Subjective QoE Assessment for Virtual Reality Cloud-based First-Person Shooter Game

Henrique Souza Rossi*, Karan Mitra*, Samuel Larsson*, Christer Åhlund*, Irina Cotanis†
*Pervasive and Mobile Computing, Department of Computer Science, Electrical and Space Engineering,
Luleå University of Technology.
†Infovista AB, Sweden.
henrique.souza.rossi@ltu.se

***Abstract*—Quality of experience (QoE) is an essential metric for stakeholders to understand how customers perceive the quality of their products or services. Gaming-as-a-Service (GaaS) is a challenging model to deliver efficiently to customers worldwide since it involves the joint force of cloud service providers, network operators, and game developers. The recent move of the cloud gaming (CG) industry to virtual reality (VR) platforms brings the benefits of the cloud to the most immersive quality of service (QoS) and QoE-sensitive VR content. Virtual reality cloud-based gaming (VRCG) necessitates understanding of stochastic broadband network connections on users' QoE so that stakeholders can deliver quality content by optimizing their services to underlying QoS conditions. Very few studies exist in the literature that study the impact of network QoS on users' QoE for VRCG. This paper presents subjective tests (N= 30) and investigates the effect of network-emulated QoS metrics (N=28) on the commercial Nvidia CloudXR service and their impact on the users' perceived QoE while playing Serious Sam VR shooter game. Our findings reveal that QoE was most affected by round trip time (RTT) $\geq$ 75 ms or packet loss (PL) $> 6\%$. Random jitter (RJ) caused QoE degradation for values more significant than one standard deviation, while the combined RTT and PL degraded QoE the most for RTT $\geq$ 25ms and PL $\geq 4\%$. Finally, based on actual network traffic data between Sweden and various data centers in Europe, we suggest VRCG can be hosted anywhere in these data centers with minimal impact on QoE for wired connections. However, for 4G and 5G networks, high jitter values could pose a challenge to VRCG services.**

***Index Terms*—Subjective tests, Quality of Experience (QoE), Virtual Reality (VR), Cloud Gaming (CG)**

I. INTRODUCTION

Extended reality (XR) is one of the axes that form Industry 4.0 and will shape future communication technologies [1]; it is a key enabling technology to build the digital metaverse where humans communicate, work, create, and consume immmersive content, merging cyber-physical worlds. VR glasses provide complete immersion in XR, as their design allows better control of the virtual experience and delivery with the highest video resolutions [2]. However, to realise the metaverse, the industry first requires understanding how each new and existing VR application or service performs and how they affect users' perceived QoE. The recent move of the 18 billion USD worth CG industry to VR platforms[1], brings a new type of gaming service to be explored within metaverse consumers by merging the advantages of the GaaS model (cost reduction, scalability, and energy efficiency), with fully immersive VR experiences. However, implementing VR cloud-based gaming (VRCG) will depend on the network infrastructure between users at home and cloud data centers. CG already requires stringent network requirements [3], which are likely to become stricter since VR technologies are sensitive to latency [4], jitter and packet loss.

To support VRCG services in current fixed and mobile broadband networks, stakeholders (e.g. network operators, cloud providers, and game developers) must learn how to optimize their infrastructures. Since the network layer (of the OSI model) is widely accessible by stakeholders and its QoS metrics are easily optimizable [5], it becomes imperative to study how these metrics will impact user's game experience or QoE. QoE is a subjective metric affected by context factors (hardware and software) under evaluation [6]. The current research in QoE for VRCG has so far focused on video metrics [7], and limited ranges for one-way delay [7], [8], RTT [9], and PL [9]. They have not considered realistic network conditions pertaining to stochastic mobile network behavior, including jitter (RJ) and the combination of RTT and PL. Further, there is a knowledge gap in understanding how VRCG services adapt to QoS metrics (considered only by [9] for open-source Air Light VR (ALVR) streaming). Most importantly, it is still unclear whether the latest broadband mobile networks, in particular 5G-SA, can deliver VRCG from QoE optics.

To address these challenges, this paper investigates the impact of actual network conditions (28 in total) RTT, PL, RJ and their combinations, in both up/down links, using a commercial VRCG streaming solution Nvidia CloudXR[2]. Our research aim is to answer the following question: *"How do mobile broadband network conditions affect VRCG services and the users' perceived QoE?"* For that, we invited 30 users to participate in subjective tests, playing the Serious Sam VR shooter game in the context of VRCG.

Major contributions of this paper are as follows:

1) To the best of our knowledge, we are the first to conduct subjective tests using a commercial VRCG solution Nvidia CloudXR and to investigate how its QoS metrics were affected by emulated network conditions.
2) We are the first to investigate the combined effect (RTT, RJ) and (RTT, PL) on QoE for VRCG users.

---

[1] https://www.uploadvr.com/xbox-cloud-gaming-quest-release-date/

[2] https://developer.nvidia.com/cloudxr-sdk

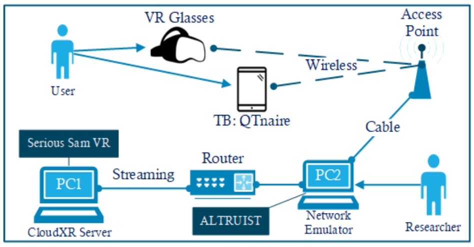


Fig. 1: The lab environment built to conduct subjective tests.

3) An assessment of whether current broadband mobile networks (wired, 4G, 5G-NSA, and 5G-SA) could possibly host VRCG in existing data centers.

This paper is organized as follows. Section II describes the technical settings and experiment design for subjective tests; Section III presents the results; Section IV discusses the novelty of this paper compared to existing research.

II. TESTBED AND SUBJECTIVE TESTS

In this section, we present the design of the VRCG experiment, including the tested conditions, as well as the technical details such as data collection and lab setup.

Laboratory Setup: The setup can be visualized in Fig. 1. It included two computers named PC1 (host the game and stream it) and PC2 to emulate the network conditions using NetEM[3]. PC2 handled the traffic between the CloudXR server and its thin client (VR Glasses), wirelessly connected to the Wi-Fi access point (AP). The baseline RTT present in the lab was on average 3 millisecond (ms). We used an Android tablet to show the questionnaires. Details of hardware specification are described in Table I. To reduce the risk of data loss, mislabeling, incorrect network emulation, and human errors, we use the open-source tool ALTRUIST [10] in PC2, to manage the tests. In particular, ALTRUIST was used to pass the correct parameters to NetEM for each condition tested.

TABLE I: Hardware used for conducting subjective tests.

Device	Description	Hardware
PC1	Hosting Serious Sam VR and CloudXR 3.2 Server	CPU i7 8700; RAM 32GB; SSD 2TB; GPU: Nvidia RTX 2070. Windows 10
PC2	Hosting ALTRUIST NetEM container and ALTRUIST ServerManager	CPU i7 8700; RAM 32GB; SSD 2TB; GPU: Nvidia RTX 2070; Ubuntu 22.04
VR Glasses	Hosting CloudXR thin Client	Oculus Quest 2. CPU SnapDragon XR2 RAM 6GB; SSD 128GB ; GPU Adreno 650; Anrdoid
Tablet	Hosting ALTRUIST questionnaire container	Samsung Galaxy Tab S3 9.7
Router	Routing all the lab network traffic	Netgear DG834
Wireless Access Point	Access Point for the Oculus Quest and Tablet	Asus RT-AX53U Wifi 6, 1800 mb/s 5Ghz.

Network Emulation: In total, 28 conditions were tested, divided into three independent variables (e.g., RTT, PL, combined (RTT, PL) and combined (RTT, RJ) listed in Table II. We first considered literature ranges [7]–[9] for RTT and PL. Then, we expanded their ranges and included the combinations for (RTT,PL) and (RTT,RJ) based on our realistic 4G measurements performed in Skellefteå [11], [12] and other cities [13], [14]. These ranges cover the heterogeneous behavior of mobile broadband networks, such as large events [14] and peak times [13]. Further, all conditions were applied in both up/down links (RTT values were halved for each link). The variability of RJ conditions follows a normal distribution.

TABLE II: Lab-emulated network conditions values.

Parameters	No. Conditions	Values
RTT	8	2,25,50,75,175,275,350,400 in ms
PL	3	6,12,24 in %
RTT and PL	8	RTT (25,50,75) ms and PL(2,4,6)%
RTT and RJ	9	RTT (25,50,75) ms and (1,3,6) std
Total	28	

Subjective Experiment Design: Each user test had a total duration of 1 hour and 30 minutes. At the start, users were asked to sign the consent form, followed by a demographic questionnaire. Then, they were invited to learn how to play the VR game for 5 min. In the testing phase, we followed a repeated measured (RM) design, where users were asked to play one game match under each tested network condition (N = 28). The network conditions were randomized, following a balanced Latin square design [15]. After completing a match, participants were asked to answer a questionnaire. Each match had a duration of 90 seconds, as suggested by ITU-T Rec. P809 [16], for game tests.

Data Collection and Questionnaire: During each user test, different sources of data were generated, including CloudXR streaming statistics logs, PCAP files from network traffic, and subjective responses from the questionnaire. The ALTRUIST tool [10] was used to copy and label each file according to the current test and user ID. The after-game-match questionnaire had six questions and covered questions about video/audio quality (1-5); overall QoE (1-5); cybersickness symptoms (multiple options checkbox); and controllability (1-5). We used a likert scale of 1 to 5 points where '1' = "very poor"; '2' = "poor"; '3' = "average"; '4' = "good"; and '5' = "very good".

Game and Experiment Settings: The VR game tested was Serious Sam VR[4], is classified within the first-person shooter genre, and has been investigated in the non-cloud context [17]. Enemies come in waves, and the player's goal is to shoot and kill as many enemies as possible to survive. The shooting is done by moving the VR controllers in the air. We choose a game in this genre since it is the most demanding in precision and sensitivity to players [18]. As such, we argue that by using this game , *we would automatically cover application factors, including precision, accuracy, and high response time, which may be or may not be available in conjunction with other less immersive VR content*. Finally, to ensure a repeatable VRCG experience within user tests and between users, Table III describes several strictly applied settings.

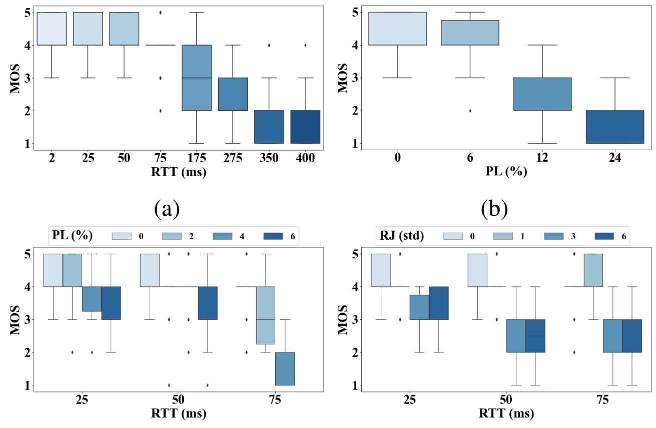
TABLE III: Experiment settings fixed during the tests.

Rule	Reason
VR Glasses Guardian turned off	Reduce interferences with video streaming performance
VR Glasses connected to portable battery	Reduce battery optimization interferences
User should stand-up while playing	Maintain similar immersive experiences between users
User should stay within determined area	Maximize signal strength with the Wi-Fi AP
Game Match level set to ”Temple” and weapon of choice set to ”laser pistols”	Maintain similar immersive experiences between users

---

[3] https://wiki.linuxfoundation.org/networking/netem

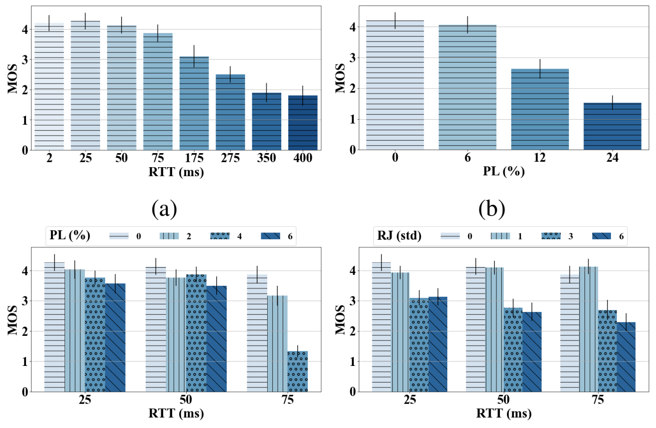
[4] https://store.steampowered.com/app/465240



(a) (b)

(c) (d)

Fig. 2: Box-plot for QoE votes distribution per condition.



(a) (b)

(c) (d)

Fig. 3: MOS for QoE question affected per condition.

III. RESULT ANALYSES

The results were grouped and explained according to the dependent variable studied and how it was affected by the lab-emulated network conditions (see Table II). In total, 30 users agreed to participate in our experiment. Their age was between 18-25 (N = 3) and 25-35 (N = 27). They reported playing games during a week between 1-5 hr. (N=12), 5-15 hr. (N=13), 15-25 hr. (N=4) and more than 25 (N=1). Most of the users (N = 26) reported having played VR games before, while all the users reported playing games on PC.

A. Network Conditions Impact on User's QoE

The distribution of the QoE question can be visualized in Fig. 2 divided into four subplots, each to show the effect of an independent variable. It shows that across all tested conditions, QoE score present consistent vote patterns by users, since the majority of votes per condition are between quartiles Q1 and Q3 and vary at most in one score difference (the width of each box). In these graphs, the small dots are observations outside the whisker line computed following the standard 1.5 interquartile range, which is a robust method to detect outliers [19]. The boxplot for the combined variables (RTT, PL) Fig. 2c and (RTT, RJ) Fig. 2d shows a slightly higher number of outliers. A more detailed analysis reveals that 3 users voted for at most 6 tests out of 28 (22%) marked as outliers, while remaining users voted for at most 3 votes as outliers (10%). Following the ITU-T P.913 [20] recommendation, we computed the Pearson correlation coefficient, to check the votes of the 3 users against the total MOS. Since the threshold for these users was $r > 0.75$, we decided not to exclude their votes and include them in the final MOS (N=30).

The MOS results can be visualized in Fig. 3, and each vertical line shows the confidence interval α=0.05 based on the normal distribution. Our analyzes assumed the data as interval and normality based on the central limit theorem [21]. The final MOS was calculated as the average opinion score of all users for the QoE question.

RTT or PL: For the RTT case (Fig. 3a), the data do not show a clear difference in MOS for RTT values smaller than 75 ms and RTT larger than 350ms. However, for RTT values between (75ms, 350ms), the MOS decreases as the RTT values increase. For the PL case (Fig. 3b), a clear change in MOS can be observed for PL greater than 6%, while there are unnoticeable differences in MOS for PL $\leq$ 6%.

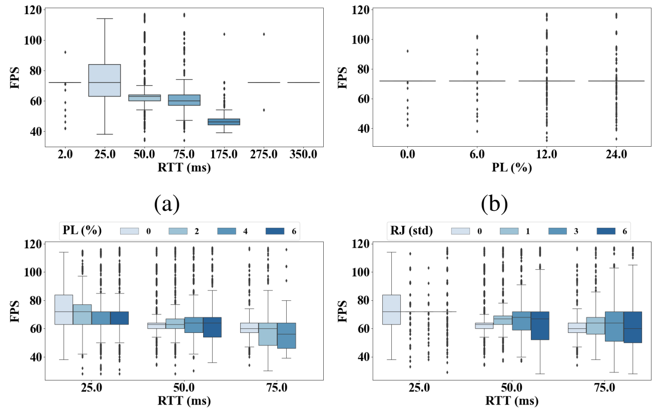
RTT and PL: In (Fig. 3c), a comparison between groups (RTT = 25ms, 50ms and 75ms) indicates that the groups RTT=25ms and 50ms, have a very similar impact on MOS regardless of the level of PL. The only noticeable difference is in the RTT = 75ms group, where each PL level (0%, 2%, and 4%) negatively impacts MOS the higher the PL values.

RTT and RJ: Regarding the levels of RJ (Fig. 3d), the MOS values are very similar for the pairs (0 std, 1 std) and (3 std, 6 std). On the contrary, there is a sharp decrease in MOS when comparing RJ values of 1 std and 3 std. These observations apply to all RTT groups. The results indicate that smaller values of RJ might not cause a perceptual difference to the users, while RJ $>$ 6 std should be further studied.

Cybersickness during tests: Cybersickness symptoms often occur when users interact with fully immersive technologies such as VR glasses [4]. We asked them to report symptoms they could have experienced during the tests. Based on this report, we identified that the most commonly chosen symptoms were "dizziness" (26 of 840 or 3%) and "fatigue" (2.3%). However, the choice of 'none' (91%) or the absence of symptoms was dominant among users in all conditions tested. These findings indicate that cybersickness occurred on a very small scale during the tests and the majority of users' tests were unaffected by it, suggesting that our methodology for conducting subjective tests had minimal impact on users.

B. Network conditions Impact on CloudXR QoS Metrics

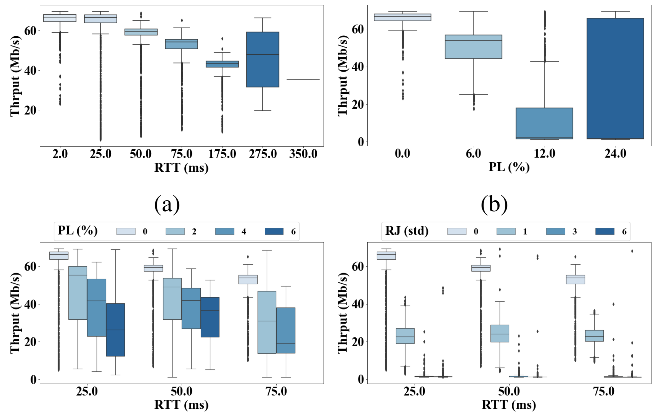
CloudXR generates streaming statistics for a variety of QoS metrics in the form of log files. Each user test (N = 30) resulted in 28 log files. The logs were sampled every second for the duration of the tests and were affected by each network



(a) (b)

(c) (d)

Fig. 4: CloudXR's frames per second (FPS) per condition.



(a) (b)

(c) (d)

Fig. 5: CloudXR's throughput in Mb/s per condition.

condition. The post-processing of the data resulted in 38,600 entries, and their distributions are depicted using boxplots for the metrics throughput in Fig. 5 and FPS in Fig. 4.

FPS: Without network degradations, the CloudXR target FPS value is set to the device limit of 72 Hz. The worse degradation of FPS occurred for the RTT tests (Fig. 4a), in particular, where RTT = 175ms causing a frame drop of 44% from the target FPS. Surprisingly, the least degradation occurred in the PL tests (Fig. 4b), where the FPS distribution was concentrated at 72 Hz, even with high levels of degradation ranging from PL values of 6% to 24%. This implies that CloudXR attempts to maintain a high FPS even in cases of heavy PL, possibly by changing bitrate [22]. The effect of combined tests (RTT, PL) on FPS (Fig. 4c) was smaller within varying PL values (a drop in 7% FPS), but greater between the RTT groups, where RTT = 75ms had the highest drop of 20%. Similarly, the effect of the combined test (RTT, RJ) on FPS (Fig. 4d) is greater between the RTT groups (up to 21% drop for RTT = 75ms) and smaller for varying RJ (up to 5% drop at RJ = 6std). The aforementioned results suggest that RTT is a major cause of FPS drops which in VR technologies affects video quality. and consequently impacts users' perceived QoE.

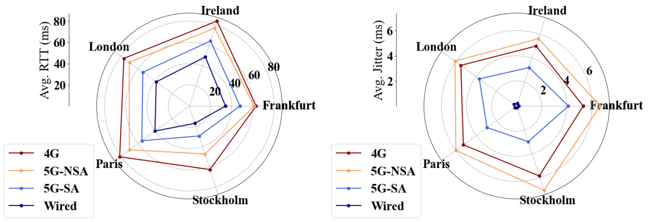
Throughput: We estimate from the baseline tests (RTT = 3ms), that CloudXR services send on average 65 Mb/s of data to the thin client (VR Glasses). Surprisingly, in the presence of jitter under conditions (RTT, RJ) (Fig. 5d), the service significantly reduces the throughput from 20 Mb/s at RJ = 1std, to a minimum of 3 Mb/s for RJ $\geq$ 3std, which is consistent regardless of the RTT group. It implies that the CloudXR service degrades the video quality in the presence of jitter, which explains why users rated the QoE question for RJ $\geq$ 3 very low. The second-worst reduction was in the PL (Fig. 5b) cases ranging from 13-25 Mb/s for PL = 12-24%, respectively. Similarly, PL affected throughput the most (Fig. 5c) for the combined (RTT, PL) tests regardless of the RTT group. The lowest throughput reduction occurred during the RTT tests (Fig. 5a). In addition, we have confirmed that there was indeed a throughput reduction in the magnitude of 10^{-1} based on PCAP files from these tests. As such, our findings suggest that CloudXR adapts to jitter and high PL by substantially reducing throughput.

C. *VRCG Network Assessment in Realistic Locations*

As a follow-up to our VRCG investigation, we now consider, from the network side, whether real-world traffic would impose a challenge should a VRCG be hosted in a distant location (cloud datacenters) from the user (e.g. their houses). For that we measured RTT (Fig. 6a) and jitter (Fig. 6b, from ICMP packets sent every second for one week, between the city of Skellefteå and five different AWS datacenters within Europe, under four broadband telecommunications standards (wired, 4G, 5G-NSA, 5G-SA) using a single major network operator. We argue, based on our reported QoE findings, that the following is true:

Wired connection: The worst-case latency scenario was between Skellefteå-Ireland (RTT = 48 ms, jitter $<$ 1 ms). These values fall within our tested condition (RTT = 75 ms, RJ = 1std) or RTT = 75 ms, and both have reported MOS = 4 (good). Regarding the best latency case (within Sweden) Skellefteå-Stockholm (RTT = 17ms, jitter $<$ 1ms), it is close to our tested (RTT = 25 ms, RJ = 1std) or RTT = 25 ms and both were reported with MOS = 4 (good). Therefore, for a wired broadband connection, even long-distance datacenters within Europe may not pose a significant challenge to VRCG.

4G and 5G NSA: The 4G worst-case latency scenario was between Skellefteå-Ireland (RTT = 84 ms, Jitter = 5 ms), while the best-case, Skellefteå-Stockholm (RTT = 65ms, Jitter= 6ms). These values are covered with our tested conditions for either the combined (RTT=75ms, RJ=6std) with MOS=2.3 (poor); or RTT conditions between 75ms and 175ms with MOS $>$ 3.8 (good) (see Fig. 3). We argue that in the presence of jitter, 4G worse-case can cause a moderate to heavy impact on QoE MOS, whereas for small jitter variation, it can cause a minimal to moderate impact. It can be seen that the 5G NSA values in Fig. 6 are slightly lower than 4G for RTT and slightly higher for jitter. Hence, we conclude that the reasoning for the 4G analysis still holds for the 5G NSA.



(a) (b)

Fig. 6: Real-world traffic measurements from a fixed location to multiple datacenters within Europe.

5G-SA: The measurements for 5G-SA worst-case latency was between Skellefteå-Ireland (RTT=62ms, Jitter=3ms) and Skellefteå-Stockholm (RTT=30ms, Jitter=3ms) in the best case. We consider the worst-case covered by our tested conditions (RTT=50ms, RJ=3std) MOS=2.8 (average), or RTT=50ms MOS=4 (good) while the best-case is close to the tested conditions (RTT=25ms,RJ=3std) MOS=3.1 (average) or RTT=25ms (good). These results suggest that 5G-SA can cause a moderate effect on QoE in the presence of jitter $\geq$ 3ms but minimal for jitter $<$ 3ms. Hence, 5G-SA performs better than 5G-NSA and 4G, worse than wired connections and it is still dependent on lower jitter to support VRCG.

IV. RELATED WORK

The state-of-the-art research in the field of VRCG has led to important contributions related to the impact of this media on QoS service metrics and QoE. Due to the nature of games requiring real/fast response from the service, we did not include research focused on VR 360 static videos.

VRCG Objective Evaluations: Several papers have so far studied VR offloading context. Maiorano *et al.* [23] and Jansen *et al.* [24] studied the effect of varying the distances between VR glasses and the access point (AP) on QoS metrics using CloudXR and Oculus Link, respectively. Distance has been found to affect both the bandwidth of the AP connection [23], [24] and increases network delay [23], while a higher frequency of the wifi signal (e.g., 2.4GHz vs 5GHz) improves bandwidth [24]. Further, Korneev *et al.* [25], focused on two different video codecs for VRCG (Nvidia Nvac vs. Pico Wireless codec), pointing out that NVAC (also included in CloudXR) is highly network optimized and produces overall better video streaming quality. When it comes to in-depth VRCG traffic analyses, Zhao *et al.* [22] demonstrated the importance of adaptive bitrate, to reduce frame latency and loss, while studying the Virtual Desktop service. When investigating the benefits of VR offloading, Nyamtiga *et al.* [26] suggest that offloading content reduces CPU and power consumption in VR headsets.

In summary, the aforementioned papers [22]–[26] provided knowledge by identifying and measuring different factors relevant to VRCG. However, since they have not studied QoS metrics from the network layer and their effect on QoE, our results will complement understanding of QoE for VRCG.

VRCG Subjective Evaluations: Kämäräinen *et al.* [8] were among the first to subjectively evaluate the VRCG scenario, using DayDream (deprecated) service. The authors focused on one-way delay and showed that users were more sensitive to latency (e.g. 200ms) when the game was rendered remotely vs. locally on the phone. Later, Li *et al.* [9] considered the early versions of the open-source VR cloud streaming ALVR, and studied the effect of varying bandwidth, RTT (5,25,45ms) and PL(0,2,4%) (applied in both links). Although these authors acknowledged ALVR limitations due to static bitrate, they highlight that high PL or limited bandwidth caused extensive video streaming blocking artifacts. Their results for the Half-Life-Alyx shooter game show no significant impact of RTT on MOS for RTT $\leq$ 45ms (similar to our results), but a considerable effect of PL=2% and 4%. In contrast, our tests shows minor impact on MOS for PL $\leq$ 6%. We account these differences due to the streaming service performance, and hence, a more thorough comparison of different VR cloud services is recommended. Lastly, Song *et al.* [27] studied the effect of VR video black-edge artifacts on users' QoE by simulating local latency (not in the network layer).

The most recent work proposed by *Lee et al.* [7] studied the effect of varying video metrics (FPS, video resolution, and bitrate) and one-way delay on users' QoE for three game genres (casual, platform, fast-peace) using ALVR. They found that the fast-peace game was the most sensitive to one-way delay, and that bitrate affects the gamer's QoE the most. We acknowledge the importance of studying the impact of video metrics on QoE for VRCG. However, in our research we instead focused on network metrics applied in both links, *that are easily accessible to network operators and cloud providers and can be easily optimized regardless of the content [5].*

In summary, we show that the current research in QoE for VRCG [7]–[9], [27] did not consider important QoS metrics, including jitter and the combination of (RTT,PL) pertaining to the stochastic behavior of broadband networks [13], [14] nor how VRCG services adapts to these degradations.

Comparison of Cloud vs non-cloud use case: Since Vlahovic *et al.* [17] studied the same VR game title as ours, but in a non-cloud context, we compared our VRCG findings with non-cloud VR for the impact of RTT conditions on MOS. In this condition, they reported MOS $\geq$ 4 for RTT $\geq$ 200 and MOS between 3 and 4 for 200 $<$ RTT $<$ 300. On the contrary, our results were on average 1 MOS score below for similar conditions, and under the hypothesis that experiment setups did not influence the results, we argue that VRCG could be more sensitive than VR. Indeed, we observed the same difference between CG and non-cloud based on state-of-the-art comparison [11].

In conclusion, this section presented the latest research in the VRCG domain and shows that our work is novel and contributes to this field by studying an extensive range of network conditions never considered before in the literature, their impact on both the CloudXR service, user's perceived QoE, and whether current broadband network (4G, 5G, and wired) can deliver VRCG.

V. CONCLUSION AND FUTURE WORK

This paper presents the first investigation of the commercial service Nvidia CloudXR from users' QoE perspective by conducting subjective tests. The following findings are summarized:

- MOS was clearly affected by RTT $\geq$ 75 ms or PL $>$ 6%. In the combined conditions (RTT, PL), MOS was significantly affected by RTT $\geq$ 25 ms and PL $\geq$ 4%. Lastly, for the combination of (RTT, RJ), RJ $>$ 1 std and RTT $>$ 25 ms caused the greatest MOS degradation.
- Analysis of the impact of network conditions on the CloudXR service indicates that throughput was mainly affected by all jitter ranges and PL $\geq$ 12%. FPS has been negatively affected the most by RTT (up 44% drop) and, surprisingly, the least by PL.
- We suggest that VRCG can be hosted within Europe dataceters (when clients are in Europe), with a minor impact on the MOS for wired connections. For 4G / 5G NSA connections, the impact would be severe in the presence of high jitter $\geq$ 6ms and minimal to moderate for jitter $\leq$ 2ms. 5G-SA performs better than 4G/5G NSA but worse than wired connections, providing good performance for jitter $\leq$ 2ms and average for jitter $>$ 2ms.
- When comparing our results with the state-of-the-art for similar RTT conditions and same game title, it indicates that VRCG is more sensitive to latency than VR gaming. Further, by comparing different VRCG services and similar games, we noticed significant differences for the PL conditions, which should be further investigated.

Our next step will be to partake in ITU-T SG-12 activities and assess how QoE for VRCG can be predicted based on network conditions.

REFERENCES

[1] L. Adriana Cárdenas-Robledo, Hernández-Uribe, C. Reta, and J. Antonio Cantoral-Ceballos, "Extended reality applications in industry 4.0. – A systematic literature review," *Telematics and Informatics*, vol. 73, p. 101863, Sep. 2022.

[2] J. Jerald, *The VR Book: Human-Centered Design for Virtual Reality*. ACM and Morgan & Claypool, Sep. 2015.

[3] O. S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco, "Measuring Key Quality Indicators in Cloud Gaming: Framework and Assessment Over Wireless Networks," *Sensors*, vol. 21, no. 4, p. 1387, Feb. 2021, publisher: MDPI AG.

[4] J.-P. Stauffert, F. Niebling, and M. E. Latoschik, "Latency and Cybersickness: Impact, Causes, and Measures. A Review," *Frontiers in Virtual Reality*, vol. 1, p. 582204, Nov. 2020.

[5] X. Marchal, P. Graff, J. R. Ky, T. Cholez, S. Tuffin, B. Mathieu, and O. Festor, "An Analysis of Cloud Gaming Platforms Behaviour Under Synthetic Network Constraints and Real Cellular Networks Conditions," *Journal of Network and Systems Management*, vol. 31, no. 2, p. 39, Feb. 2023.

[6] K. Mitra, A. Zaslavsky, and C. Ahlund, "Context-Aware QoE Modelling, Measurement, and Prediction in Mobile Computing Systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 920–936, May 2015.

[7] K.-Y. Lee, J.-W. Fang, Y.-C. Sun, and C.-H. Hsu, "Modeling Gamer Quality-of-Experience Using a Real Cloud VR Gaming Testbed," in *Proceedings of the 15th International Workshop on Immersive Mixed and Virtual Environment Systems*. Vancouver BC Canada: ACM, Jun. 2023, pp. 12–17.

[8] T. Kämäräinen, M. Siekkinen, J. Eerikäinen, and A. Ylä-Jääski, "CloudVR: Cloud Accelerated Interactive Mobile Virtual Reality," in *Proceedings of the 26th ACM international conference on Multimedia*. Seoul Republic of Korea: ACM, Oct. 2018, pp. 1181–1189.

[9] Y. C. Li, C. H. Hsu, and C. H. Hsu, "Performance Measurements on a Cloud VR Gaming Platform," in *QoEVMA 2020 - Proceedings of the 1st Workshop on Quality of Experience (QoE) in Visual Multimedia Applications*, vol. 20. New York, NY, USA: Association for Computing Machinery, Inc, Oct. 2020, pp. 37–45.

[10] H. S. Rossi, K. Mitra, C. Åhlund, I. Cotanis, N. Ögren, and P. Johansson, "ALTRUIST: A Multi-platform Tool for Conducting QoE Subjective Tests," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, Jun. 2023, pp. 99–102, iSSN: 2472-7814.

[11] H. S. Rossi, N. Ögren, K. Mitra, I. Cotanis, C. Åhlund, and P. Johansson, "Subjective Quality of Experience Assessment in Mobile Cloud Games," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Dec. 2022, pp. 1918–1923.

[12] H. S. Rossi, K. Mitra, C. Åhlund, I. Cotanis, N. Ögren, and P. Johansson, "Objective QoE Models for Cloud-based First Person Shooter Game over Mobile Networks," in *Accepted paper for IEEE 21th Consumer Communications & Networking Conference (CCNC)*, Jan. 2024.

[13] D. Baltrunas, A. Elmokashfi, and A. Kvalbein, "Measuring the Reliability of Mobile Broadband Networks," in *Proceedings of the 2014 Conference on Internet Measurement Conference*. Vancouver BC Canada: ACM, Nov. 2014, pp. 45–58.

[14] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A first look at cellular network performance during crowded events," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1, pp. 17–28, Jun. 2013.

[15] D. C. Howell, *Statistical methods for psychology*, 7th ed. Australia : Belmont, CA: Thomson Wadsworth, 2010, oCLC: ocn321045335.

[16] I.-T. ITU-T Recommendation P809, "Subjective evaluation methods for gaming quality," 2018.

[17] S. Vlahovic, M. Suznjevic, and L. Skorin-Kapov, "Challenges in Assessing Network Latency Impact on QoE and In-Game Performance in VR First Person Shooter Games," in *2019 15th International Conference on Telecommunications (Con)*, Jul. 2019, pp. 1–8.

[18] M. Claypool and D. Finkel, "The effects of latency on player performance in cloud-based games," in *2014 13th Annual Workshop on Network and Systems Support for Games*. Nagoya, Japan: IEEE, Dec. 2014, pp. 1–6.

[19] Tukey, John W and others, *Exploratory data analysis*. Reading, MA, 1977, vol. 2.

[20] ITU-T Reomendation P.913, "Recommendation ITU-T P.913 (06/2021) Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment," 2021.

[21] S. G. Kwak and J. H. Kim, "Central limit theorem: the cornerstone of modern statistics," *Korean Journal of Anesthesiology*, vol. 70, no. 2, p. 144, 2017.

[22] S. Zhao, H. Abou-zeid, R. Atawia, Y. S. K. Manjunath, A. B. Sediq, and X.-P. Zhang, "Virtual Reality Gaming on the Cloud: A Reality Check," in *2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021, pp. 1–6.

[23] G. Maiorano, G. P. Mattia, and R. Beraldi, "Local and Remote Fog based Trade-offs for QOE in VR Applications by using CloudXR and Oculu Air Link," in *2022 International Conference on Edge Computing and Applications (ICECAA)*, Oct. 2022, pp. 95–101.

[24] M. Jansen, J. Donkervliet, A. Trivedi, and A. Iosup, "Can My WiFi Handle the Metaverse? A Performance Evaluation Of Meta's Flagship Virtual Reality Hardware," 2023.

[25] E. S. Korneev, M. V. Liubogoshchev, and E. M. Khorov, "Studying Cloud-Based Virtual Reality Traffic," *Journal of Communications Technology and Electronics*, vol. 67, no. 12, pp. 1500–1505, Dec. 2022.

[26] B. W. Nyamtiga, A. A. Hermawan, Y. F. Luckyarno, T.-W. Kim, D.-Y. Jung, J. S. Kwak, and J.-H. Yun, "Edge-Computing-Assisted Virtual Reality Computation Offloading: An Empirical Study," *IEEE Access*, vol. 10, pp. 95 892–95 907, 2022.

[27] J. Song, X. Mao, and F. Yang, "The Impact of Black Edge Artifact on QoE of the FOV-based Cloud VR Services," *IEEE Transactions on Multimedia*, pp. 1–16, 2022.