The Double-Edged Impact of User Customisation on QoE in Personalised Media Experiences

Tomasz Lyko*, Edward Austin*, Alexander Lee*, Yehia Elkhatib†*, 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

***Abstract*—User-driven experience customisation can potentially enhance the Quality of Experience (QoE) in personalised multimedia. Such experiences could be, for example, delivered using HTTP Adaptive Streaming - the most prominent way of consuming media over the Internet. In this paper, we present a subjective study designed to investigate the QoE impact of user-driven customisation in personalised media experiences. We offered participants the option to customise their video layout, after which we asked them to score a range of quality impairments found in HTTP Adaptive Streaming. Based on our analysis of the collected Mean Opinion Scores (MOS), we found that experience customisation impacts the QoE in a surprising way. We found that experience personalisation caused participants' expectations to increase when it comes to QoE, as they perceived the most severe quality impairments worse than the control. This establishes the need for specialised QoE models that take into account different levels of user expectations.**

***Index Terms*—QoE assessment, subjective study, Picture-In-Picture, videos, experience customisation, personalisation**

I. Introduction

Multimedia contains a broad spectrum of experiences, with video streaming over the Internet being the most prominent, delivered using HTTP Adaptive Streaming standards such as DASH [1] or HLS [2]. This includes both on-demand and live streaming scenarios. Quality of Experience (QoE) plays an important role in HTTP Adaptive Streaming, modelling the user's perception of a streaming session, including factors such as video quality, changes in quality, and stalling. Such QoE models are used in the design of Adaptive Bitrate (ABR) algorithms, which monitor network conditions and adapt the video quality bitrate to provide the best QoE possible given the bandwidth available.

Recently, the attention in multimedia has been shifting towards flexible media, such as Object-Based Media (OBM) [3]. In OBM, individual components of a multimedia stream can be be transmitted to and combined at the client side. This opens the door to user-driven customisation of media experiences. With it comes the potential for personalisation, which adds a new dimension to QoE, leading to the following research questions:

- `RQ1`: How does user-driven customisation impact the QoE in personalised media experiences?
- `RQ2`: Furthermore, does user-driven personalisation make quality impairments such as video quality changes and rebuffering less perceptible?

We constrain our exploration to Picture-In-Picture (PiP) gaming content, also known as "let's play" videos, commonly found in live streaming platforms such as Twitch and YouTube. PiP gaming videos consist of two main elements, gameplay and an accompanying player video/audio stream where the person playing the game provides their commentary. Usually, both of these elements are first combined and then transmitted as a single AV stream using DASH or HLS. With OBM, these two elements could be transmitted as separate layers, allowing for customisation at the client.

In this paper, we present the results of a subjective study designed to answer our research questions. We offered participants experience personalisation by enabling them to choose their preferred video layout, allowing them to place the Player element in 9 positions around the screen or to completely hide it, resulting in 10 personalisation options. Then we showed the participants 9 test conditions based on the common quality impairments found in HTTP Adaptive Streaming around video quality and stalling. The video layout used to show the quality impairments was based on the group assigned to participants and their selected preference. This is the first work to look into the QoE impact of user-driven customisation in personalised media experiences.

II. Background

The flexibility afforded by OBM has engendered significant interest in producing media experiences that adapt to the user's physical context e.g."The Living Room of the Future" [4] or personal preferences [5], [6]. By deferring media object composition to the point of user-consumption, OBM offers a degree of customisability which allows media experiences to be tailored to fit the preferences of individual users [7], [8].

Streaming platforms such as YouTube and Twitch have sky rocketed the popularity of PiP content. It encompasses a whole host of "let's play" videos, where a content creator streams or records their experience playing a video game. With gaming being consumed by an estimated 1 billion audience members creating revenue valued at US$11.7 billion [9] and expected to increase to 1.65 billion and US$17.4 billion by 2027 [10].

While comprehensive literature exists on the QoE in HTTP Adaptive Streaming [11], that investigates various QoE factors such video quality changes [12]–[16] and rebuffering [17]–[20], there is no literature on the QoE impact of user-driven personalisation in media experiences. Inspired by the study by Norton et al. [21] on the IKEA Effect, which refers to a correlation between a consumer putting effort into a product and their perceived value of said product, we set out to

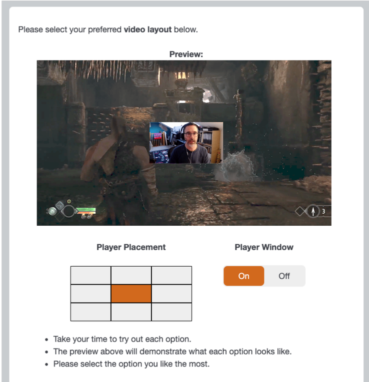

Fig. 1: The layout selection page used by participants to personalise their experience, with a live preview showing the currently selected option.

Fig. 2: All possible permutations of video layout that can be selected by participants.

investigate the QoE impact of user-driven customisation in personalised media experiences.

III. Methodology

We conducted a subjective study to investigate the QoE impact of user-driven experience customisation in personalised media. In this section, we detail our methodology. First, we give an overview of our study's design, including the personalisation options offered to participants, and the quality impairments tested. Second, we describe the generation process for the test sequences, based on the quality impairments under investigation. We then describe our online survey tool, specifically developed to show participants the test sequences without any unintentional quality impairments. Finally, we report on the recruitment process and demographic characteristics of our participants.

A. Study Design

In order to test the QoE impact of experience customisation, we offered some of the participants the option to personalise their experience by selecting their preferred video layout for a Picture-In-Picture gaming video, as seen in Figure 1. This layout selection page included a live preview of the currently selected option, with two looped GIF images representing the gameplay and player elements, where the player was overlaid on top of the gameplay in the desired location or completely hidden if the player window was selected to be "Off". In total, there were 10 possible permutations of video layout as seen in Figure 2. After layout selection, participants were asked to score test sequences that showcased a variety of quality impairments that can be found in adaptive video streaming using a similar Picture-In-Picture gaming video.

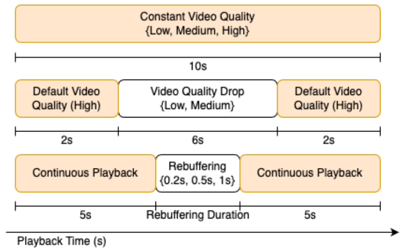

Fig. 3: The test conditions used in our study which contain variations of: constant video quality, drops in video quality, and rebuffering.

We divided the participants into three groups: A, B, and C. In Group A, participants were asked to select their preferred layout, after which they were presented with test sequences created based on the video layout they had selected. In Group B, participants were not offered any personalisation, and instead were shown all of the test sequences with the player positioned in the top right – this group acting as our control. In Group C, participants were asked for their preference which was then ignored, and instead a different layout was shown in the test sequences.

Figure 3 shows the test conditions in our study which contain the following quality impairments. First, three test conditions contained a constant video quality of Low, Medium, and High. The next two conditions included a short drop in video quality for 6s in the middle of the clip. The last three conditions contained a single rebuffering event in the middle of the clip, lasting for 0.2, 0.5, and 1s. All clips were 10s long, with the rebuffering test sequences being slightly longer - by the duration of rebuffering.

We used the Absolute Category Rating (ACR) method [22] as the testing procedure. Participants were shown one test sequence at a time and were asked to score each test sequence immediately after its presentation using the 5-point Mean Opinion Score (MOS) scale: Bad, Poor, Fair, Good, and Excellent. The order in which the test sequences were presented was random for each participant. The experimental setup was approved by the Ethics Advisory Board.

B. Test Sequences & Encoding

All of the test sequences were created using ffmpeg[1], with quality impairments introduced at the encoding stage. `H.264` and `AAC` codecs were used to encode all of the test sequences. We defined three video quality levels based on the recommended HLS specifications [23]: low, medium, and high, corresponding to the following resolution and bitrate

[1]https://ffmpeg.org/

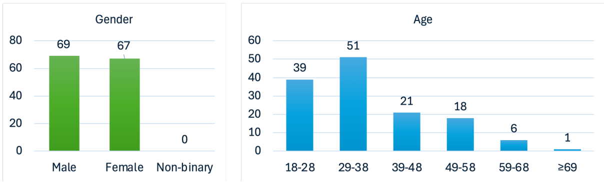

Fig. 4: The gender and age distributions of the participants recruited for our subjective study.

settings: 416x234 at 145kbps, 768x432 at 1100kbps, and 1920x1080 at 4500kbps respectively. All three quality levels were set to 25 frames per second.

To create test sequences, clips were first divided into segments using `trim` and `atrim` filters for video and audio, respectively. Each segment was then adjusted to introduce the quality impairment required for the studies.

First, constant video quality was achieved by encoding all segments at the desired resolution and bitrate, with lower resolution segments additionally upscaled to the highest setting. Second, video quality drops were replicated by encoding the appropriate segments at lower bitrates and resolution settings, and upscaled to the highest resolution tested. Third, rebuffering events were emulated by repeating the last frame of a segment with silent audio for a period lasting the intended duration of rebuffering. No other visual indication of rebuffering was introduced as only short stalling events were tested.

In the final stage, we combined the altered segments using the `concat` filter, and added an 'End of Clip' message at the end. Final clips were compressed further to reduce the file size by setting the `CRF` parameter to 23.

C. Survey Tool

We developed an online survey tool capable of pre-loading full videos before playback to avoid any unintended network-driven quality degradation. Additionally, the survey contained methods of detecting whether a participant had tried to cheat during the study – by not watching clips fully, not with sound, not paying attention to the clips, not watching in full-screen mode, or trying to skip clips all together.

At the beginning of the survey, participants were asked to self-report their age and gender. Additionally, they were asked whether they play video games, or watch videos/streams of others playing video games. Subsequently, each participant was assigned by the survey tool to one of three groups: A, B, or C. Participants in Groups A and C were then shown the personalisation page, when they had to select their preferred video layout to continue. Participants in Group B were not shown this page. Next, the participants would proceed to the clip assessment stage where they would be presented and asked to score one test sequence at a time. The video layout shown in the test sequences was different for each participant, depending on their assigned group and selected preference.

D. Participant Recruitment

We recruited 136 participants through Prolific [24], an ethical crowd sourcing platform. Figure 4 shows the gender and age distributions of the participants. 82% reported playing video games, and 66% stated that they watch videos or streams of people playing video games. The participants were required to complete the survey using a desktop/laptop with a minimum screen resolution of 1920x1080 to match the video resolution of the test sequences being used. 40 participants were assigned to Group A, 42 to Group B, and 54 to Group C.

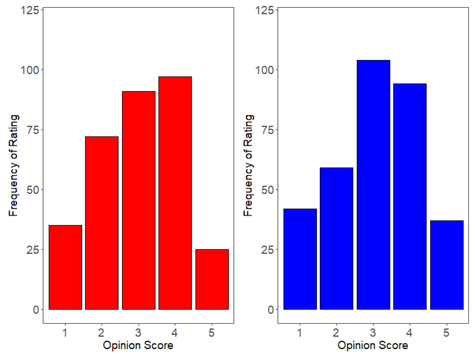

Fig. 5: Comparison of the distribution of subjective test opinion scores for Group A (red) and Group B (blue).

IV. Results

To establish the QoE impact of experience customisation in our subjective tests, this section provides a statistical analysis of the experimental data. We begin by presenting a detailed comparison of the subjective test results for Groups A and B, followed by analyses of the differences between Groups A and C, and B and C, respectively. We conclude with an investigation into how the number of clicks, and the time spent clicking, correlate with QoE.

A. Comparison of Groups A and B

We begin with the overall results from Groups A and B, depicted as bar charts in Figure 5. The MOS for Group B is 0.06 higher, which is a subtle increase; however, a more significant difference can be identified by performing a chi-squared test between the two distributions of opinion scores. The assumptions for this hypothesis test are met as the subjective test participants are a sufficiently large sample (greater than 50) of randomly selected prolific users who rated each video sequence independently. Note that for some videos the counts of some opinion score ratings are less than 10, and so we correct the p-values using a Monte Carlo method [25]. In this paper we use the chi-squared test to determine if there is evidence that two datasets have a different categorical characteristic - in this case, whether the opinion score proportions for Groups A and B differ. The p-value for this test is 0.048; hence, there is evidence to suggest that the two distributions are dissimilar. In other words, the option to choose the player location has some impact on QoE.

Based on the overall results from our subjective tests, the impact of the choice of player location is perhaps surprising. Not only is the MOS slightly higher for Group B, but participants are 41% more likely to score a video as a 5 in Group

Scenario	High Quality	Medium Quality	Low Quality	Quality Drop Medium
p	0.098	0.176	0.017	0.624
Scenario	Quality Drop Low	Rebuffering 0.2s	Rebuffering 0.5s	Rebuffering 1s
p	0.638	0.271	0.530	0.075

TABLE I: Table showing the p-values from the chi-squared tests comparing the difference in the opinion score distributions for the eight permutations of the subjective tests.

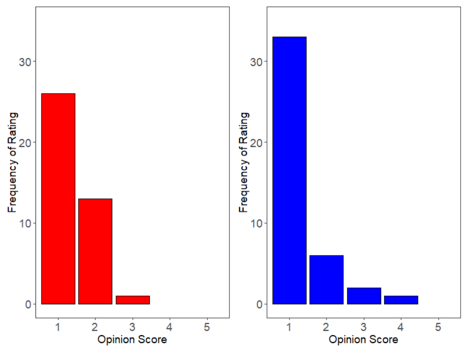

Fig. 6: Barplot comparing the distribution of subjective test opinion scores for Group A (red) and Group B (blue) when the media content quality is low.

B than in Group A, and 10% less likely to score it as a 1 or 2. From this, we infer that the overall effect of being able to select the player location appears to be a negative one. The goal for the remainder of this section is to further interrogate the data to understand why this arguably surprising result has been obtained.

To further explore the initial conclusion that choosing the player location can negatively impact QoE, we investigate the difference between the opinion score distributions for Groups A and B under the different quality permutations tested in the study. The p-values for these chi-squared tests are presented in Table I, and show that there is a significant difference in three scenarios: high content quality, low content quality, and 1.0s rebuffering. We investigate each of these in turn.

With a chi-squared test p-value of 0.017, the strongest evidence of difference comes between the distributions for A and B when the quality is low. A plot comparing opinion scores is presented in Figure 6. The headline here is that a viewer in Group is 185% more likely to rate a video as a 3+, with no ratings of 4 given by any participants in Group A at all. This further reinforces the argument that viewers who have been able to select their player location are scoring content lower than those who have not; however, this seems to be due to the fact that the viewers in Group A have a worse perception of low-quality media. This may be because, for example, their expectations are higher due to being given the option to choose the location of the in-picture video.

Analysis of the rebuffering results, depicted in Figure 7, also supports the conclusion that viewers who are allowed to select their player location are more severely affected by

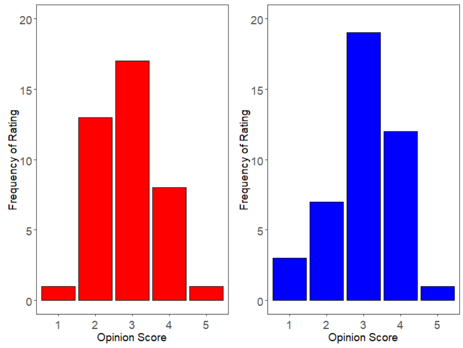

Fig. 7: Barplot comparing the distribution of subjective test opinion scores for Group A (red) and Group B (blue) when there is one second of rebuffering.

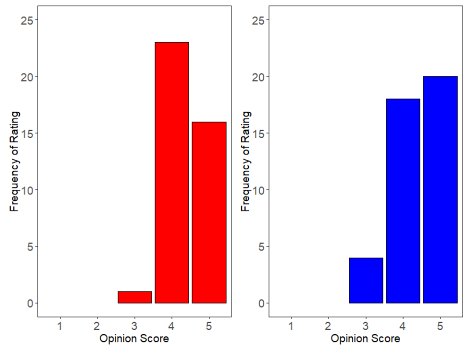

Fig. 8: Barplot comparing the distribution of subjective test opinion scores for Group A (red) and Group B (blue) when the media content quality is high.

quality impairments. Indeed, with a chi-squared p-value of 0.075, there is evidence to suggest that the distribution of opinion scores between A and B differ, with viewers in Group A 47% being more likely to rate a video as a 1 or 2. This illustrates how viewers who have selected their player location are less tolerant of quality impairments, such as a one second rebuffering event, and is indicative of the fact that giving the viewer an option to personalise their media increases their expectations for its overall quality. In turn, this leads to a lower opinion score when the quality does not meet these expectations.

The final test condition where a difference between distributions was identified was when the quality was high. The results are illustrated in Figure 8. In this case, there is limited evidence, with a chi-squared p-value of 0.098, to suggest that the distributions of A and B differ. The plots provide a comparison of these distributions, and we see that a user was 8% more likely to rate a video as a 4+ in Group A, and 10% less likely to rate it as a 3, than in Group B. This result, in contrast to the other results in this section, provides subtle evidence that users who have been able to customise their player location, and have been provided with high quality content, have a better opinion of their experience than those who have not.

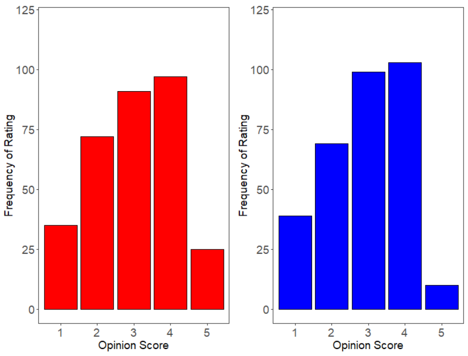

Fig. 9: Barplot comparing the overall distribution of subjective test opinion scores for Group A (red) and Group C (blue).

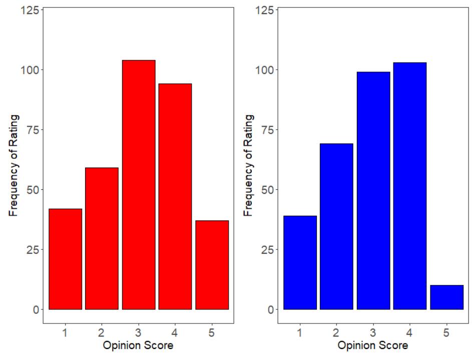

Fig. 10: Barplot comparing the overall distribution of subjective test opinion scores for Group B (red) and Group C (blue).

Combining these results from the different test conditions, we have acquired far greater insight than by solely comparing the difference between the overall distributions. From the results, we observed that when the media quality is high and there are no events affecting this, giving a viewer the option to choose their player location increases the probability that they will have a positive view of the content on the MOS scale. On the other hand, asking the viewer to customise their experience has increased their expectations for the media, and so these viewers perceive low quality footage or rebuffering events more harshly than if they had not been offered the customisation opportunity.

As a final comment, given the discussion on how negative QoE events affect viewers in Group A more than those in Group B, the reader may wonder why there has not been a difference identified between the two groups when considering quality drops. With chi-squared p-values of 0.624 and 0.638 for the medium and low quality drops, respectively, there is no evidence to suggest that the two distributions are different. Looking at the drop in MOS for the two groups clarifies why, however with, both groups reporting a drop of over 2 in the MOS. This shows that both groups are more concerned with the negative impact of the low quality than they are with any positive opinions arising from personalisation.

B. Comparison of Groups A and C

The opinion score distributions for Groups A and C can also be compared using a chi-squared test, and with a p-value of 0.001 we find that there is significant evidence that these two distributions are different. A plot comparing the distributions is shown in Figure 9. We observe that Group A are 151% more likely to rate a video as a 5, 10% less likely to rate it as a 1, and have a slightly higher MOS value. From this we conclude that users who are able to personalise the player location have a higher QoE than those who are offered the chance to personalise their selection but then given something else. Whilst this may seem unsurprising, it is important to have demonstrated that it is not solely the option to personalise the media that leads to an increase in the QoE, but also the viewer seeing their choices reflected in the video.

C. Comparison of Groups B and C

Our final comparative analysis assesses the difference between Groups B and C, with Figure 10 showing the two distributions of opinion scores. Performing a chi-Squared test, we obtain a p-value of 0.001, and so there is strong evidence that the two groups have different opinions. We find that the MOS for Group B is 0.15 higher, and a two-sample t-test returns a p-value of 0.089, indicating that there is some evidence that the mean for Group B is higher than that for Group C. Furthermore, users in Group B were 12% more likely to rate a video as a 4 or higher, and 10% less likely to rate it as a 2 or lower, than those in Group C.

From this statistical analysis, we conclude that if a user is offered the opportunity to personalise their viewing experience, but the content does not reflect their choices, then this results in a lower QoE. This therefore illustrates a potential challenge when delivering personalised content: personalisation has increased expectations for the media, and getting the user's customised options wrong may negatively impact their QoE to the point that it would have been better not to offer it.

D. Click Data Analysis

Our final analysis concerns the relationship between click times, number of clicks, and MOS. To assess this, we measure the association between the former two quantities and the MOS for the high quality video. We cannot measure the correlation between these variables using ordinary correlation coefficients, as they are discrete observations; therefore, we measure the association using Cramer's V statistic, which is designed for arbitrary quantities. We find that $V = 0.529$ for the number of clicks, and $V = 0.383$ for the time spent clicking.

As per [26], a V statistic above 0.4 can be considered evidence of a relatively strong association. Thus, our findings indicate that there is an association between the number of clicks and the MOS. Alone Cramer's V Statistic does not identify whether the association is positive or negative, and so to determine this we fit a logistic regression model that predicts the probability a user gives a rating of 5 based on their number of clicks. We see that the coefficient of this model is positive, which shows that as the number of clicks increases, so does the MOS.

Click data analysis provides some evidence that a user who makes more actions when personalising their media has a better QoE. Interestingly, the stronger association between the number of clicks than the time spent suggests that it is not necessarily the length of time spent that enhances the viewer experience, but the ability to make as many clicks as they need to get it “just right”.

V. Discussion

Personalised media presents a significant opportunity for improvements to the Quality of Experience in multimedia, by allowing the user to tailor their viewing experience according to their own preferences. In our subjective study, we have tested a Picture-In-Picture gaming use case, where the footage of the player is overlaid on top of the gameplay video. This allows for considerable customisation of video layout, as the player element can be positioned anywhere within the bounds of the gameplay element, or hidden altogether, ensuring that the gameplay is never obscured. Figure 11 shows the video layout preference expressed by the participants in Groups A and C. While the most common option was ‘Top Right’, the layout shown to all participants in Group B, there were plenty of other preferences chosen by our participants.

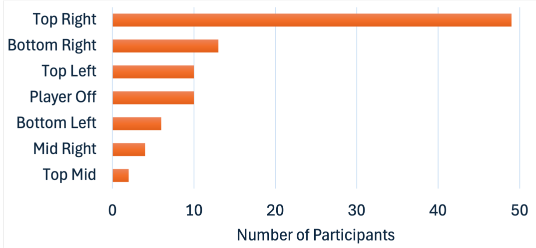

Fig. 11: The frequency of preferred video layouts by participants in Groups A and C.

In the results of our study we observe the double-edged QoE impact of experience customisation. Participants in Group A, who selected and viewed all clips with their preferred video layout, have scored the test condition showing constant high quality better than the participants in Group B, our control. Additionally, participants who performed more clicks when selecting their preference were more likely give higher scores. These two observations answer our first research question.

However, surprisingly, the test conditions containing the most severe quality impairments were rated in the opposite manner, with participants in the control group finding them less annoying. This unexpected finding answers our second research question and indicates the double edged effect of personalisation in multimedia. This contradicts conventional assumptions and highlights the depth of the QoE challenge in the delivery of personalised media. These results have significant impact on the potential delivery of personalised media experiences. For example, if such content were delivered using HTTP adaptive video streaming, such as DASH [1] or HLS [2], it would require more sophisticated ABR logic that takes into account the increased expectations of users, to minimise the occurrence of the most annoying quality impairments.

For content producers, this insight can encourage the development of new interfaces that provide consumers with customisation options in order to improve the QoE of multimedia experiences. Our findings suggest that the priority dictated by consumers might be different from that of producers, and having such interfaces would allow content service providers to find a sweet spot between consumer preference and production vision. This creates another challenge, as there needs to be a reasonable balance between producers’ intent and the impact of offered customisation options on the narrative of experiences. Future work should explore this tension between customisation and narrative integrity, particularly in media contexts where user agency is an established element of the experience.

As a final point, we consider the limitations of our subjective study. It was constrained to a single experience with a limited amount of customisation, as well as a small sample size per group. Therefore, we note that the presented findings are preliminary, offering initial insights into the QoE impact of user-driven customisation in personalised media experiences. Future work should explore a greater variety of experiences, along with different degrees of customisation.

VI. Conclusion

In this study, we investigated the QoE impact of user-driven customisation in personalised media experiences. We conducted an extensive subjective study, which recruited 136 participants divided into three experimental groups. We designed 10 personalisation options for a Picture-In-Picture gaming video, allowing the user to select their preferred video layout by hiding or changing the position of the player element on the screen. We found subtle evidence for the double-edged QoE impact of personalisation in the results of our studies, with the highest quality of video being perceived better by users who personalised their experience. On the other hand, the most severe quality impairments were perceived as better by the control group. We conclude that personalisation can improve QoE; however, QoE modelling needs to carefully consider the paradoxical impact on user expectations, where personalisation simultaneously enhances satisfaction with high-quality content while amplifying disappointment with quality degradations. This not only has implications for content producers, but also for providers as this will affect ABR algorithms and content delivery mechanisms.

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] ISO/IEC, "ISO/IEC 23009-1: Information technology-dynamic adaptive streaming over HTTP (DASH)-part 1: Media presentation description and segment formats," 2019.

[2] IETF, "HTTP live streaming (RFC 8216)," 2017. [Online]. Available: https://tools.ietf.org/html/rfc8216

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

[4] J. A. Colley and A. Crabtree, "Object based media, the IoT and Databox," in *Living in the Internet of Things: Cybersecurity of the IoT*. IET, 2018.

[5] M. Ursu, D. Smith, J. Hook, S. Concannon, and J. Gray, "Authoring interactive fictional stories in object-based media (OBM)," in *ACM International Conference on Interactive Media Experiences (IMX)*, 2020, pp. 127–137.

[6] J. Cox, R. Jones, C. Northwood, J. Tutcher, and B. Robinson, "Object-based production: a personalised interactive cooking application," in *ACM International Conference on Interactive Experiences for TV and Online Video*, 2017, pp. 79–80.

[7] M. Armstrong, M. Brooks, A. Churnside, M. Evans, F. Melchior, and M. Shotton, "Object-based broadcasting-curation, responsiveness and user experience," BBC R&D, Tech. Rep. WHP 285, Sep. 2014.

[8] M. Armstrong, S. Bowman, M. Brooks, A. Brown, J. Carter, A. Jones, M. Leonard, and T. Preece, "Taking object-based media from the research environment into mainstream production," *SMPTE Motion Imaging Journal*, vol. 129, no. 5, pp. 30–38, 2020.

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

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

[11] N. Barman and M. G. Martini, "QoE modeling for HTTP adaptive video streaming – a survey and open challenges," *IEEE Access*, vol. 7, pp. 30 831–30 859, 2019.

[12] H. T. T. Tran, T. Vu, N. P. Ngoc, and T. C. Thang, "A novel quality model for HTTP adaptive streaming," in *International Conference on Communications and Electronics (ICCE)*, 2016, pp. 423–428.

[13] M. N. Garcia, W. Robitza, and A. Raake, "On the accuracy of short-term quality models for long-term quality prediction," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015.

[14] N. Staelens, J. De Meulenaere, M. Claeys, G. Van Wallendael, W. Van den Broeck, J. De Cock, R. Van de Walle, P. Demeester, and F. De Turck, "Subjective quality assessment of longer duration video sequences delivered over http adaptive streaming to tablet devices," *IEEE Transactions on Broadcasting*, vol. 60, no. 4, pp. 707–714, 2014.

[15] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual quality of HTTP adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2141–2153, 2016.

[16] T. Lyko, M. Broadbent, N. Race, M. Nilsson, P. Farrow, and S. Appleby, "Improving quality of experience in adaptive low latency live streaming," *Multimedia Tools and Applications*, pp. 1–27, 2023.

[17] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of HTTP video streaming," in *IFIP/IEEE International Symposium on Integrated Network Management (IM) and Workshops*, 2011, pp. 485–492.

[18] K. Yamagishi and T. Hayashi, "Parametric quality-estimation model for adaptive-bitrate-streaming services," *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1545–1557, 2017.

[19] D. Zegarra Rodríguez, R. Lopes Rosa, E. Costa Alfaia, J. Issy Abrahão, and G. Bressan, "Video quality metric for streaming service using DASH standard," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 628–639, 2016.

[20] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang, "A quality-of-experience index for streaming video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 154–166, 2017.

[21] M. I. Norton, D. Mochon, and D. Ariely, "The IKEA effect: When labor leads to love," *Journal of Consumer Psychology*, vol. 22, no. 3, pp. 453–460, 2012.

[22] I.-T. S. Group *et al.*, "P. 910: Subjective video quality assessment methods for multimedia applications," *International Telecommunications Union: Geneva, Switzerland*, 2008.

[23] Apple. (2023) HTTP live streaming (HLS) authoring specification for Apple devices. [Online]. Available: https://developer.apple.com/documentation/http-live-streaming/hls-authoring-specification-for-apple-devices

[24] Prolific. [Online]. Available: https://www.prolific.co/

[25] A. Agresti, *Categorical Data Analysis*, ser. Wiley Series in Probability and Statistics. Wiley, 2012. [Online]. Available: https://books.google.co.uk/books?id=UOrr47-2oisC

[26] D. K. Lee, "Alternatives to P-value: confidence interval and effect size," *Korean Journal of Anesthesiology*, vol. 69, no. 6, p. 555, 2016.