Drop or Stop: Investigating the Impact of Playback Rate on QoE in Adaptive Video Streaming

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*—Quality of Experience (QoE) is a crucial component of adaptive bitrate (ABR) streaming, with the effects of abrupt changes in playback quality or rebuffering, caused by delivery disruptions, being widely studied. However, the collective ABR community has a limited understanding of the effects of changes in playback *rate* on QoE. In this pioneering work, we investigate two aspects of playback rate fluctuations. In particular, we carry out two subjective studies to assess if a change in playback rate is more or less acceptable than a drop in video quality or a rebuffering event. Furthermore, we examine the effect of the transition in playback rate on QoE, comparing gradual and instant variations. Our subjective studies recruited 120 participants who evaluated 102 test sequences. In summary, we find that playback rate drops of 0.8-0.9 are imperceptible for most content, and rated similarly to a video quality drop to medium level. In contrast, lower playback rates of 0.6-0.7 were perceived as poorly as rebuffering events. Gradual changes in playback rate can offer better QoE, but only in limited cases depending on the content, target playback rate, as well as magnitude of change.**

***Index Terms*—playback rate, adaptive streaming, QoE, assessment, subjective study**

I. Introduction

Video streaming over HTTP (such as DASH [1] and HLS [2]) typically makes use of adaptive bitrate (ABR) algorithms in video players that continually learn which video quality (bitrate representation) optimally fits the bandwidth of the connection and react to network events such as congestion and drop in signal-strength. For maximum Quality of Experience (QoE), these algorithms must maintain high video streaming bitrate, and achieve low rebuffering and bitrate oscillations. To do so, client-side ABR algorithms employ strategies like estimating available bandwidth and measuring metrics such as playback buffer occupancy, number of previous buffer stalls, and frequency of quality oscillations, to select the appropriate bitrate level for the next segment to be downloaded.

As such, the bulk of research efforts in client-side bitrate adaptation tends to focus on either *bandwidth-estimation-based adaptation* schemes, where bandwidth fluctuations and congestion are detected based on video segment download times, or *playback-buffer-based adaptation* schemes, where playback buffer occupancy is used as a criterion to select the next segment's bitrate [3]. While moving down the video encoding ladder to select a lower bitrate is the primary strategy to respond to network congestion and avoid rebuffering, there has been scant attention paid to the alternative of adjusting video playback rate to avoid buffer underruns, and the impact this strategy may have on perceived quality.

This work is motivated by the potential of *client-side playback-rate adaptation* as a complementary approach to bitrate switching for rebuffering avoidance in HTTP-delivered video playback. The viability of this approach is explored via answers to the following research questions:

- How are changes in playback rate perceived by users across different content genres compared to other QoE-affecting factors?
- Are gradual changes in playback rate perceived better than more drastic adjustments? Can such a strategy enable greater variations in altered playback rate?

The former question is motivated by the exploration of alternative means of dealing with delivery issues whilst maintaining high QoE. The latter question is exemplified by recent studies on the impact on QoE by the magnitude of bitrate oscillations in adaptive bitrate streaming.

Our contributions in this paper are as follows.

1) Evaluation of playback rate drops when compared to two main QoE factors in adaptive video streaming: video quality and rebuffering, using a subjective study consisting of 54 test sequences, 6 pieces of content and 40 participants.
2) Investigation of gradual changes in playback rate, across two magnitudes of change, using a subjective study compromising of 48 test sequences and 80 participants.

Both studies presented in this paper expand on the related work by evaluating two aspects of playback rate that have not been considered in the existing literature.

II. Background and Related Work

Quality of Experience (QoE) is an important aspect of ABR streaming. It can be divided into the following main factors as outlined in the survey by Barman et al. [4]. Rebuffering events which have strong negative impact on QoE [5]–[8]. Video quality is another important QoE factor [9]–[11], with multi-level video quality switches, where the quality changes across more than one level between two consecutive segments, capable of having negative impact on QoE [11]–[13].

In the recent years, new ABR algorithms for low latency live adaptive streaming have been proposed [14]–[19]. These ABRs have been designed to operate in scenarios where client buffer is severely restricted by the need to maintain low latency. In such conditions, playback rate might be adjusted by the ABR algorithm to aid rebuffering avoidance, or by the video player to maintain a specific target latency.

There is very limited literature on the relationship between playback rate and QoE. Two studies, both published at QoMEX, investigated this to different extents. The first, by Rainer and Timmerer [20], they studied the QoE impact of decreasing and increasing playback rate, using a single piece of content. They found that audio plays an important role in the perceptibility of playback rate changes, and that decreases were more perceptible than increases. The other study was carried out by Pérez et al [21], in which they investigated increases and decreases in playback rate across a variety of content, finding that its perceptibility is content-dependent and that a 10% change is safe. Both studies have investigated playback rate in isolation and did not consider gradual changes. In this paper, we aim to expand on the related work by evaluating playback rate when compared to the impact of video quality drops and rebuffering events in the same study, as well as, by investigating the magnitude of changes in playback rate.

III. Methodology

To investigate the impact of playback rate drops in relation to other QoE factors, as well as the impact of employing changes in playback rate gradually instead of instantly, we designed two subjective studies, one for each research question outlined in Section I. Both studies share some common methodologies, which we detail in this section.

Our goal was to perform both studies using an online crowdsourcing platform to gather responses at reasonable time and cost overheads. In the following subsections, we first outline the testing procedure selected for our studies, then we describe an online survey tool we developed to gather reliable results. Thereafter, we describe the selected clips of various content genres for stimuli used in the studies, and finally we detail the encoding process behind the test sequence creation for our studies to produce multiple types of quality impairments.

A. Testing Procedure

We selected the Absolute Category Rating (ACR), outlined in ITU Rec. P.910 [22], as our video quality assessment method. In ACR, test sequences are presented one at a time, with participants asked to score each sequence immediately after presentation. The standard 5-point Mean Opinion Score (MOS) scale was used: *Bad*, *Poor*, *Fair*, *Good*, and *Excellent*. For the first study, we used the Absolute Category Rating with Hidden Reference (ACR-HR) method, which adds an additional test sequence that is presented without any quality impairments.

B. Survey Tool

In order to perform the studies remotely, we created an online survey capable of clip playback without unintended quality impairments, such as rebuffering. In the survey, the test sequences were always fully fetched before their playback could begin. Additionally, clips are always played in full-screen mode with the sound switched on.

The survey begins with a consent form as well as questions about gender, age, and past experience with video quality assessment studies. It then proceeds to the instructions page, demonstrating the video playback mechanism along with a single example clip for each tested video quality impairment; different content is used for the rest of the survey.

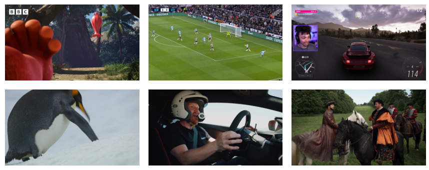

Fig. 1. Thumbnails of the clips used in the studies (from top left): Vegesaurs, Football, Forza, Frozen Planet, Top Gear, and Wolf Hall.

TABLE I
A summary of the properties of the content used in our studies.

Clip	Genre	SI	TI
Vegesaurs	Cartoon	33	35
Football	Sports	49	19
Forza	Gaming	39	33
Frozen Planet	Documentary	21	44
Top Gear	Entertainment	24	37
Wolf Hall	Drama	19	33

Next, the survey proceeds to the video quality assessment stage, where the participant is shown and asked to score test sequences, one at a time. The first few test sequences were a training set, presenting a subset of the tested conditions using different content to allow the participant to become accustomed to the scoring scale. The remaining test sequences were arranged at random, with a single constraint to ensure that the same content was never presented twice in a row.

The survey ends with an attention-check question, querying the participant about one of the scenes shown in one of the clips. Additionally, the participant is asked to confirm whether they had watched the clips in full-screen mode and with the sound on as instructed.

Both studies were performed using an ethical crowdsourcing platform Prolific[1]. To make sure that the responses given by participants were reliable, we incorporated a mechanism to detect whether the participant had watched all of the test sequences completely before rating them. We also tracked whether the participant tried to adjust the playback rate or seek when watching clips. Responses in which participants did not watch clips in full or tried to seek or adjust playback rate were discarded. This study was approved by our ethics committee.

C. Content

As seen in a previous study on this topic [21], the impact of changes in playback rate can vary across different content genres. Given this observation, we sourced six clips of different genres. Figure 1 shows the thumbnails of the selected clips and Table I describes the clips in terms of calculated Spatial Information (SI) and Temporal Information (TI) according to ITU P.910 [22], as well as content genre. Each clip was 11s long, 25FPS, and had resolution and bitrate greater or equal to 1920x1080 at 6000kbps.

D. Encoding

In order to prepare the test sequences, we used ffmpeg[2] to introduce the desired video and audio quality impairments.

[1]https://www.prolific.com/
[2]https://ffmpeg.org/

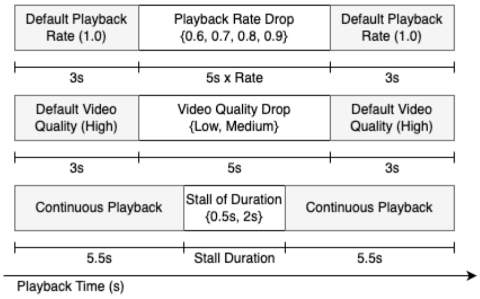

Fig. 2. Test conditions used in the first study, describing playback rate drops, video quality drops, and rebuffering events.

Clips were encoded using the `H.264` and `AAC` codecs.

To create test sequences, clips were first divided into segments using `trim` (for video) and `atrim` (for audio) filters. Each segment was then adjusted to introduce the quality impairment required for the studies. First, changes in playback rate were achieved using `setpts` (for video) and `atempo` (for audio) filters. Second, video quality changes were executed by encoding the segments at different bitrate and resolution settings and upscaled to the highest resolution tested. Third, rebuffering events were introduced by repeating the last frame of a segment with silent audio for a period of time matching the desired rebuffering duration.

Finally, the adjusted segments were stitched together (using the `concat` filter) along with an 'End of Clip' message at the end. Further compression was applied to reduce file size by setting the `CRF` parameter to 23. For video quality impairment sequences, three video quality levels were defined, based on the recommended HLS specifications for Apple devices [23]: low, medium, and high, corresponding to the following resolution and bitrate pairings: 416x234 at 145kbps, 960x540 at 2000kbps, and 1920x1080 at 6000kbps respectively. All test sequences that did not contain any video quality impairments were encoded at the highest quality level.

IV. Playback Rate vs. Other QoE Factors

In this section, we present and describe the results of the first subjective study, designed to investigate the QoE impact of changes in playback rate when compared to drops in video quality and rebuffering events.

A. Study Set-up

We created eight test conditions, as illustrated in Figure 2, which were shown to each participant along with a hidden reference test sequence that did not contain any quality impairments. Four test conditions focused on changes in playback rate and consisted of slowed down playback for the middle section of the clip. The next two test conditions contained a single rebuffering event in the middle of the clip, of 0.5s and 2s. In the last two conditions, the video quality of the middle section of the clip was reduced to a Low or Medium quality level.

All six sourced clips were used for this study, resulting in 54 test sequences, which were divided into two sets with each containing all test sequences belonging to 3 out of 6 clips. Set A contained all test sequences based on clips: Vegesaurs, Football, and Forza, while set B contained all test sequences based on clips: Frozen Planet, Top Gear, and Wolf Hall.

We recruited 40 participants in total on the Prolific platform. Each set of test sequences was watched by 20 participants. 50% of participants were male, and 50% female. The age groups can be described as 18-28 for 5%, 29-38 for 40%, 39-48 for 35%, 49-58 for 15%, and 59-68 for 5% of participants. 90% of the participants reported having no previous experience in video quality assessment studies. All participants were required to complete the survey on a desktop/laptop device with a screen resolution of 1920x1080 – the video resolution of our test sequences.

B. Results

Figure 3 presents the comprehensive outcomes of the first study, organised by content type. Differential Mean Opinion Scores (DMOS), calculated using the hidden reference test sequence without quality impairments, were plotted against the test conditions, with confidence intervals of 95% plotted in the error bars. Two-point crushing function was applied as outlined in ITU Rec. P.910 [22]. Additionally, we performed a paired T-test ($\alpha = 0.05$) to determine whether differences in MOS between test conditions were statistically significant.

Impact of Individual Quality Impairments. In the results presented in Figure 3, we observe the impact of changes in playback rate alone. For all of the playback rate test sequences, the first and the last 3s of the clip would play at normal rate (1.0x) with the middle of the clip playing at the target playback rate. These test sequences were encoded at the highest level and did not contain any rebuffering.

For test condition PR-0.9, where the playback rate dropped to 0.9, the average DMOS across all content equalled 4.7; differences between these test sequences and the hidden reference are statistically insignificant - except for the Vegesaurs clip. At a lower playback rate of 0.8, presented in test condition PR-0.8, the average DMOS decreased to 4.5, with differences still statistically insignificant for half of the clips: Vegesaurs, Frozen Planet and Top Gear. This rate was the most noticeable for the Wolf Hall clip with a DMOS of 3.9. Compared to the playback rate of 0.9, this resulted in reduced DMOS by only 0.2 on average, with all differences between the two sets of test sequences being statistically insignificant - except for the Wolf Hall content, where the DMOS dropped by 0.9. For playback rates of 0.7 and 0.6, presented in PR-0.7 and PR-0.6, the average DMOS further decreased to 3.4 and 2.9 respectively, with all differences being statistically significant. The differences between the scores of both these rates were statistically significant for all content.

When comparing the scores of test sequences featuring quality drops, we can observe the impact of this quality impairment alone. For test condition QD-M, where the middle of the clip was presented at medium quality, the average DMOS across all content was equal to 4.8, meaning it was only 0.2 lower than the score of the hidden reference clip which was encoded at high quality in its entirety. All of the differences between these test sequences and the hidden reference were statistically insignificant. This suggests that this quality impairment was

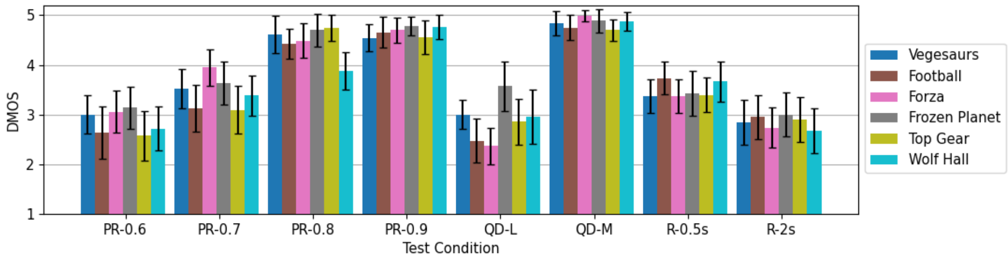

Fig. 3. Results of the first subjective study, with Differential Mean Opinion Scores (DMOS) plotted for each test condtion and content.

imperceptible to participants and therefore did not have an impact on QoE, despite the birate decreasing significantly, from 6000 to 2000kbps. For the test sequences featuring a quality drop to the lowest level, QD-L, the average DMOS dropped significantly to 2.9 on average across all content with all differences being statistically significant. However, the range of scores between pieces of content was at 1.2, with Frozen Planet scoring the highest at 3.6 and Forza scoring the lowest at 2.4.

We can observe the impact of rebuffering alone, of various durations, by comparing the scores of test sequences for the two test conditions featuring rebuffering events, R-0.5s and R-2s. In these clips, there was a single rebuffering event in the middle of the video, of duration of 0.5s or 2s. These test sequences were encoded at the highest quality level, with a regular playback rate of 1.0.

For the test sequences with a rebuffering event of 0.5s, the average DMOS across all content equalled to 3.5, with all differences between the test sequences and the hidden reference being statistically significant. This test condition was perceptible and decreased QoE across all content. When the rebuffering increased to 2s, the average DMOS further decreased to 2.9 on average across all content, with all differences being statistically significant again. As expected, the negative impact on QoE grew as the rebuffering duration increased, with differences between 0.5 and 2s test sequences being statistically significant for all content except the Frozen Planet clip.

Finding 1. *A medium quality drop as well as playback rate drops of 0.8-0.9 had little to no impact on the perceived quality.*

Playback Rate vs. Quality Drops. When comparing the test sequences containing variations in playback rate and quality drops, the following observations can be made. The test condition QD-M, where the video quality dropped to medium level in the middle of the clip, was imperceptible with an average DMOS of 4.8 across all content. Similarly, PR-0.9, where the playback rate of the middle section of the clip was reduced to 0.9, was also imperceptible but with a slightly lower average DMOS of 4.7. Differences in DMOS between these two test conditions were statistically insignificant for all content except for Vegesaurs and Forza.

PR-0.8 was scored only 0.3 lower than QD-M on average across all content, with only Forza and Wolf Hall clips showing statistically significant differences. Wolf Hall had notably lower DMOS for PR-0.8 when compared to other content, suggesting that dropping the quality level to medium would be much better in this case. The next lowest-scored test condition, PR-0.7 still achieved 0.5 higher DMOS on average than QD-L, where the video quality of the middle segment dropped to the lowest level. The differences were statistically significant for half of the content: Vegesaurs, Football, and Forza.

The two lowest-scored test conditions, PR-0.6 and QD-L, were rated similarly with an average DMOS of 2.9 across content and differences statistically insignificant for all clips except Forza. This suggests that a playback rate of 0.6 is as perceptible and detrimental to QoE as dropping the video quality to the lowest level.

Finding 2. *Playback rate drop of 0.6 was perceived as poorly as a low quality drop.*

Playback Rate vs. Rebuffering. We now compare the test conditions containing playback rate changes and rebuffering events.

Test conditions PR-0.9 and PR-0.8, containing playback rates drops of 0.9 and 0.8 respectively, were rated better than both test conditions containing rebuffering events R-0.5s and R-2s – where the rebuffering duration was 0.5s and 2s respectively. PR-0.9 had a 1.2 and 1.8 higher DMOS on average across all content than R-0.5s and R-2s respectively, while the average DMOS for PR-0.8 was only 0.2 lower than for PR-0.9. All differences between PR-0.9 and both R test conditions were statistically significant.

PR-0.7 and R-0.5s were rated similarly with average DMOS of 3.4 and 3.5 respectively across all content. Most of the differences were statistically insignificant, except for clips of Football and Forza, suggesting that a playback rate of 0.7 has a similar detrimental impact on QoE as a single rebuffering event of 0.5s.

Similar observations are made about test conditions PR-0.6 and R-2s, which had an average DMOS of 2.9. with the differences being statistically insignificant for all clips. This suggests that a single rebuffering event of 2s is perceived similarly to a playback rate of 0.6, with both having a notable negative impact on QoE.

Finding 3. *Playback rate drops of 0.6 and 0.7 were rated as poorly as rebuffering events of 2s and 0.5s respectively.*

V. Gradual vs. Instant Changes in Playback Rate

We turn our attention to the second subjective study, investigating the impact of gradual changes in playback rate. Inspired by the existing literature on the impact of changes in video quality in adaptive streaming, where in some cases gradual changes, one level at a time, offer better QoE than instant changes, of more than one quality level between two consecutive segments [11], [12].

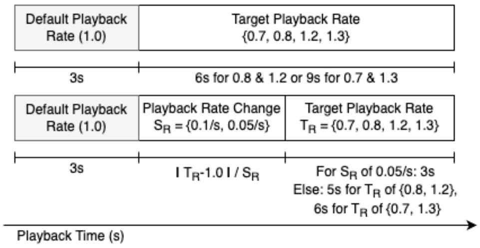

Fig. 4. Test conditions used in the second study, describing instant and gradual changes in playback rate.

A. Study Set-up

For this study, we created 4 sets of test conditions, grouped by target playback rate: 0.7, 0.8, 1.2, and 1.3. In each group there are 3 test conditions: I, where the playback rate change is instant, G-0.05 and G-0.1, where the playback rate is gradual, changing by 0.05/s and 0.1/s, respectively, towards the target playback rate. Figure 4 illustrates the test conditions. For both instant and gradual variations, the test sequence always begins with playback at the default rate (1.0) and finishes at the target playback rate. The duration of test sequences varies, however, the extracted amount of content is the same between instant and gradual variations of each playback rate: 9s for 0.8, 1.2, and 11s for 0.7, 1.3. In total 12 test conditions were created, assessing not only a drop in playback rate (as in our first study), but also an increase in playback rate. This study does not include a test condition without quality impairments, as it aims to compare the differences between instant and gradual changes only. Additionally, it differs from the the first study, where only temporary drops in playback rate were tested.

Four out of six pieces of content were used for this study. We selected the clips for which changes in playback rate were the most perceptible, based on the results of our first study: Football, Frozen Planet, Top Gear, and Wolf Hall. For each clip, test sequences covering all 12 test conditions were created, resulting in 48 test sequences in total. Test sequences were divided into 4 sets, each set containing all pieces of content, with each content presented using a different group of test conditions, meaning, each content was seen at a different playback rate along with its one instant and two gradual variations.

We recruited 80 participants using the same crowdsourcing platform as in our first study. 20 participants watched each set of test sequences. The gender of participants was reported as 50% male and 50% female. The age group representation was 16.25% for 18-28, 26.25% for 29-38, 26.25% for 39-48, 16.25% for 49-58, and 15% for 59-68. 92.5% of participants reported having no previous experience in video quality assessment. Just as with the first study, all participants were required to complete the survey on a desktop/laptop device with a screen resolution of 1920x1080.

B. Results

Figure 5 shows the full results of the second study, broken down by content. Mean Opinion Scores were plotted against the test conditions, with confidence intervals of 95% plotted in the error bars. Additionally, we performed a paired T-Test ($\alpha = 0.05$) to determine whether differences in MOS between test sequences were statistically significant. We can analyse the impact of gradual changes in playback rate across the different variations tested: two rate decreases of 0.7 and 0.8, as well as two rate increases of 1.2 and 1.3.

At a playback rate of 0.7, the instant change variation presented in 0.7-I had an average MOS of 2.7 across all content. A gradual change of 0.1/s, present in 0.7-G-0.1, improved MOS by 0.3 on average, however, the differences were statistically significant for only two clips: Top Gear and Wolf Hall which improved by 0.4 and 0.8 respectively. A gradual change of 0.05/s, present in 0.7-G-0.05 and offering even more subtle change, further improved the MOS by 0.7 on average with all differences being statistically significant. The improvement varied across content, with increased MOS by 0.8, 0.4, 1, and 0.9 for Football, Frozen Planet, Top Gear, and Wolf Hall respectively.

For a playback rate of 0.8, when comparing the test sequences containing gradual changes and instant changes, most differences were statistically insignificant with an average MOS only 0.2-0.3 higher. The only exception was 0.8-G-0.05 for the Top Gear clip, which scored 0.7 higher than the corresponding 0.8-I test sequence.

When looking at increases in playback rate, starting with 1.2, gradual changes offered little to no improvement, with average MOS at 3.9, 3.9, and 4 for test conditions 1.2-I, 1.2-G-0.1, and 1.2-G-0.05 respectively. In the case of the Wolf Hall clip, MOS stayed the same for a gradual change of 0.1/s and decreased by 0.2 for 0.05/s - however, both of these differences were statistically insignificant. For the Frozen Planet clip, MOS decreased by 0.3 and 0.5 for gradual changes of 0.1/s and 0.05/s respectively, however, only the latter was statistically significant. For the Football clip, there was only one decrease in MOS, of 0.2, for a gradual change of 0.1/s, however, it was not statistically significant. For a gradual change of 0.05/s, the MOS improved by 0.3, with the difference statistically significant. In the case of Top Gear content, the MOS improved by 0.4 and 0.5 for gradual changes of 0.1/s and 0.05/s respectively.

For a playback rate of 1.3, we can observe some improvement as the average MOS increased by 0.2 and 0.4 for gradual changes of 0.1/s and 0.05/s respectively. However, not all of the differences were statistically significant. For the Football clip, only a gradual change of 0.05/s achieved a statistically significant MOS improvement of 0.4. In the case of Frozen Planet content, both improvements were statistically significant, resulting in 0.5 and 0.7 better MOS for gradual changes of 0.1/s and 0.05/s respectively. For the two remaining clips, Top Gear and Wolf Hall, the differences were statistically insignificant.

Overall, the more gentle gradual changes of 0.05/s were perceived better than the gradual changes of 0.1/s. However, both variations of gradual changes offered improvement in only some cases, suggesting that it is dependent on the playback rate and content. Improvement was the most notable for the lowest and highest playback rates tested, 0.7 and 1.3. In terms of content, Wolf Hall benefited the most from gradual changes, however, the improvement in MOS was not universal across all playback rates with no difference at the rate of 1.3.

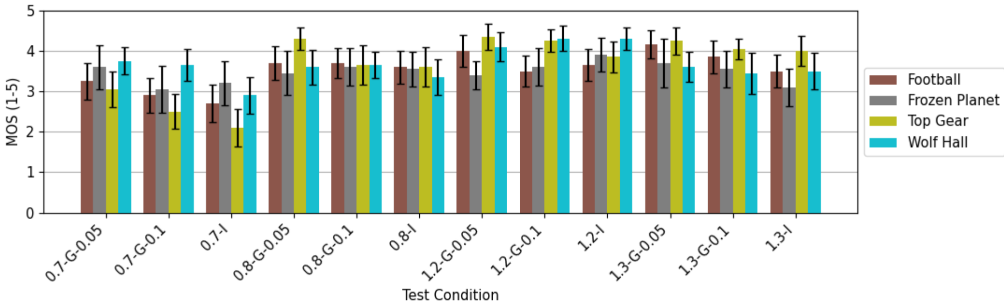

Fig. 5. Results of the second subjective study, with Mean Opinion Scores (MOS) plotted for each test condition and clip.

VI. Discussion

In our first study, we investigated the QoE impact of playback rate in comparison to two other important QoE factors in adaptive video streaming: video quality and rebuffering. A quality drop to medium level (where bitrate reduced from 6000 to 2000kbps) and a playback rate drop of 0.9 were both imperceptible to participants. A playback rate drop of 0.8 was still imperceptible for half of the content. Overall, these three quality impairments were rated the highest in our test. This suggests that temporarily reducing the playback rate to 0.8 or higher could be a viable option in ABR algorithm design as it has little to no negative impact on overall QoE while allowing the client to reduce its buffer depletion rate. For example, this might be preferred in a scenario where the video quality is already at a medium level and dropping it further would have a negative impact on QoE.

For half of the content, a playback rate drop of 0.7 offered some improvement over the quality drop to the lowest level - which was perceived as the worst, along with a playback rate of 0.6 and a rebuffering event of 2s. A smaller rebuffer duration of 0.5s was perceived similarly to a playback rate drop of 0.7, however, both of these conditions still had a significant negative impact on overall QoE. These results indicate that a playback rate of 0.7 and lower should be avoided in ABR design as a method for rebuffering avoidance, since its negative impact is comparable to rebuffering itself.

The second study focused on investigating gradual changes in playback rate and whether they can absorb some of the negative QoE impact caused by the target playback rate. This varied greatly across content, target playback rates, and magnitude of gradual change. In the case of most content, for target playback rates of 0.8, 1.2, 1.3 gradual changes offered no significant improvement over the instant variations. However, there were exceptions, with Wolf Hall improving at 0.8, and Top Gear clips being perceived better in gradual variations for target playback rates of 1.2 and 1.3. At the lowest target playback rate tested, 0.7, there was a significant improvement in perceived quality for both variations of gradual change, but only for half of the content in the case of the less subtle variant.

These results suggest that gradual changes in playback rate can be less perceptible but only in some cases as this improvement is not universal across all content and target playback rates. Additionally, the magnitude of gradual playback rate change is an important factor.

VII. Conclusion

In this paper, we presented the results of two subjective studies designed to investigate two aspects of playback rate in adaptive video streaming.

First, we investigated how drops in playback rate are perceived by users when compared to two main QoE factors: video quality and rebuffering. We conducted a study that involved 54 test sequences and 40 participants, evaluating 4 forms of playback rate drops, 2 quality drops, and 2 rebuffering variations across 6 content genres. We found that playback rate drops of 0.8 and 0.9, along with a medium video quality drop, were imperceptible for most content, resulting in little to no negative impact on QoE. We also found that lower playback rates, of 0.6 and 0.7, were comparable to rebuffering events as well as to a low quality drop, all of which significantly reduced the perceived quality.

In the second study, we investigated gradual and instant changes in playback rate in terms of their impact on QoE. We performed a study consisting of 48 test sequences, evaluating 4 target playback rates with 3 variations of each, across 4 pieces of content, with 80 participants recruited. We found that gradual changes in playback rate can be beneficial, but only in limited cases, with the greatest improvement observed at the lowest target playback rate of 0.7. The impact of gradual changes varied with the magnitude of change across different target playback rates and content genres.

In summary, in this paper we presented the results of two subjective studies to further investigate the QoE impact of changes in playback rate, with 102 test sequences evaluated and 120 participants recruited in total. We plan to expand this work in two different directions. In our studies, we observed differences in the perceptibility of test conditions across different content genres. More work is needed to determine which content and scene characteristics correlate with the perceptibility of changes in playback rate. Additionally, in the future we plan to evaluate more variations in playback rate transitions.

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] I. 23009-1, "Information technology-dynamic adaptive streaming over http (dash)-part 1: Media presentation description and segment formats," 2019.

[2] R. 8216, "Http live streaming," 2017. [Online]. Available: https://tools.ietf.org/html/rfc8216

[3] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann, "A survey on bitrate adaptation schemes for streaming media over http," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 562–585, 2019.

[4] 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.

[5] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of http video streaming," in *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*, 2011, pp. 485–492.

[6] 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.

[7] 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.

[8] 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.

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

[10] M. N. Garcia, W. Robitza, and A. Raake, "On the accuracy of short-term quality models for long-term quality prediction," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, pp. 1–6.

[11] 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.

[12] 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.

[13] 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.

[14] O. F. Aladag, D. Ugur, M. N. Akcay, and A. C. Begen, "Content-aware playback speed control for low-latency live streaming of sports," in *Proceedings of the 12th ACM Multimedia Systems Conference*, 2021, pp. 344–349.

[15] L. Sun, T. Zong, S. Wang, Y. Liu, and Y. Wang, "Tightrope walking in low-latency live streaming: optimal joint adaptation of video rate and playback speed," in *Proceedings of the 12th ACM Multimedia Systems Conference*, ser. MMSys '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 200–213. [Online]. Available: https://doi.org/10.1145/3458305.3463382

[16] T. Karagkioules, R. Mekuria, D. Griffioen, and A. Wagenaar, "Online learning for low-latency adaptive streaming," in *Proceedings of the 11th ACM Multimedia Systems Conference*, ser. MMSys '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 315–320.

[17] A. Bentaleb, M. N. Akcay, M. Lim, A. C. Begen, and R. Zimmermann, "Catching the moment with lol+ in twitch-like low-latency live streaming platforms," *IEEE Transactions on Multimedia*, pp. 1–1, 2021.

[18] C. Gutterman, B. Fridman, T. Gilliland, Y. Hu, and G. Zussman, "Stallion: Video adaptation algorithm for low-latency video streaming," in *Proceedings of the 11th ACM Multimedia Systems Conference*, ser. MMSys '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 327–332. [Online]. Available: https://doi.org/10.1145/3339825.3397044

[19] T. Lyko, M. Broadbent, N. Race, M. Nilsson, P. Farrow, and S. Appleby, "Llama - low latency adaptive media algorithm," in *2020 IEEE International Symposium on Multimedia (ISM)*, 2020, pp. 113–121.

[20] B. Rainer and C. Timmerer, "A quality of experience model for adaptive media playout," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 177–182.

[21] P. Pérez, N. García, and A. Villegas, "Subjective assessment of adaptive media playout for video streaming," in *International Conference on Quality of Multimedia Experience (QoMEX)*, 2019.

[22] Recommendation ITU-T P. 910, "Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union*, 2023.

[23] 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