm that always follows the current lateral position of the puck. This was chosen to increase the focus on player behavior rather than producing the undefeatable opponent. For the *VR Video* modes, a virtual screen is positioned above the digital twin air hockey table and environmental feedback from the twin is not provided. In the *VR DT* modes, only digital twin feedback using semantic communication is shown to the player.

TABLE II
EXPERIMENTAL MODES FOR IMMERSIVE TELEOPERATION EVALUATION.

Mode	Network Characteristics
6G VR Video	12 Mbit/s H.264 stream with RISE multi-link stabilization [2]
Impaired VR Video	Same stream without stabilization; $\approx 1\%$ avg. loss with 20–25% bursts; one-way delay 10 ms ± 5 ms
6G VR DT	900 kbit/s position/state data rendered via digital twin; SEAMLESS for robust multi link capabilities [1]
Impaired VR DT	Same digital twin data without SEAMLESS; identical loss/delay as impaired VR Video [1]

A. Data Types and Logging

The system exchanges multiple classes of data with distinct requirements: Robot pose information is transmitted as joint state messages at a nominal frequency of 10 Hz, providing timestamped position, velocity, and effort values. For the VR operation, teleoperator hand poses are transmitted at 30–90 Hz as lightweight JSON-style messages. The puck and pusher positions, updated at 30 Hz, are treated as latency-critical state information and prioritized in the analysis. For the *VR Video* modes, one camera delivers a Point of View (PoV) stream, encoded on the local device and forwarded to the edge server.

All relevant data streams and game events are recorded in unified log files on the central server. Logged topics include puck and player positions, robot joint states, VR hand poses and control inputs. Game events, such as touches, goals, and reflections, are also captured. These logs serve as the basis for both objective performance metrics and QoE indicators.

As illustrated in Fig. 2, packet loss in the impaired VR Video mode leads to severe encoder artifacts and reduced visual stability. By contrast, RISE stabilization in the *6G VR Video* mode ensures much more consistent visual quality. We explicitly model packet loss because, in wireless networks, degraded channel conditions or mobility often manifest first as connectivity issues, which in turn cause packet loss. This makes loss a representative stressor for assessing robustness of both video and semantic visual feedback. From the recorded log files, we derive task-level metrics, including the number of reflections, scoring and defensive success as well as player positioning. These values serve as proxies for Quality of Service (QoS) from the player’s perspective.

B. Experimental Modes

To systematically compare visual feedback modes, we define four setups shown in Fig. 2 and parameterized in Tab. II. Two modes are based on high-quality, near-optimal connections (*6G VR Video* and *6G VR DT*), while two modes emulate degraded wireless conditions (*impaired VR Video* and *impaired VR DT*) using vSTING [3]. The impaired profiles are aligned with findings from empirical wireless measurement studies [12], which report throughput inefficiencies, burst losses, and variable delay particularly under mobility and handover.

For the near-optimal conditions, we assume that our proposed 6G mechanisms such as SEAMLESS [1] multi-link transmission and RISE [2] proactive video streaming stabilize the connection. Accordingly, both *6G VR Video* and *6G VR DT* are expected to run without noticeable artifacts or instability, serving as the reference cases for comparison. This design allows direct comparison between video-based and digital twin-based teleoperation, both under stabilized (6G) and impaired network conditions.

C. Experimental Protocol and Tasks

The evaluation follows a within-subjects design. Each participant experiences all four modes in randomized order. Each trial includes a short warm-up phase, followed by a single air hockey game per mode. The task communicated to the participants was to achieve the best possible score by blocking the opponent’s shots and pushing the puck into the opponent’s goal as often as possible within a two-minute time

span. Objective measures include number of touches, types of reflections, time to goal, scoring & defensive success and actual active playtime.

D. Participants and Subjective Measures

We involve 25 participants in the study, which were mainly students aged 25–35 (with a few outliers) and reported moderate prior VR experience (mean 2/5, where 1 indicates no prior experience and 5 indicates regular personal VR use). After each of the four modes, players complete a post-trial questionnaire on a 7-point Likert scale. It consists of seven categories with two questions each, addressing different aspects of the user experience: workload/effort (NASA-TLX [13]), usability/control (SUS [14]), presence/immersion (IPQ [15]), trust/confidence (RoSAS [16]), comfort/side effects (SSQ [17]). We further include reliability/stability and safety considerations, which are rarely covered in prior work but are crucial for evaluating immersive teleoperation under differing conditions. Qualitative feedback is additionally gathered through short interviews after all modes that the participants are free to take.

In summary, the methodology combines objective system-level metrics with subjective participant ratings, ensuring that both technical and experimental aspects of teleoperation are captured. This mixed-method approach enables a robust comparison of *VR Video* versus *VR DT* control, under both 6G-enhanced and impaired conditions.

V. EXPERIMENTAL RESULTS AND PLAYER EXPERIENCE EVALUATION

This section presents a quantitative and qualitative analysis of the gameplay sessions, focusing on offensive and defensive performance as well as subjective QoE. The central benchmark is the *goal ratio* (GR), which captures how reliably players were able to defend their goal while scoring against the robotic opponent. Supporting metrics explain why the four modes differ, before the subjective ratings are summarized.

A. Player Performance Analysis

Fig. 3 highlights the core finding of this work. Both *6G VR Video* and *6G VR DT* delivered strong results under near-optimal conditions, confirming that stabilized links provide reliable control. However, under degraded network conditions, the gap between modalities became decisive. While *impaired VR DT* retained a competitive GR of 1.9 close to the optimal cases, *impaired VR Video* breaks down, achieving a substantially worse GR of 1.2. In fact, immersive digital twin feedback achieved up to 70% higher offensive and defensive performance compared to *impaired VR Video*. The share of own goals was consistently low across all conditions, showing that the vast majority of goals were deliberate player actions rather than random events. This confirms that differences in goal ratio stem from the feedback stability.

Fig. 4 shows that all modes have a balanced mix of player-generated and player-independent wall reflections. However, the absolute interaction counts differ strongly: *impaired VR Video* saw far fewer total reflections, while the other conditions had higher counts. This suggests that participants engaged in more active play under VR than video. In other words, VR

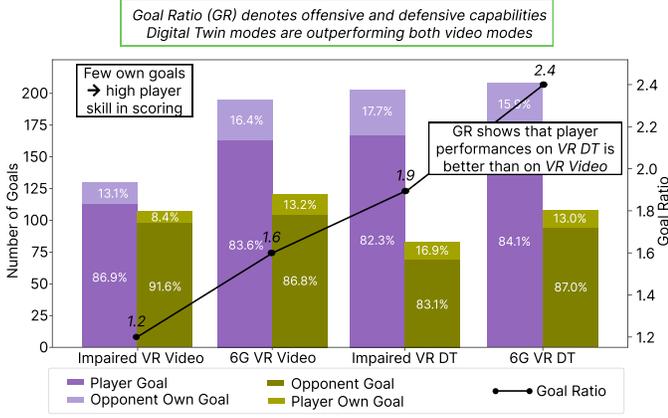


Fig. 3. Player Performance: Goal Ratio (scored vs received) across modes. Higher GR shows better combined offensive and defensive performance.

feedback not only preserved the proportions of defensive versus offensive interactions, but also increased overall activity.

In addition, the downtime analysis (Fig. 5) further explains the break down in *impaired VR Video*. When the puck was hidden by either the robotic arm or encoder artifacts, players hesitated until it reappeared. This effect was particularly severe in *impaired VR Video*, where reset positions after a goal were often invisible due to artifacts. As a result, stationary phases doubled to 34% compared to only 16% in *6G VR Video*, while VR modes remained substantially lower. Occasionally, brief puck stalls were observed in *impaired VR DT*, but they did not materially affect playability. Across measurements, video breaks down mainly due to inter-frame dependencies (artifact propagation) and stalls, while DT feedback remains stable because it relies on lightweight state updates that degrade more gracefully under loss. However, the visual stability of digital twins under packet loss can mask network issues, so operators may not immediately notice a loss of control, especially when unexpected objects or uncertainty are present.

B. Quality of Experience (QoE) Analysis

While objective metrics explain differences in player performance, the subjective ratings (Fig. 6) capture how participants experienced them and provide additional insight into player acceptance. For this purpose, we also report the Mean Opinion Score (MOS), which condenses the questionnaire ratings into a single representative value per question¹. The overall MOS for each mode is defined as:

$$\text{MOS}_{\text{Total}} = \frac{1}{Q} \sum_{j=1}^Q \underbrace{\frac{1}{N} \sum_{i=1}^N s_{i,j}}_{\text{MOS for each question}}, \quad (1)$$

where $s_{i,j}$ denotes the score of participant i for question j , N the number of participants and Q the total number of questionnaire items. This averaging first computes the MOS per question and averages each item equally weighted to yield the global $\text{MOS}_{\text{Total}}$ for each mode.

¹Questionnaire results and long form questions online available: https://github.com/tudo-cni/qoe_vr_digital_twin_results/ (accessed 31. Oct. 2025)

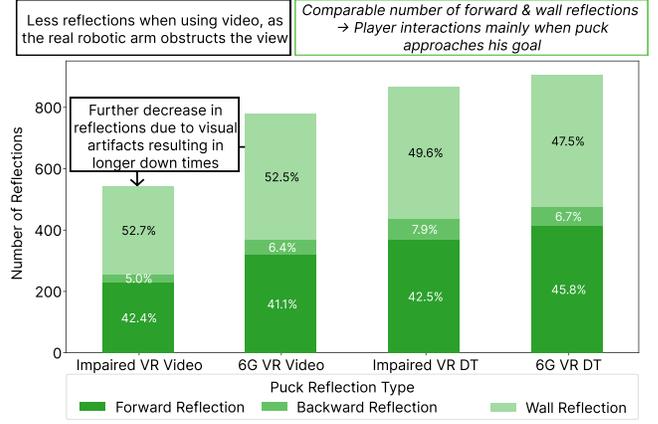


Fig. 4. Distribution of puck reflection types by mode. Equal shares of direct and wall reflections indicate balanced play.

As expected, both *6G Video* ($\text{MOS}_{\text{Total}} \approx 5.0$) and *6G VR DT* ($\text{MOS}_{\text{Total}} \approx 5.18$) received high scores across all categories, confirming that stable links allow both visual feedback modalities to support effective teleoperation. Notably, *6G VR DT* achieved slightly higher $\text{MOS}_{\text{Total}}$ despite requiring only ~ 900 kbit/s compared to ~ 12 Mbit/s for *6G VR Video*, underlining the efficiency of semantic rendering in terms of bandwidth and energy consumption. Under impairments, however, the divergence became striking: *impaired VR Video* breaks down in nearly every dimension ($\text{MOS}_{\text{Total}} \approx 2.71$), while *impaired VR DT* still preserved usability and confidence at a moderate level ($\text{MOS}_{\text{Total}} \approx 4.51$). Participants consistently described impaired video as difficult to follow and often “unplayable.” In contrast, *impaired VR DT* maintained mid-level ratings for *Trust*, *Immersion*, and *Presence*, showing that semantic rendering degrades more gracefully under packet loss. The difference is especially evident in immersion-related dimensions: while both optimal modes achieved high scores for *Presence* and *Immersion*, only *VR DT* feedback preserved these values under degraded conditions. Participants emphasized that even when frames were lost, they could still anticipate puck motion and maintain a coherent spatial model of the game. Video degradations, by contrast, caused abrupt breakdowns. With only 1% packet loss, streams already suffered from artifacts, reducing trust in the system.

Overall, the subjective ratings and MOS confirm the cen-

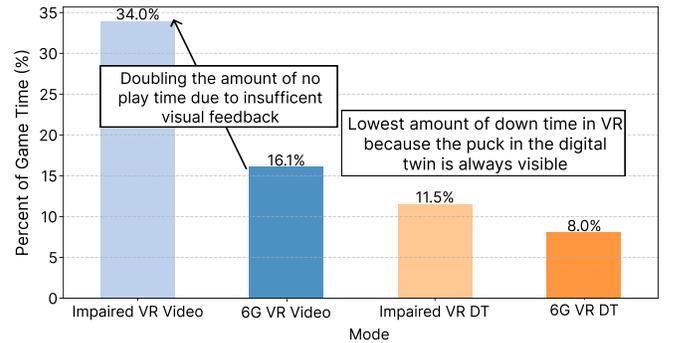


Fig. 5. Percentage of time in which the puck remained stationary. Impaired VR Video shows doubled downtime.

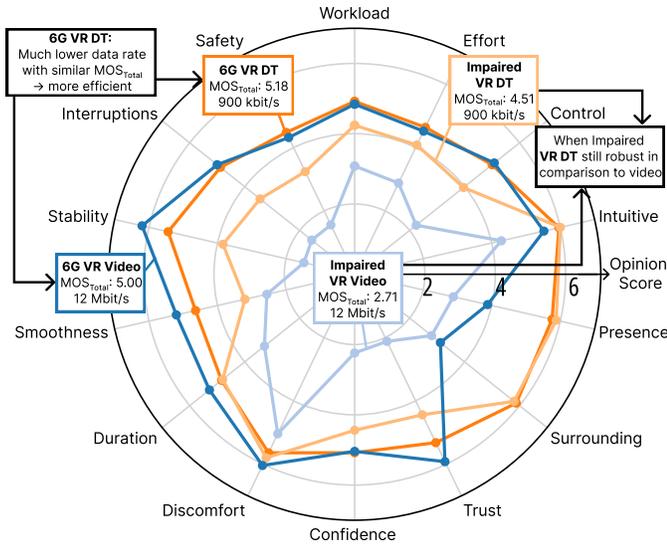


Fig. 6. Quality of Experience: Mean response for each questionnaire item, grouped by mode. MOS confirms that impaired VR Video scores the lowest across all categories, while VR DT sustains higher values even under loss.

tral result: whereas *impaired VR Video* fails abruptly under adverse conditions, *impaired VR DT* rendering sustains player experience by safeguarding embodiment and continuity. This positions *VR DT* feedback as a more resilient and resource-efficient foundation for teleoperation in 6G, reducing reliance on fragile raw video streams even when advanced countermeasures are unavailable.

VI. CONCLUSION AND OUTLOOK

Utilizing our previously proposed 6G mechanisms SEAMLESS [1] and RISE [2], both *6G VR DT* and *6G VR Video* delivered high playability and QoE. When impairments were introduced via the use of vSTING [3], however, the modalities diverged: *impaired VR DT* remained close to optimal, whereas *impaired VR Video* became nearly unplayable due to flicker, artifacts, and stalls. Despite identical 1% packet loss, VR preserved *immersion*, *control*, and *trust* sufficiently to maintain strategies such as central defensive positioning and goal-to-goal shots, while video degradations undermined spatial orientation and timing. The decisive advantage of *VR DT* feedback was confirmed, showing up to 70% higher offensive and defensive performance under loss compared to degraded video. In addition, the MOS_{Total} values highlight this effect: while *6G VR Video* reached 5.0, *6G VR DT* performed slightly better with 5.18 at only 900 kbit/s compared to 12 Mbit/s, underscoring its efficiency. Even under impairments, *impaired VR DT* achieved 4.51, close to its optimal baseline, while *impaired VR Video* dropped to 2.71. This demonstrates both the bandwidth savings and the robustness of digital twins, making them a more sustainable foundation for next-generation teleoperation. Future studies should broaden network conditions (e.g. loss patterns, latency, jitter, uplink constraints), increase sample sizes, and probe longitudinal effects to capture learning and fatigue. Transferring the paradigm to manipulation and inspection tasks will test generality and safety margins. A

central direction is replacing raw video with real-time object detection and 3D pose tracking that feed a lightweight, temporally consistent digital twin. Key elements include instance-level IDs, low-latency pose encodings, temporal smoothing, sensor fusion, and occlusion handling, ideally on-device with adaptive update rates. Future evaluations should benchmark such semantic encoders against stabilized video baselines in terms of bandwidth, latency, energy, QoE, and safety under networks impairments.

ACKNOWLEDGMENT

This work has been funded by the German Federal Ministry of Research, Technology and Space (BMFTR) in the course of the *6GEM+ Transfer Hub* under grant number 16KIS2412.

REFERENCES

- [1] T. Gebauer, M. Patchou, and C. Wietfeld, "SEAMLESS: Radio Metric Aware Multi-Link Transmission for Resilient Rescue Robotics," in *Proc. IEEE SSRR*, Fukushima, Japan, Nov. 2023, best Paper Award.
- [2] H. Schippers, T. Gebauer, K. Heimann, and C. Wietfeld, "RISE: Multi-link proactive low-latency video streaming for teleoperation in fading channels," in *Proc. IEEE VTC-Spring*, Oslo, Norway, Jun. 2025.
- [3] M. Patchou, J. Tiemann, C. Arendt, S. Böcker, and C. Wietfeld, "Realtime wireless network emulation for evaluation of teleoperated mobile robots," in *Proc. IEEE SSRR*, Sevilla, Spain, Nov. 2022.
- [4] T. Luu *et al.*, "Enhancing real-time robot teleoperation with immersive virtual reality in industrial IoT networks," *Int. J. Adv. Manuf. Technol.*, pp. 6233–6257, Aug. 2025.
- [5] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang, "Open-television: Teleoperation with immersive active visual feedback," in *Proc. CoRL*, ser. Proc. Mach. Learn. Res., P. Agrawal, O. Kroemer, and W. Burgard, Eds., vol. 270. PMLR, Jul. 2024.
- [6] P. Barattini, C. Morand, and N. M. Robertson, "A proposed gesture set for the control of industrial collaborative robots," in *Proc. IEEE RO-MAN*, Paris, France, Sep. 2012.
- [7] D. Calandra, F. G. Praticò, A. Cannavò, C. Casetti, and F. Lamberti, "Digital twin- and extended reality-based telepresence for collaborative robot programming in the 6G perspective," *Digit. Commun. Netw.*, vol. 10, no. 2, pp. 315–327, Oct. 2022.
- [8] C. Zheng, K. Wang, S. Gao, Y. Yu, Z. Wang, and Y. Tang, "Design of multi-modal feedback channel of human-robot cognitive interface for teleoperation in manufacturing," *J. Intell. Manuf.*, vol. 36, no. 6, pp. 4283–4303, Aug. 2024.
- [9] S. E. Zaatari, M. Marei, W. Li, and Z. Usman, "Cobot programming for collaborative industrial tasks: An overview," *Robot. Auton. Syst.*, vol. 116, pp. 162–180, Mar. 2019.
- [10] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *Int. J. Rob. Res.*, vol. 38, no. 12–13, pp. 1513–1526, Oct. 2019.
- [11] W3C, "WebXR Device API," <https://www.w3.org/TR/webxr/>, 2020, accessed: 2025-09-23.
- [12] D. Xu *et al.*, "Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption," in *Proc. ACM SIGCOMM Conf.*, Jul. 2020.
- [13] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Adv. Psychol. North-Holland, 1988, vol. 52, pp. 139–183.
- [14] J. Brooke, "SUS: a quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, 1995.
- [15] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators Virtual Environ.*, vol. 10, no. 3, pp. 266–281, Jun. 2001.
- [16] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (RoSAS): Development and validation," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact.*, 2017, pp. 254–262.
- [17] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *Int. J. Aviat. Psychol.*, vol. 3, no. 3, pp. 203–220, 1993.