Differential QoE in Picture-in-Picture Gaming Videos: A Subjective Study

Tomasz Lyko*, Yehia Elkhatib†*, Rajiv Ramdhany‡‡ and Nicholas Race*
*School of Computing and Communications, Lancaster University, United Kingdom, {i.lastname}@lancaster.ac.uk
†School of Computing Science, University of Glasgow, United Kingdom ‡‡BBC R&D North, United Kingdom

***Abstract*—Video streaming continues to be the largest service delivered on the internet. This includes gaming videos, delivered both on-demand and live, where gaming footage is usually accompanied by a video of the player overlaid on top of the gameplay – resulting in Picture-In-Picture (PiP) content. Currently, PiP content is usually combined into a single video before being delivered to the client via technologies such as HTTP Adaptive Streaming (HAS). In this study, we investigated the QoE importance of gameplay and player elements in PiP gaming videos by varying the video quality of these elements individually. We conducted a subjective study, testing nine quality permutations based on three quality levels across three pieces of content from different gaming genres, with 30 participants recruited using an ethical crowdsourcing platform. We found that gameplay was significantly more important in terms of overall QoE, while the player element made a difference in only a few cases.**

***Index Terms*—QoE, assessment, subjective study, Picture-In-Picture, gaming, videos**

I. Introduction

Gaming videos are becoming increasingly popular, both as live events and on-demand content. This is made possible by many platforms, the most prominent of which are Twitch and YouTube Live. The current global audience is estimated at 1 billion, and revenue at US$11.7 billion [1]. These figures are expected to grow to 1.65 billion users and US$17.4 billion, respectively, by 2027 [2]. Such video content typically involves more than one element, such as the gameplay video and an accompanying commentary and reaction video/audio stream where the player comments on the gameplay.[1] This is referred to as Picture-in-Picture (PiP) content.

To date, most PiP content has been delivered as one stream that consolidates two (or more) elements. However, these elements are of categorically different types; e.g. vector-based graphics in an artificial world versus video capture of a talking head. There is growing interest in having more flexibility in the way these elements are merged together, specifically towards doing so in distributed rather than cloud-centric fashion [3]. However, the effect of network transmission artefacts on either stream would be different due to their contrasting natures. This raises several open questions:

- When consolidating more than one stream of different types, what prioritization, if any, needs to be performed to optimize the overall Quality of Experience (QoE)?
- Is this prioritization content-sensitive?
- Can better QoE at the same bitrate be achieved by deploying Perceptual Video Compression (PVC) with Region-Of-Interest (ROI) encoding for PiP gaming content?

As such, we are motivated to explore how adapting video delivery to the achievable Quality of Service (QoS) translates to QoE perceived by viewers. In this paper, we present the results of our subjective study, designed to assess the QoE importance of gameplay and player elements in PiP gaming videos. We varied the video quality of these elements individually, resulting in 9 quality variations per 3 quality levels tested across 3 pieces of content. In total, 27 test sequences were assessed by 30 participants recruited using an ethical crowdsourcing platform.

[1] Additional streams such as text chat could also be included.

Fig. 1. Thumbnails of the clips viewed by the experiment participants. From the left, these are Forza, Red Dead Redemption 2 (RDR2), and PGA.

TABLE I
A SUMMARY OF THE PROPERTIES OF THE CONTENT USED IN OUR STUDY.

Clip	SI	TI	Player Size and Position
Forza	77.3	37.2	350x370, middle left
RDR2	47.7	22.3	394x422, top right
PGA	56.0	32.8	356x380, top left

II. Background

QoE of gaming video streams has been studied in different works, both in the academic literature (e.g., [4], [5]) and standardization efforts (e.g. [6], [7]). However, none looked into the effect of the quality of the *individual streams* on the overall QoE, as perceived by the user.

In the following, we give a brief overview of the adjacent research. Pires and Simon [4] as well as Bilal and Erbad [8] investigated transcoding strategies (specifically regarding resolution and bitrate) for Twitch feeds to optimize cloud resource utilization and viewer QoE. Wahab et al. [9] studied the effect of the packet loss rate on gaming streams of different genres. Madanapalli et al. [10] devised a machine learning method to detect gaming streams and deduce QoE-affecting network events.

III. Methodology

Our subjective study was conducted using the Absolute Category Rating (ACR) method, outlined in ITU Rec. P.910 [11]. In ACR, participants are shown one test sequence at a time and asked to score it individually immediately after presentation. The standard 5-point Mean Opinion Score (MOS) scale was used: *Bad*, *Poor*, *Fair*, *Good*, and *Excellent*.

A. Content and Encoding

We sourced 3 '*let's play*' videos from YouTube, each containing gameplay footage along with player footage overlaid on top, as shown in Figure 1. Table I describes the selected content in terms of the calculated Spatial Information (SI) and Temporal Information (TI) according to ITU P.910 [11], as well as the size and position of the player overlay. Each video was ten seconds long, and had greater bitrate and resolution than our highest tested quality setting.

To vary the quality of the gameplay and player independently, we extracted player footage as a separate clip using ffmpeg[2] and its `crop` filter. Gameplay and player clips were then encoded into the desired quality levels using ffmpeg and H.264 codec, after which the player clip was overlaid back on top of the gameplay clip in its original position, using ffmpeg's `overlay` filter.

Three quality levels were used: low, medium, and high, corresponding to the following bitrates: 730, 2000, and 6000 kbps, respectively, based on the recommended specifications for Apple devices [12], with a video resolution of 1920x1080 at 60 FPS. These settings were scaled down for the Player clips to match their much smaller screen resolution, using the ratio of Player to Gameplay in terms of resolution. This resulted in 9 possible quality permutations of gameplay and the player.

B. Experiment Set-up

We created an online survey capable of displaying videos without rebuffering or unintended quality degradation. Test sequences were fully fetched before playback and were always played in full-screen mode with sound on. Each participant watched all the test sequences, 27, arranged in a pre-generated random order with one constraint: the same clip was never shown twice in a row. Additionally, 3 training sequences, based on different content and showing a subset of the tested quality variations, were shown before the main set of test sequences to allow the participants to become accustomed to the scoring scale. The survey also incorporated a mechanism to detect whether the participant had watched all of the test sequences completely before rating them. The experimental setup was approved by our ethics advisory board.

C. Participant Recruitment

Participants were recruited using the ethical crowdsourcing platform Prolific[3]. We recruited 30 participants: 16 male, 13 female, and 1 other, with a median age group of 29-38. 83% of the participants reported playing video games, and 73% reported watching videos of people playing video games. The participants were required to complete the survey on a desktop or laptop with a screen resolution of 1920x1080 – the video resolution of our test sequences.

IV. Results and Discussion

Figure 2 shows the MOS of each test sequence, with a Confidence Interval of 95% plotted in the error bars, with the top chart showing the average MOS across all 3 clips tested. Each test sequence contained a different variation of video quality of the gameplay and player elements. Paired T-Test ($\alpha = 0.05$) was used to determine the statistical significance

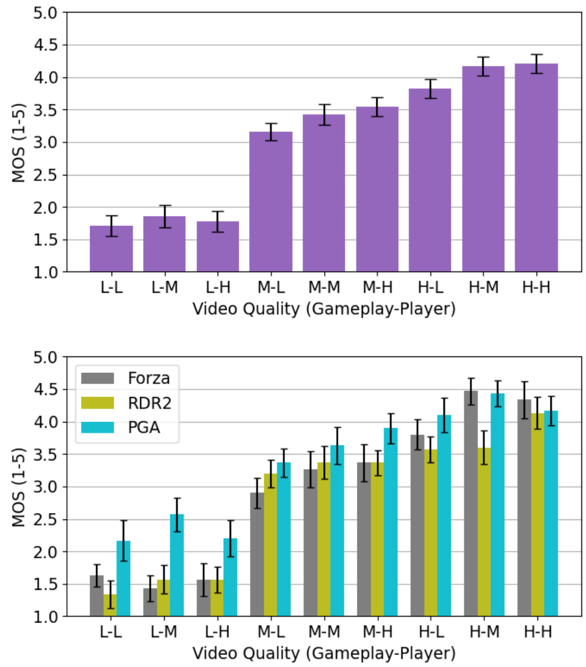

Fig. 2. Results of the subjective study, with Mean Opinion Scores (MOS) plotted for each video quality permutation of gameplay and player. The top chart shows the average MOS across all content, while the bottom chart shows the individual MOS values for each test sequence.

of differences between MOS of test sequences. All 27 test sequences were watched and rated by all the 30 participants.

A. Gameplay Importance

When the video quality of the gameplay element improved and the player quality remained constant, the observed MOS significantly improved in the majority of cases across all the three clips tested. When the quality of the player element was low, the increase in gameplay quality from low to medium and from low to high resulted in 1.44 and 2.11 respectively better MOS on average across the three clips tested. In the case of medium quality player, the improvement in gameplay quality resulted in 1.57 and 2.31 better MOS for medium and high quality respectively, when compared to low quality. When the player quality was high, the improvement in gameplay quality from low to medium and from low to high resulted in 1.77 and 2.43 better MOS respectively.

The improvement in MOS varied across the three tested clips, with an average improvement of 2.14, 2.05, and 1.62 for Forza, RDR2, and PGA clips respectively. All three low quality gameplay variants of the PGA clip were rated better than Forza and RDR2, on average achieving 0.77 and 0.82 higher MOS respectively, explaining the observed lower rate of improvement in MOS when the gameplay quality increased from the lowest level. While the PGA clip had lower Spatial and Temporal Information than Forza, it was still higher than RDR2, making this observation unexpected, as videos with lower Spatial and Temporal Information typically result in better QoE at the same bitrate owing to better compression efficiency.

[2]https://ffmpeg.org/
[3]https://www.prolific.co/

B. Player Importance

When the video quality of the player element alone increased, there was a marginal improvement in the observed MOS; however, most differences were statistically insignificant. In the case where the gameplay quality was low, there was only one statistically significant improvement in MOS: for the PGA clip, player at medium quality was rated 0.4 better than at low quality. For gameplay at medium quality, the MOS for Forza clip improved by 0.37 and 0.47 when the player quality increased from low to medium and low to high respectively. For the PGA clip, player at high quality resulted in 0.53 better MOS than those at low quality. The remaining differences at this gameplay quality were statistically insignificant. In the case of high-quality gameplay, only the following differences were statistically significant. For the Forza clip, the MOS improved by 0.67 and 0.53 for medium and high quality variations of the player respectively. The RDR2 clip with high quality player was scored 0.57 higher than the low quality player version.

The improvement in MOS varied slightly between the tested clips, with averages of 0.29, 0.23, and 0.27 for the Forza, RDR2, and PGA clips, respectively. In the case of the PGA clip, there were two cases in which the medium quality variation of the player scored higher than the high quality. When the gameplay was at low quality, medium quality player was scored 0.37 better, and at high quality, medium quality player was scored 0.27 better. Both of these differences are statistically significant. This observation was unexpected and was only present in this specific clip.

For all three clips tested, the player element varied slightly in terms of size and position, as shown in Table I. The player element for the RDR2 clip was approximately 26% larger when compared to the other two clips, but it did not result in greater improvement in MOS when only the video quality of the player improved. However, this observation is not clear, as a complete comparison is not possible because most differences were statistically insignificant. This also applies to the position of the player element, which appears to have no impact on the MOS improvement.

C. Overall Importance

From the results presented above, we can observe that the gameplay element is significantly more important in terms of QoE. Improvement in the gameplay's video quality resulted in a 1.94 better MOS on average, with all differences being statistically significant, while the increased video quality of the player element resulted in only a marginal improvement of 0.27 on average, with most of the differences being statistically insignificant. This suggests that the gameplay element should be prioritized to achieve optimal QoE. Additionally, in some cases, improving the video quality of the player element increased the QoE, especially when the gameplay was already of medium or high quality. However, more work is needed to determine when increasing the player quality is optimal.

V. Conclusion

In this paper, we presented the results of a subjective study designed to investigate the QoE importance of gameplay and player elements in Picture-In-Picture gaming videos. The study included 9 quality permutations of gameplay and player based on 3 quality levels using three different pieces of content, resulting in a total of 27 test sequences. All test sequences were watched and rated by 30 participants on a standard 5-point MOS scale, following the ACR method.

We found that gameplay was significantly more important in terms of overall QoE, with an increase in video quality resulting in 1.94 better MOS on average. An increase in the video quality of the player element offered little to no improvement in QoE – only 0.27 better MOS on average. However, in a few cases, increasing the video quality of the player achieved a significant improvement in QoE, especially when the gameplay was already of acceptable quality. Further studies are needed to determine when a player element becomes significant in terms of QoE. The results of this study could be used to improve video compression of Picture-In-Picture gaming videos in scenarios where Perceptual Video Compression (PVC) with Region-Of-Interest (ROI) encoding is applied, or when content is delivered as independent objects, i.e., Object Based Media (OBM) [13].

Acknowledgment

This research was supported by UKRI EPSRC and BBC Prosperity Partnership AI4ME: Future Personalised Object-Based Media Experiences Delivered at Scale Anywhere EP/V038087.

References

[1] L. Chiovato, "Gaming's live-streaming audience will hit one billion next year & 1.4 billion by 2025," May 2022. [Online]. Available: https://newzoo.com/resources/blog/gamings-live-streaming-audience-will-hit-one-billion-next-year-1-4-billion-by-2025

[2] Statista, "Games live streaming - worldwide," Apr. 2023. [Online]. Available: https://www.statista.com/outlook/amo/media/games/games-live-streaming/worldwide

[3] K. Krewell, "AMD builds breakthrough AV1 encoder chip for massive streaming services," https://www.forbes.com/sites/tiriasresearch/2023/04/06/amd-builds-breakthrough-av1-encoder-chip-for-massive-streaming-services/, Apr. 2023.

[4] K. Pires and G. Simon, "DASH in twitch: Adaptive bitrate streaming in live game streaming platforms," in *Workshop on Design, Quality and Deployment of Adaptive Video Streaming (VideoNext)*. ACM, 2014, p. 13–18.

[5] F. Metzger, S. Geißler, A. Grigorjew, F. Loh, C. Moldovan, M. Seufert, and T. Hoßfeld, "An introduction to online video game QoS and QoE influencing factors," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1894–1925, 2022.

[6] S. Zadtootaghaj, S. Schmidt, and S. Möller, "Influence factors on gaming quality of experience (QoE)," https://handle.itu.int/11.1002/1000/13396, Tech. Rep. ITU-T G.1032, Oct. 2017.

[7] S. Schmidt, S. Zadtootaghaj, and S. Möller, "ITU-T standardization activities targeting gaming quality of experience," *SIGMultimedia Rec.*, vol. 13, no. 1, dec 2022.

[8] K. Bilal and A. Erbad, "Impact of multiple video representations in live streaming: A cost, bandwidth, and QoE analysis," in *IEEE International Conference on Cloud Engineering (IC2E)*, 2017, pp. 88–94.

[9] A. Wahab, N. Ahmad, and J. Schormans, "Variation in QoE of passive gaming video streaming for different packet loss ratios," in *International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–4.

[10] S. C. Madanapalli, A. Mathai, H. H. Gharakheili, and V. Sivaraman, "ReCLive: Real-time classification and QoE inference of live video streaming services," in *IEEE/ACM International Symposium on Quality of Service (IWQOS)*, 2021, pp. 1–7.

[11] ITUTP, "P.910: Subjective video quality assessment methods for multimedia applications," Tech. Rep. Recommendation ITU-T P.910, 1999.

[12] Apple, "HTTP Live Streaming (HLS) Authoring Specification for Apple Devices," 2023. [Online]. Available: https://developer.apple.com/documentation/http_live_streaming/http_live_streaming_hls_authoring_specification_for_apple_devices

[13] T. Lyko, Y. Elkhatib, M. Sparks, N. Race, and R. Ramdhany, "QoE assessment for multi-video object based media," in *International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, Sep. 2022.