QoE in Multi-user Collaborative Virtual Reality Games: Impact of Network and Avatar Quality

Gulnaziye Bingol[1,2], Lazizjon Suyunov[3], Zukhriddin Kamolov[3], Alessandro Floris[1,2], Simone Porcu[1,2], and Luigi Atzori[1,2]
[1]DIEE, University of Cagliari, 09123 Cagliari, Italy
[2]CNIT, University of Cagliari, 09123 Cagliari, Italy
[3]Sapienza University of Rome, 00185 Rome, Italy
gulnaziye.bingol@unica.it, lazizjon.suyunov@uniroma1.it, kamolov.2006993@studenti.uniroma1.it,
{alessandro.floris84, simone.porcu, l.atzori}@unica.it

***Abstract*—This paper presents the results of a subjective assessment investigating the impact of multiple factors (network, avatar, and player role) on the perceived Quality of Experience (QoE) in a multi-user collaborative virtual reality (VR) game. Forty test participants collaborated in pairs to complete a cooking VR game, wearing a Meta Quest Pro headset, under variable network conditions (no impairments or delayed network traffic), using diverse types of avatars (cartoon-style and humanoid), and interpreting different roles (teacher and student). A humanoid custom avatar has been implemented that replicates the user's facial expressions and body movements to investigate whether the introduction of non-verbal emotional communication within a VR environment influences the perceived user experience. Quality- and emotion-related subjective metrics were rated by test participants at the end of each test session, and the computed results show that the cartoon-like avatar, being lightweight, provides the highest perceived QoE even when the network was impaired. On the other hand, while the QoE using the humanoid avatar lowers with the introduction of network distortions (because of the larger amount of data required to replicate the facial expressions), the ability to see the facial expressions of the partner prevents a greater reduction of user experience due to the network issues.**

***Index Terms*—Quality of Experience, Virtual Reality, Subjective Assessment, Avatar, Facial Expressions.**

I. Introduction

Virtual reality (VR) applications have emerged as transformative platforms that offer immersive experiences in diverse domains, including gaming, education, industry, and specialized training scenarios [1], [2]. The widespread development of these applications has encouraged significant research interest in understanding how the Quality of Experience (QoE) is perceived in VR contexts, particularly in collaborative scenarios where multiple users engage in synchronized interactions within shared virtual environments.

Recent studies on VR have primarily concentrated on analysing system performance in terms of usability, interaction, and network transmission efficiency [3]. The network transmission, especially, is of paramount interest since good coordination between users in the gaming environment allows consistent experiences for the users. For example, in [4], a narrow range of network parameters was examined to identify which transmission impairments affect the perceived QoE the most. Similarly, the study in [5] focused on optimizing QoE by analysing how a wider range of network conditions influence users' quality perception. Still, [6] investigated the critical role of network stability, specifically variations in packet loss, latency, and baseline throughput, while participants engaged in VR cloud gaming sessions. In addition, recent studies have also examined communication in social XR contexts. The study in [7] investigated delay thresholds in volumetric XR communication, while in [8], the impact of network quality across metaverse platforms is assessed. However, these studies focused on general network performance without systematically examining avatar expressiveness, particularly facial expressions and body movement, which may help users interpret intentions and positively influence the quality of collaboration and communication in shared virtual spaces [9].

Unlike previous studies that have mainly examined network impairments or tested limited experimental conditions, our approach goes further by considering the influence of network distortions (delay and jitter), type of avatar (cartoon-like human, humanoid, and humanoid equipped with facial expression capabilities), and social role (teacher and student) on the perceived user quality and emotional experience within collaborative VR environments. In particular, we have deployed a humanoid custom avatar that replicates the user's facial expressions and body movements to investigate whether the introduction of non-verbal emotional communication within a VR environment influences the perceived user experience.

By combining these variables (network, avatar, and role) within the same study, we move beyond isolated analyses and explore how they interact to influence communication and collaboration. We conducted a subjective quality assessment involving 40 participants (20 pairs with 10 females and 30 males), who were located in separate rooms and engaged in a collaborative VR cooking game through Meta Quest Pro headsets while experiencing systematically varied test conditions. Participants rated seven quality-related metrics (QoE, visual

This work has been partially supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP C29J24000300004, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART") and by the European Union under the Italian NRRP of NextGenerationEU, "Sustainable Mobility Center" Centro Nazionale per la Mobilità Sostenibile, CNMS, CN_00000023).

and audio quality, game completion efficiency, sense of presence, comfort, and collaboration efficiency) using the 5-level Absolute Category Rating (ACR) scale, and three emotion-related metrics (valence, arousal, and dominance) using the Self-Assessment Manikin (SAM) technique. The collected subjective data were analysed to compute the Mean Opinion Score (MOS) and processed using statistical analysis (Kruskal-Wallis test) to reveal significant influences of the considered factors on the considered subjective metrics concerning the perceived user experience and emotion.

The paper is structured as follows: Section II discusses the related work in QoE assessment methodologies for VR frameworks. Section III presents the developed system, whereas in Section IV we describe the design and procedures implemented for the subjective assessment. Section V discusses the achieved results, and we conclude in Section VI.

II. Related Work

The QoE assessment of eXtended Reality (XR) applications has become increasingly important to understand how users perceive and interact with immersive content under variable network conditions. In [10], a cross-lab investigation involving over 300 participants across 10 laboratories is focused on the subjective quality assessment of 360° videos as a function of sequence duration, headset, and coding degradations. The integration of simulator sickness measurements was included as a component of immersive experience evaluation, recognizing that physiological comfort affects quality perception. The study in [5] conducted a systematic examination of multiple network parameters in Virtual Reality Cloud Gaming. Through controlled tests with 30 participants, this study established critical threshold values for round-trip-time, packet loss, and random jitter, both individually and in combination. Their findings demonstrated differentiated impact patterns across these parameters, with particular sensitivity to combined conditions. The differential impact of network impairment types was further elucidated in [6] through a within-subjects study with 16 participants, comparing baseline conditions against delay (40 ms) and packet loss (0.3%) impaired scenarios. Their findings revealed that packet loss degraded user experience significantly more than delay, resulting in MOS reductions of 1.7 versus 0.9 points, respectively. They also found that impairment sensitivity varied significantly across content types, demonstrating that quality perception in interactive VR is highly context-dependent. While these studies provide an overview of the impact of network distortions on 360° videos and VR applications, only single-user scenarios were considered.

However, some studies have also explored various interaction scenarios under network constraints in multi-user collaborative VR applications. The impact of network distortions on a collaborative VR game is investigated in [4], involving 20 participants working in pairs in a subjective assessment. The experimental design controlled specific network conditions, i.e., latency (0 ms, 500 ms) and burst latency (0 ms, 500 ms with 50% probability). The results indicated that while participants could distinguish between impaired and unimpaired scenarios, they demonstrated limited ability to differentiate between specific types of network impairments. In [11], a telerehabilitation system is investigated, requiring users to maintain a horizontal bar during exercises, establishing 0-130 ms as the latency threshold for acceptable visual QoE. Another study in [12] found that in a networked haptic hockey game, quality ratings dropped below acceptable levels with just 100 ms delay. Similarly, the authors demonstrated in a collaborative balloon-bursting game that the satisfaction level declined significantly when latency exceeded 300 ms, with contextual elements like the game object properties substantially influencing perception [13]. The technical measurements by the study [14] documented substantial real-world delays in collaborative VR systems (414-605 ms), while, in [15], they revealed how progressive latency (0-450 ms) in a cooperative cube placement task degraded performance and understanding while co-presence perception remained relatively stable. While these studies provide valuable insights into network effects on collaborative VR, they focus primarily on communication infrastructure parameters.

Unlike previous studies, in addition to network quality, our approach also incorporates a controlled investigation of the avatar type on the perceived user experience of a collaborative VR application. In particular, besides a default avatar reproducing a cartoon human with no facial expressions, we developed a humanoid avatar capable of dynamically mirroring users' facial expressions in real-time. The emotional functionality of this avatar enables us to investigate whether emotional expressivity can influence collaborative interpersonal communication within VR environments. By simultaneously manipulating both network parameters and avatar capabilities, our research offers a comprehensive analysis of how technical and human-centered factors interact to shape the perceived user quality of collaborative VR experiences.

III. The Developed System

Fig. 1 illustrates the server-client architecture of the developed multi-user VR environment. The users are located in physically separated rooms (Room 1 and Room 2) and interact within a shared audio-visual virtual space through their respective Meta Quest Pro headsets. The developed system comprises three primary functional components: (i) a VR cooking application with two types of avatars; (ii) a dedicated network with management interface; iii) a dedicated Web server for data acquisition and management.

A. VR application and Avatars

The developed VR application is based on the collaborative Cooking VR environment available in [16]. It consists of a virtual kitchen, where two players can bake a pizza together. The kitchen is provided with all the necessary ingredients and tools, such as flour, water, a rolling pin, oven, which can be used by the players to bake the pizza. Moreover, a blackboard is attached to a wall to show the players the step-by-step tasks to be completed to create the pizza. The developed application was installed on the Meta Quest Pro devices.

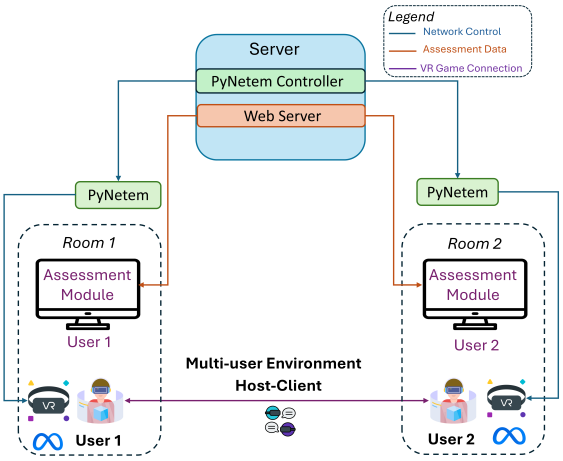


Fig. 1: The architecture of the developed system.

(a) Chef avatar. (b) Humanoid avatar.

Fig. 2: The Chef and Humanoid avatars.

In our study, we considered two types of avatars: the *Chef* and *Humanoid* avatar. Both avatars are capable of navigating the developed VR application scene and interacting with its game objects, supporting a wide range of actions essential to the test execution (e.g., grabbing ingredients and using kitchen tools). However, they differ in appearance and functionality.

The Chef avatar is the default avatar provided in [16], representing a cartoon-style character designed to appear like a professional chef, as shown in Fig. 2a. It is composed of the head, wearing a characteristic chef's hat, and the hands; it has no body. It was rigged to allow for hand movements and body transposition, though it does not support individual finger tracking. This avatar cannot reproduce any facial expression, but it always shows the same *neutral* face.

The Humanoid avatar is a realistic representation of an African-American human figure (including head, body, arms, and hands) with comprehensive blendshape support for detailed facial animations, as shown in Fig. 2b. This avatar was developed using Meta's Movement SDK [17], which allows it to reflect the user's body movements and individual finger movements accurately. Precisely, the Body Tracking module of the Meta SDK [18] constructs a skeletal model of 84 bones, which receive real-time data from the headset and controllers to replicate joint rotations and positions in the virtual environment. In the same way, the Eye Tracking API [19] captures subtle eye movements and gaze direction, enhancing avatar gaze realism. Moreover, the blendshape technique was used to dynamically map 63 different facial action units, enabling the representation of a range of facial expressions, such as smiling, frowning, and winking, in real time. Facial mapping is performed through the use of the Face Tracking API [20], which directly accesses the cameras of the Meta Quest Pro to capture the 63 facial action units and their movement intensity.

To synchronize the avatar movements, facial expressions and interactions with the game objects within the VR environment, a networking layer was implemented using Unity's Netcode [21], which supports the synchronization of game object states in multiplayer environments. The *NetworkTransform* component was used to transmit the 3D position and rotation of the avatars' components across the two headsets. However, the facial blendshape values could not be transmitted using this component because they are embedded within a single mesh and not exposed as discrete transformable objects. Furthermore, while Unity supports Remote Procedure Calls (RPCs) [22] for data transmission, the achievable update rate (approximately 3-4 FPS) was inadequate for smooth facial animation. To overcome these issues, the facial expression values were encoded into the position and rotation of empty game objects, which were then synchronized across the network using *NetworkTransform*. This approach allowed high-frequency updates (up to 40 FPS) while preserving the visual integrity of avatar expressions. All the implemented toolkits have been made public in [23]. In addition, the multiplayer architecture supports real-time vocal communication between the two players. Voice data is captured locally via the Meta Quest Pro headset's built-in microphones and transmitted across the network to remote players using Unity's low-latency audio streaming capabilities. This feature enables users to speak naturally within the virtual environment, allowing them to collaborate during the experiment.

B. Network Control and Data Management

The developed system (Fig. 1) leverages a centralized computing node implementing a Web server and creating a dedicated local Wi-Fi network used to establish communication between the two Meta Quest Pro VR headsets, preventing extraneous traffic variables. A custom interface was developed to manipulate the network through the NetEm (Network Emulator) software [24], which provides programmatic control over critical network variables at both connection endpoints, thereby achieving comprehensive bidirectional impairment simulation. In particular, the PyNetEm Controller custom interface was used to add precise network impairments, such as delay and jitter, into the dedicated network by embedding different NetEm rules, creating reproducible experimental conditions essential for comparative analysis.

A dedicated Web server was developed to manage user authentication, session parameters, and assessment data persistence. This server implements a relational database structure

that maintains hierarchical relationships between participant profiles, experimental conditions, and corresponding assessment metrics. The authentication system establishes unique participant identifiers that persist throughout the experimental session, enabling precise association for subjective assessment metrics across all test conditions. The data acquisition system facilitates the delivery of assessment questionnaires to users via the Assessment Modules, collecting responses in real-time and storing them with test condition identifiers. The workflow follows a sequential process: 1) session initiation through VR headset activation, user authentication, and network control; 2) test game session within the VR environment under controlled network conditions; 3) session termination upon headset removal; 4) collection of the subjective assessment questionnaires; and 5) data archiving with structured parameter encoding within the Web server's database.

IV. Subjective assessment

This section details the experimental design and the procedures employed in the subjective assessment.

A. Experimental Design

The experiment was designed to systematically assess the impact of three factors on the user experience in a collaborative VR cooking game: i) the network quality; ii) the type of avatar; and iii) the social role. The selection of network parameters was based on previous experimental studies [4], [5] and the result of empirical tests aimed to provide different levels of perceived QoE to the users. Finally, we selected three levels of network conditions (NCs): (NC0) no impairments, (NC1) continuous packet transmission delay (500 ms), and (NC2) variation in packet transmission delay using NetEm's normal distribution (500 ms delay $\pm$ 500 ms jitter variation).

Concerning the avatar, we considered two types of avatar: the cartoon Chef avatar and the Humanoid avatar. As shown in Fig. 2, the Chef avatar only has the head and the hands, and always shows the same neutral face with no expressions. On the other hand, the Humanoid avatar can reproduce the user's movements of the body, the facial expressions and eye gaze. To better investigate whether the emotional functionality of the Humanoid avatar influences the user's perceived experience, we have considered the Humanoid standard avatar (HST) that does not reproduce facial expressions and the Humanoid avatar that reproduces the facial expressions (HFE) of the user.

Concerning the social role, we have considered the user to assume the role of the *teacher* or the *student*, to foster communication and interaction between test participants within the VR environment. Specifically, the teacher had to provide comprehensive verbal instructions and procedural guidance for pizza creation to the student, including detailed specifications on ingredient selection, preparation techniques, and baking procedures. Concurrently, the student had to listen and execute these instructions through direct manipulation of virtual objects within the VR environment, following the teacher's guidance to complete the tasks to bake the pizza.

TABLE I: Test Conditions (TCs) for the three session tests.

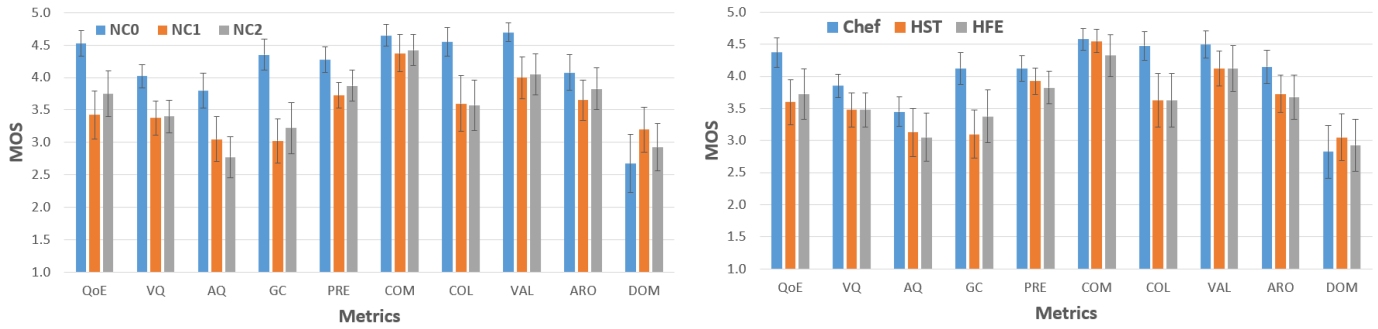
Session	NC	Delay [ms]	Jitter [ms]	Teacher Avatar	Student Avatar
1	0	0	0	Chef	Chef
	1	500	0	Chef	Chef
	2	500	500	Chef	Chef
2	0	0	0	HST	HFE
	1	500	0	HST	HFE
	2	500	500	HST	HFE
3	0	0	0	HFE	HST
	1	500	0	HFE	HST
	2	500	500	HFE	HST

Table I presents the considered test conditions (TCs) that combine the NCs, the avatar type, and the interpreted role. Each pair of test participants completed three sessions, and for each session, only an NC was set (randomly chosen among the possible three NCs) with the constraint that each session should undergo a different NC. This way, each pair of participants experienced all the different NCs. Each participant used one of the three different avatars in Table I; if the Chef avatar was used, both participants used this avatar; if the Humanoid avatar was used, one participant used the HST, while the other one used the HFE. Moreover, test participants interpreted both the student and teacher roles alternately (e.g., user1 randomly chosen as student for session1, then interpreted teacher for session2, and again student for session3, while opposite roles are interpreted by user2 among the three test sessions). Finally, note that the session order was randomized for each different pair of participants to mitigate potential learning effects according to the ITU-T P.800 [25].

B. Subjective Assessment

The subjective assessment involved 40 participants (20 pairs with 10 females and 30 males, mean 22.10 years, standard dev. 3.57). They were mainly university students (97.5%) with varying VR experience levels (65% basic knowledge, 20% no experience, 10% moderate experience, 5% expert proficiency). The ITU-T Rec. P.920 [26] suggests that familiarity between conversing participants enhances the naturalness of audiovisual interactions. Thus, we selected individuals who had prior collaborative experience and already knew each other.

Before test sessions, participants had to complete a pre-assessment protocol including informed consent documentation, demographic questionnaires, and authentication processes. Participants were explicitly informed about the number of test sessions, the different types of avatars, the alternation of teacher-student role assignments, the possibility of impairments disturbing the gaming activity, and the meaning of the subjective metrics they had to rate after each test session. Moreover, they had to participate in a training session to be familiar with the cooking VR application and learn how to move and interact with objects within the VR environment. The actual assessment was conducted in laboratory conditions with participants situated in separate rooms, where they had to wear headphones and the Meta Quest Pro headset. Each test session lasted between 5 and 15 minutes, depending on



(a) MOS vs. Network Conditions.

(b) MOS vs. Avatar.

Fig. 3: MOS with 95% CI for the considered metrics as a function of the network quality (a) and type of used avatar (b).

the network impairments and participants' ability to complete the collaborative tasks. The complete experimental procedure required approximately 50 minutes per participant pair, including the training session.

Each pair of test participants had to complete three test sessions under the test conditions summarized in Table I. The aim was to provide each participant with different levels of experience and emotions due to the combination of different NCs, avatars, and interpreted social roles. At the end of each test session, each participant had to take off the headset and complete a subjective questionnaire where he/she had to rate the level of perceived experience for 10 different metrics:

1) (QoE): the overall QoE, including all aspects of the perceived VR game experience.
2) (VQ): the visual quality of the 3D game environment, including the avatar.
3) (AQ): the audio quality concerning the communication with the partner.
4) (GC): the game completion efficiency.
5) (PRE): the sense of presence in the VR environment.
6) (COM): the level of comfort, opposite to cybersickness.
7) (COL): the collaboration efficiency with the partner to complete the game tasks.
8) (VAL): the valence, the perceived pleasantness.
9) (ARO): the arousal, the perceived intensity of emotions.
10) (DOM): the dominance, the degree of control of the perceived emotions.

The five-level Absolute Category Rating (ACR) scale [25] was used for rating the first 7 metrics, whereas the Self-Assessment Manikin (SAM) pictorial technique [27] was used for assessing the last 3 emotion-related metrics with a 5-level scale. For valence, higher values indicate more positive emotions (1 = negative, 5 = positive). For arousal, higher values indicate higher activation (1 = calm, 5 = excited). For dominance, lower values indicate higher perceived control (1 = high control, 5 = low control).

V. Results

In this section, we provide the subjective assessment results. No outliers were found. Then, we have computed the MOS with a 95% confidence interval (CI) on the scores provided for all 10 considered metrics. Then, to identify significant variations in MOS results under different test conditions, we have computed the Kruskal-Wallis test, an equivalent non-parametric test requiring the respect of three assumptions for the data under test: i) ordinal or continuous response variables, such as survey responses measured on a Likert scale; ii) independence of individual scores in each group; and iii) similar distribution shape in each group. In the next sections, the influence of network quality, type of avatar, and social role on the considered user experience and emotional metrics is discussed, respectively.

A. Influence of network

Fig. 3a shows the MOS with a 95% CI as a function of the 3 network test conditions: NC0, NC1, and NC2. As expected, when no network distortions are applied (NC0), all metrics achieved the best result. This also applies to DOM, although lower values in this case indicate a higher degree of control of the perceived emotions according to our 5-level scale. On the other hand, when packet transmission delay is introduced into the network, either continuously (NC1) or with variations (NC2), the achieved MOS results are comparable. Kruskal-Wallis test results indicate that for most of the metrics, the MOS achieved when no distortions are applied is significantly different from the MOS achieved when packet delay is introduced, regardless of the delay intensity. Specifically, this applies to QoE, VQ, AQ, GC, and COL with a $p < 0.001$, and to PRE and VAL with a $p < 0.01$. No significant difference is observed between the MOS achieved when one of the two network distortions is applied. Moreover, by computing the Pearson correlation coefficient (PCC) between the subjective results collected for the different metrics, it was found that a positive PCC exists between QoE and GC (0.61), QoE and COL (0.64), and QoE and VAL (0.63). A PCC of 0.58 is also found between COL and GC, whereas all the other PCC values were lower than 0.5.

Based on these results, we can state that the packet transmission delay impairments had a major negative influence on the visual and audio quality and on the player collaboration and efficiency to complete the game as well, and consequently on the perceived overall QoE. Sense of presence and valence

TABLE II: MOS for the subjective metrics as a function of both network quality and avatar.

Avatar	NC	QoE	VQ	AQ	GC	PRE	COM	COL	VAL	ARO	DOM
Chef	**NC0**	4.57	4.07	3.71	4.50	4.29	4.64	4.50	4.57	4.14	2.50
	NC1	4.00	3.79	3.36	3.64	3.93	4.37	4.29	4.43	3.93	3.07
	NC2	4.58	3.67	3.25	4.25	4.17	4.50	4.67	4.50	4.42	2.92
HST	**NC0**	4.31	3.92	3.85	4.15	4.31	4.62	4.46	4.69	4.08	2.69
	NC1	3.23	3.23	2.92	2.62	3.69	4.46	3.31	3.92	3.77	3.38
	NC2	3.29	3.29	2.64	2.57	3.79	4.57	3.14	3.79	3.36	3.07
HFE	**NC0**	4.69	4.08	3.85	4.38	4.23	4.69	4.49	4.85	4.00	2.85
	NC1	3.00	3.08	2.85	2.77	3.54	4.08	3.15	3.62	3.23	3.15
	NC2	3.50	3.29	2.50	3.00	3.71	4.21	3.07	3.93	3.79	2.79

were also affected, although to a lesser extent. In particular, the perceived QoE and valence were mostly reliant on a fruitful collaboration between the players, leading to successful game completion. In case of network distortions, the players were hindered by delays impairing their communication and collaboration, which resulted in lower QoE and valence.

B. Influence of avatar

Fig. 3b shows the MOS as a function of the 3 types of avatar: Chef, humanoid standard (HST), and humanoid reproducing facial expressions of the user (HFE). It can be seen that the Chef avatar achieved the highest MOS for each metric, whereas the MOS for the humanoid avatars are comparable. By computing the Kruskal-Wallis test, it was found that the MOS achieved for the Chef avatar is significantly higher than the MOS achieved for the humanoid avatars for GC with a $p < 0.001$, QoE and COL with a $p < 0.01$, and ARO with a $p < 0.05$. Also, it is interesting to note that, on average, when using the Chef avatar, the players were able to collaborate satisfactorily, complete the game, and perceive an optimal QoE even when network impairments were applied.

To investigate the reason why some metrics achieved a significantly lower MOS with humanoid avatars than the Chef avatar, we have computed the Kruskal-Wallis test on the subjective results collected for each avatar under different network conditions, whose MOS are in Table II. For the Chef avatar, only GC was rated significantly lower (with $p < 0.05$) when network distortions were applied. For the HST avatar, the MOS achieved when no distortions are applied is significantly higher than the MOS achieved when distortions are applied for GC with a $p < 0.001$, and for QoE, AQ, PRE, COL, and VAL with a $p < 0.05$. For the HFE avatar, significantly higher MOS are achieved for QoE, VQ, GC, and COL with a $p < 0.01$, and for AQ and VAL with a $p < 0.05$.

Thus, both the humanoid avatars suffer from the introduction of network distortions, likely because the exchange of data between the two players is larger than that occurring when the Chef avatar is used, since humanoid avatars are required to show body motions and facial expressions of the user. The packet transmission delay leads to a reduction of audio and video quality and collaboration efficiency, with consequent difficulty in completing game tasks and achieving sufficient QoE and valence. No significant differences are observed between HST and HFE avatars, although HST suffer a bit less from the introduction of network distortions. This may be because players using HST can see the facial expressions of partners using HFE, which may alleviate the reduction of user experience due to the game issues.

C. Influence of role

By computing the Kruskal-Wallis test, it was found that, for each considered metric, the subjective results achieved for the players interpreting the role of the teacher were not significantly different from those achieved for the players acting as students. Thus, the role did not influence the perceived user experience and emotions for this experiment.

VI. Conclusion

This study evaluated the user experience and emotions in a collaborative VR cooking game under the influence of network distortions (delay and jitter), type of avatar (cartoon-like human, humanoid, and humanoid equipped with facial expression capabilities), and social role (teacher and student).

The introduced packet transmission delay impairments had a major negative influence on the visual and audio quality of the players, leading to difficulties in collaboration for completing the game. As a consequence, the perceived overall QoE and valence (perceived pleasantness) were significantly reduced in the presence of such network distortions. However, further analysis on the influence of the avatar revealed that this QoE decrease occurred only when the humanoid avatars were used because of the larger amount of data required to replicate the facial expressions. The cartoon-like avatar, being lightweight, was more robust to the introduced network distortions. Nonetheless, the results also highlight that the players who were able to see the facial expressions of the partner's avatar achieved a lower reduction of user experience due to the network issues. This suggests that emotional non-verbal communication can contribute to enhancing the perceived QoE within VR game environments. Finally, no influence of the player's role was found on the user experience.

Future work will focus on further experimental studies, including the developed humanoid avatar replicating the facial expressions of the user, but in different VR contexts. The aim is to collect more data to validate the capability of non-verbal communication, such as facial expression and body gestures, to improve the perceived QoE in VR applications. Moreover, realistic avatars representing the physical aspect of the users will be considered.

REFERENCES

[1] A. Paszkiewicz, M. Salach, P. Dymora, M. Bolanowski, G. Budzik, and P. Kubiak, "Methodology of implementing virtual reality in education for industry 4.0," *Sustainability*, vol. 13, no. 9, p. 5049, 2021.

[2] S. K. Renganayagalu, S. C. Mallam, and S. Nazir, "Effectiveness of VR head mounted displays in professional training: A systematic review," *Technology, Knowledge and Learning*, pp. 1–43, 2021.

[3] H. M. Fornes, E. H. Birketvedt, C. Griwodz, M. Skjegstad, M. Welzl, and O. Alay, "Acceptable latency in predictable first-person vr cloud games," in *Proceedings of the 17th International Workshop on IMmersive Mixed and Virtual Environment Systems*, ser. MMVE '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 58–64. [Online]. Available: https://doi.org/10.1145/3712677.3720467

[4] S. Van Damme, J. Sameri, S. Schwarzmann, Q. Wei, R. Trivisonno, F. De Turck, and M. Torres Vega, "Impact of latency on QoE, performance, and collaboration in interactive Multi-User virtual reality," *Applied Sciences*, vol. 14, no. 6, p. 2290, 2024.

[5] H. S. Rossi, K. Mitra, S. Larsson, C. Åhlund, and I. Cotanis, "Subjective QoE Assessment for Virtual Reality Cloud-based First-Person Shooter Game," in *ICC 2024 - IEEE International Conference on Communications*, 2024.

[6] M. Warsinke, T. Kojić, M. Vergari, J.-N. Voigt-Antons, and S. Möller, "VR Cloud Gaming UX: Exploring the Impact of Network Quality on Emotion, Presence, Game Experience and Cybersickness," in *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2024, pp. 578–581.

[7] C. Cortés, I. Viola, J. Gutiérrez, J. Jansen, S. Subramanyam, E. Alexiou, P. Pérez, N. García, and P. César, "Delay threshold for social interaction in volumetric extended reality communication," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 7, pp. 1–22, 2024.

[8] R. D. Tripathi, M. Lyu, and V. Sivaraman, "Assessing the impact of network quality-of-service on metaverse virtual reality user experience," in *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*. IEEE, 2024, pp. 206–213.

[9] C. Pelachaud, "Studies on gesture expressivity for a virtual agent," *Speech Communication*, vol. 51, no. 7, pp. 630–639, 2009, research Challenges in Speech Technology: A Special Issue in Honour of Rolf Carlson and Björn Granström.

[10] J. Gutierrez, P. Perez, M. Orduna, A. Singla, C. Cortes, P. Mazumdar, I. Viola, K. Brunnström, F. Battisti, N. Cieplińska *et al.*, "Subjective Evaluation of Visual Quality and Simulator Sickness of Short 360 Videos: ITU-T Rec. P. 919," *IEEE transactions on multimedia*, vol. 24, pp. 3087–3100, 2021.

[11] K. Venkatraman, S. Raghuraman, Y. Tian, B. Prabhakaran, K. Nahrstedt, and T. Annaswamy, "Quantifying and improving user quality of experience in immersive tele-rehabilitation," in *2014 IEEE International Symposium on Multimedia*, 2014, pp. 207–214.

[12] Y. Kusunose, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Qoe assessment in networked air hockey game with haptic media," in *2010 9th Annual Workshop on Network and Systems Support for Games*, 2010, pp. 1–2.

[13] M. Sithu, Y. Ishibashi, P. Huang, and N. Fukushima, "Qoe assessment of operability and fairness for soft objects in networked real-time game with haptic sense," in *2015 21st Asia-Pacific Conference on Communications (APCC)*, 2015, pp. 570–574.

[14] D. Roberts, T. Duckworth, C. Moore, R. Wolff, and J. O'Hare, "Comparing the end to end latency of an immersive collaborative environment and a video conference," in *2009 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*, 2009, pp. 89–94.

[15] A. Becher, J. Angerer, and T. Grauschopf, "Negative effects of network latencies in immersive collaborative virtual environments," *Virtual Reality*, vol. 24, pp. 369–383, 2020.

[16] CVRcooking, https://github.com/mj-sam/CVR_cooking, accessed: 2025-04-03.

[17] Meta, https://developers.meta.com/horizon/documentation/unity/move-overview/, accessed: 2025-04-03.

[18] ——, https://developers.meta.com/horizon/documentation/unity/move-body-tracking/, accessed: 2025-04-03.

[19] ——, https://developers.meta.com/horizon/documentation/unity/move-eye-tracking/, accessed: 2025-04-03.

[20] ——, https://developers.meta.com/horizon/documentation/unity/move-face-tracking/, accessed: 2025-04-03.

[21] Unity, https://docs-multiplayer.unity3d.com/netcode/current/about/, accessed: 2025-04-03.

[22] ——, https://docs-multiplayer.unity3d.com/netcode/current/advanced-topics/message-system/rpc/, accessed: 2025-04-03.

[23] "Facial Expressions In A Multi-user Collaborative VR Game," https://github.com/Net4uCA/Facial-Expressions-In-A-Multi-user-Collaborative-VR-Game.

[24] S. Hemminger *et al.*, "Network emulation with NetEm," in *Linux conf au*, vol. 5. Citeseer, 2005, p. 2005.

[25] ITU, "Methods for subjective determination of transmission quality." Recommendation ITU-T P.800, 1996.

[26] ——, "Interactive test methods for audiovisual communications." Recommendation ITU-T P.920, 2000.

[27] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49 – 59, 1994.