

Beyond Bitrate: Understanding the QoE Impact of Playback Rate and Seeking in Adaptive Video Streaming

TOMASZ LYKO, School of Computing and Communications, Lancaster University, United Kingdom

YEHIA ELKHATIB, School of Computing Science, University of Glasgow, United Kingdom

RAJIV RAMDHANY, BBC R&D North, United Kingdom

NICHOLAS RACE, School of Computing and Communications, Lancaster University, United Kingdom

Quality of Experience (QoE) is a key component in adaptive bitrate (ABR) streaming. While the effects of delivery disruptions - such as changes in video quality or rebuffering - have been extensively studied, the impact of playback rate variations remains relatively unexplored. Existing work has examined the QoE impact of playback rate in isolation, without comparing it to other common ABR streaming artifacts such as video quality variations, rebuffering, or seeking. Moreover, the role of gradual playback rate transitions has not been explored. This paper addresses these gaps through four large-scale subjective studies that provide a systematic QoE evaluation of playback rate. We compare the acceptability of playback rate changes against quality degradations and rebuffering, and assess rate increases relative to seeking, both of which are strategies for maintaining the desired latency in low-latency streaming. We further investigate how gradual versus instantaneous playback rate transitions affect QoE. Through our subjective studies we identify levels of slowed-down playback that are imperceptible to users, and can be combined with video quality adaptation in ABR algorithms to reduce rebuffering. Moreover, we identify imperceptible rates of speeded-up video playback, which can be used as part of catch-up mechanisms to maintain a desired latency in live streaming, offering a less detrimental alternative to seeking events, which we found to significantly degrade QoE. This work presents the first systematic comparison of playback rate variations with video quality degradation, rebuffering, and seeking. The findings extend our understanding of QoE in adaptive video streaming and provide actionable design guidelines for video players to improve the user experience of streaming services.

CCS Concepts: • **Information systems** → **Multimedia streaming**.

Additional Key Words and Phrases: playback rate, seeking events, adaptive streaming, QoE, assessment, subjective study

ACM Reference Format:

Tomasz Lyko, Yehia Elkhatib, Rajiv Ramdhany, and Nicholas Race. . Beyond Bitrate: Understanding the QoE Impact of Playback Rate and Seeking in Adaptive Video Streaming. In *Proceedings of ACM Trans. Multimedia Comput. Commun. Appl. (ACM TOMM)*. ACM, New York, NY, USA, 24 pages. <https://doi.org/10.1145/3796724>

1 Introduction

Video streaming over HTTP, such as DASH [18] and HLS [16], typically makes use of adaptive bitrate (ABR) algorithms in video players that continually learn which video quality (bitrate representation) optimally fits the prevailing network conditions and reacts to key network events such as congestion and packet loss. For maximum Quality of Experience (QoE), these algorithms must maintain a balance of achieving high video quality whilst minimising rebuffering and bitrate oscillations. To do so, client-side ABR algorithms will measure a range of metrics including playback

Authors' Contact Information: Tomasz Lyko, t.lyko@lancaster.ac.uk, School of Computing and Communications, Lancaster University, Lancaster, United Kingdom; Yehia Elkhatib, School of Computing Science, University of Glasgow, Glasgow, United Kingdom; Rajiv Ramdhany, BBC R&D North, Manchester, United Kingdom; Nicholas Race, School of Computing and Communications, Lancaster University, Lancaster, United Kingdom.

ACM TOMM,

©

<https://doi.org/10.1145/3796724>

buffer occupancy, frequency of rebuffering and quality oscillations, whilst attempting to estimate the available bandwidth, in order to select the most appropriate bitrate level for the next segment to be downloaded.

The majority of research in client-side bitrate adaptation focuses on either *bandwidth-based adaptation* schemes, where bandwidth fluctuations and congestion are detected based on video segment download times, or *buffer-based adaptation* schemes, where playback buffer occupancy is used as a criterion to select the next segment's bitrate [9]. While selecting a lower bitrate is the primary strategy to respond to network congestion and avoid rebuffering, less attention has been paid to a complementary strategy: adjusting the video and audio playback *rate*. Client-side playback rate adaptation can be used *alongside* existing bitrate switching mechanisms in order to reduce the occurrence of rebuffering in adaptive video streaming. However, it is also critical to understand the impact of slowing down video and audio playback on QoE.

In recent years, attention in video streaming has shifted to *live* streaming, and specifically the challenge of achieving low latency in live streaming scenarios. This has resulted in a number of low latency adaptive streaming standards (CMAF [19], LL-DASH [10], LL-HLS [17]). In live streaming, rebuffering events typically result in an increase in a client's latency; as a pause in the playout of content leads to the client moving further away from live. Once the client's latency drifts away from live, it can only be reduced through the use of a catch-up mechanism employed by the video player. A catch-up mechanism can utilise two methods to reduce a client's latency. The first involves increasing the playback rate, which will result in the client slowly converging towards the desired latency. The second involves the video player skipping ahead (seeking) by a number of seconds, quickly achieving the desired latency but at the cost of content not being shown to the viewer.

Playback rate adaptations, such as slowing down playback in order to avoid rebuffering, or speeding up playback to reduce latency, have significant *potential* to enhance video streaming experiences. However, the impact this range of adaptations has on QoE is not well understood. This research presents a detailed investigation into the impact of playback rate adjustments on QoE in adaptive video streaming. While previous studies have explored the effects of video quality drops and rebuffering events, this research is the first to systematically compare these QoE factors with playback rate fluctuations. Additionally, this research is the first to evaluate the QoE impact of seeking events in adaptive video streaming, and to compare them against playback rate adjustments. Through extensive subjective user studies, we provide a comprehensive evaluation of the QoE impact of playback rate changes in relation to other quality impairments, offering new insights for ABR and latency management.

This work explores and evaluates the QoE impact of *client-side playback-rate adaptation* in adaptive video streaming, addressing the following research questions (RQs):

- RQ1: How are drops in playback rate perceived by users across different content genres compared to other QoE-affecting factors in adaptive video streaming?
- RQ2: Conversely, how are increases in playback rate perceived by users in relation to other QoE-affecting factors?
- RQ3: Furthermore, how are increases in playback rate perceived by the user when compared to seeking events?
- RQ4: Are gradual changes in playback rate perceived better than more drastic adjustments, and can such a strategy enable greater variations in altered playback rate?

RQ1 is motivated by the exploration of alternative means of dealing with delivery issues while maintaining high QoE. RQ2 and RQ3 investigate the QoE impact of the two methods used by catch-up mechanisms to reduce latency in live adaptive streaming. RQ4 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 include:

- (1) Evaluation of playback rate drops and increases when compared to two main QoE factors in adaptive video streaming: video quality and rebuffering, using two subjective studies, in total consisting of 108 test sequences, 6 pieces of content and 82 participants.
- (2) Evaluation of seeking events and playback rate increases, using a subjective study made of 54 test sequences, 6 pieces of content and 42 participants.
- (3) Investigation of gradual changes in playback rate, across two magnitudes of change, using a subjective study consisting of 48 test sequences and 80 participants.

Based on the results of the four subjective studies outlined above, this paper also contributes a number of improvements to the design of video players in adaptive video streaming. It highlights cases where playback rate adaptation can safely be used to supplement bitrate adaptation to reduce the amount of rebuffering and where playback rate adaptation is superior to seeking in reducing the client's latency in live streaming scenarios. This work significantly extends our conference paper [25], where only drops in playback rate and gradual changes were evaluated (answering RQ1 and RQ4). This paper offers a complete evaluation of playback rate and its impact on QoE, by presenting two additional subjective studies and addressing two more research questions, focusing on increases in playback rate and seeking events - as both can be found in low latency streaming. Our final contribution is the release of the dataset containing the full results from all four subjective studies for the benefit of the research community [1].

2 Background and Related Work

Quality of Experience (QoE) is an important aspect of ABR streaming. A number of factors influence QoE, as outlined in the survey by Barman et al. [7]. Rebuffering events, which occur when the client buffer is fully depleted causing the playback to stall, have a strong negative impact on QoE, with both the duration and frequency being relevant [3, 5, 6, 11, 12, 14, 27, 32, 36, 38, 39]. Video quality of each segment is another important factor, with lower bitrates having negative impact on QoE; however, the relationship between the bitrate and its perceptibility is non-linear [13, 26, 33, 37]. Furthermore, quality switches, where the video quality between two consecutive segments changes, can potentially impact the QoE, with high frequency of switches being detrimental [13, 33, 37]. This is especially the case with multi-level switches, where there is a change in quality across more than one level between two consecutive segments [24, 33, 35].

In recent years, new ABR algorithms for low latency live adaptive streaming have been proposed [2, 8, 15, 20, 21, 23, 34]. These ABRs have been designed to operate in scenarios where the client buffer is severely restricted by the need to maintain low latency. In such conditions, the playback rate might be adjusted by the ABR to aid rebuffering avoidance, or by the video player to maintain a specific target latency. Additionally, the playback rate might be increased to maintain a specific target latency by the client, using a catch-up mechanism [22].

There is very limited literature on the relationship between playback rate and QoE. The following three studies investigated this issue to different extents. Rainer and Timmerer [30] studied the QoE impact of decreasing and increasing playback rates 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. Mu et al. [28] evaluated the QoE impact of increases in playback rate, in terms of magnitude and duration of altered playback rate. The authors found that both of these factors had non-linear impact on the perceptibility of increases in playback rate, and that a 20% increase was imperceptible to most users. Pérez et al. [29] conducted a study 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 was considered safe. All three studies

investigated playback rate in isolation; that is, without considering any variation in video quality or the occurrence of rebuffering during playback. Current literature has also not investigated the use of seeking events during playback and the associated impact on QoE.

In this paper, we address this gap in the literature by evaluating playback rate, reductions in video quality and rebuffering events in the same study. Furthermore, we compare playback rate increases and seeking events, as both can be used to maintain target latency in low latency adaptive streaming. Finally, we investigate whether gradual changes in playback rate are less perceptible – an area of study currently unexplored in the literature.

3 Methodology

In order to perform a comprehensive evaluation of playback rate, we have designed four subjective studies - one for each RQ outlined in Section 1. All four studies share some common methodologies, which are detailed in this section.

First, we outline the testing procedure selected for our studies, then we describe an online survey tool we developed to gather reliable results, and the recruitment of participants using an online crowdsourcing platform. Thereafter, we describe the selected clips of various content genres for stimuli used in the studies. Finally, we detail the encoding process behind the test sequence creation for our studies to produce multiple types of quality impairments.

3.1 Testing Procedure

We selected the Absolute Category Rating (ACR), outlined in ITU Rec. P.910 [31], as our video quality assessment method. In ACR, test sequences are presented individually, and participants are 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 three studies, 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.

In order to collect reliable results and avoid participant fatigue, we limited the duration of each test to a maximum of 10 minutes. For studies where the number of test clips exceeded this limit, we divided the test clips into smaller subsets based on content and showed each participant only a single subset of clips. This required more participants to be recruited overall. We describe this in more detail in the set-up section of each study.

3.2 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 playback began. Additionally, clips are always presented in full-screen mode with the sound enabled.

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 remainder of the survey.

Next, the survey proceeds to the video quality assessment stage, where the participant is shown and asked to score the test sequences, one at a time. The first few test sequences are 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 are arranged at random with a single constraint to ensure that the same content was never presented twice in a row.



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

Table 1. A summary of the properties of the content used in our studies.

Clip	Genre	SI	TI
Vegesaur	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

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.

All four studies were performed using an ethical crowdsourcing platform Prolific¹. 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. To ensure that the responses given by participants were reliable, we incorporated mechanisms to detect whether the participant had watched all of the test sequences completely before rating them. We also tracked whether the participant attempted to adjust the playback rate or seek whilst watching the clips. Any responses where participants did not watch the clips in full or tried to seek or adjust the playback rate were discarded. The study was approved by our ethics committee.

3.3 Content

As observed in a previous study on this topic [29], the impact of changes in playback rate can vary across content genres. Given this, we sourced six clips of different genres. Figure 1 shows the thumbnails of the selected clips and Table 1 describes the clips in terms of calculated Spatial Information (SI) and Temporal Information (TI) according to ITU P.910 [31], as well as content genre. Each clip was 11s long at 25FPS, and had resolution and bitrate greater than or equal to 1920x1080 at 6000kbps. We selected content to cover both on-demand and live streaming scenarios, aiming to provide a comprehensive evaluation of playback rate that applies to both. The clip *Football* represents sports content, typically found in live streaming services, while clip *Forza* represents gaming with commentary – content commonly found on live streaming platforms such as YouTube or Twitch.

3.4 Encoding

In order to prepare the test sequences, we used `ffmpeg`² to introduce the desired video and audio quality impairments. Clips were encoded using the H.264 and AAC codecs.

To create the test sequences, the 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

¹<https://www.prolific.com/>

²<https://ffmpeg.org/>

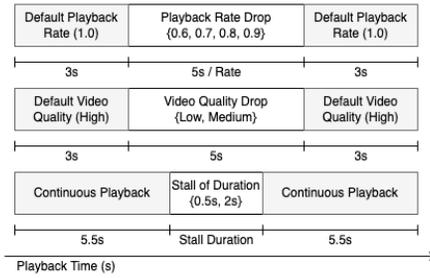


Fig. 2. Test conditions used in the first study, describing playback rate drops, video quality drops, and rebuffering events.

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. Fourth, seeking events were introduced by the removal of segments where seeking occurs, emulating the behaviour of a real video player where a seeking event would cause part of the video to be skipped and not displayed to the user.

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 [4]: 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.

4 Playback Rate Drops 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 drops in playback rate when compared to drops in video quality and rebuffering events.

4.1 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 decreases in playback rate, whereby playback was slowed to one of four rates throughout the middle section of the clip. The next two test conditions contained a single rebuffering event during the middle of the clip, of 0.5s or 2s duration. 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: Vegesaur, Football, and Forza, while set B contained all test sequences based on clips: Frozen Planet, Top Gear, and Wolf Hall. This resulted in each participant rating 27 clips.

We recruited 40 participants in total. 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%,

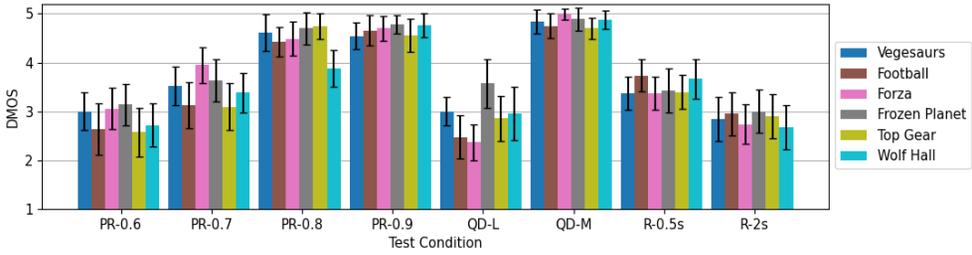


Fig. 3. Results of the first subjective study, with Differential Mean Opinion Scores (DMOS) plotted for each test condition and content.

Table 2. Table showing the p-values for the paired T-test between all test sequences in the first study. Each cell contains the values for each content in the following order: Vegesaurus, Football, Forza, Frozen Planet, Top Gear, and Wolf Hall.

	PR-0.6	PR-0.7	PR-0.8	PR-0.9	Ref	QD-L	QD-M	R-0.5s	R-2s
PR-0.6		0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	1.0,0.3,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.1,0.0,0.2,0.0,0.0,0.0	0.5,0.1,0.2,0.0,0.0,0.0
PR-0.7	0.0,0.0,0.0,0.0,0.0,0.0		0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.5,0.0,0.0,0.0,0.0,0.0	0.0,0.2,0.0,0.0,0.0,0.0
PR-0.8	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0		0.7,0.1,0.3,0.6,0.3,0.0	0.2,0.0,0.0,0.8,1.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.4,0.1,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0
PR-0.9	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.7,0.1,0.3,0.6,0.3,0.0		0.0,0.2,0.1,0.1,0.3,0.6	0.0,0.0,0.0,0.0,0.0,0.0	0.1,0.8,0.0,0.1,0.8,0.3	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0
Ref	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.2,0.0,0.0,0.8,1.0,0.0	0.0,0.2,0.1,0.1,0.3,0.6		0.0,0.0,0.0,0.0,0.0,0.0	0.8,0.2,0.3,0.5,0.2,0.5	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0
QD-L	1.0,0.3,0.0,0.1,0.2,0.3	0.0,0.0,0.0,0.8,0.3,0.1	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0		0.0,0.0,0.0,0.0,0.0,0.0	0.1,0.0,0.0,0.6,0.0,0.1	0.6,0.0,0.1,0.0,0.9,0.3
QD-M	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.4,0.1,0.0,0.4,0.4,0.0	0.1,0.8,0.0,0.1,0.8,0.3	0.8,0.2,0.3,0.5,0.2,0.5	0.0,0.0,0.0,0.0,0.0,0.0		0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0
R-0.5s	0.1,0.0,0.2,0.2,0.0,0.0	0.5,0.0,0.0,0.4,0.1,0.2	0.0,0.0,0.0,0.0,0.0,0.3	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.1,0.0,0.0,0.6,0.0,0.1	0.0,0.0,0.0,0.0,0.0,0.0		0.0,0.0,0.0,0.1,0.0,0.0
R-2s	0.5,0.1,0.2,0.5,0.2,1.0	0.0,0.2,0.0,0.0,0.3,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.0,0.0,0.0	0.6,0.0,0.1,0.0,0.9,0.3	0.0,0.0,0.0,0.0,0.0,0.0	0.0,0.0,0.0,0.1,0.0,0.0	

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.

4.2 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 [31]. Additionally, we performed a paired T-test ($\alpha = 0.05$) to determine whether differences in MOS between test conditions were statistically significant, with p-values shown in Table 2.

Impact of Individual Quality Impairments. The results presented in Figure 3 show the impact of drops in the playback rate alone. For all playback rate test sequences, the first and last 3s of the clip would play at a normal rate (1.0x) with the middle of the clip playing at the reduced playback rate. These test sequences were encoded at the highest level and did not contain 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 Vegesaurus 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: Vegesaur, Frozen Planet and Top Gear. This rate was most noticeable for the Wolf Hall clip with a DMOS of 3.9. Compared to the playback rate of 0.9, PR-0.8 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 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 differences between these test sequences and the hidden reference were statistically insignificant. This suggests that this quality impairment was imperceptible to participants and therefore did not have an impact on QoE, despite the bitrate 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 0.5s or 2s duration. 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 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 perceived quality.

Playback Rate Drops 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 Vegesaur 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: Vegesaur, 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. *A playback rate drop of 0.6 was perceived as poorly as a low drop in quality.*

Playback Rate Drops 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 for 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.*

5 Playback Rate Increases vs. Other QoE Factors

In this section, we analyse the results of the second subjective study, designed to investigate the QoE impact of increases in playback rate. Similar to the first study, we compare the playback rate changes to video quality drops and rebuffering events, with these QoE factors acting as reference points between the two studies, allowing for further comparison between the temporary drops and increases in playback rate. Furthermore, playback rate adaptation for the purposes of rebuffering avoidance in live streaming will require both, a playback rate drop to avoid rebuffering, and a subsequent playback rate increase to reduce the latency to the original level prior to the playback rate drop. Therefore, it is important to compare both playback rate drops and increases against other QoE factors.

5.1 Study Set-up

Figure 4 illustrates the eight test conditions created for this study. These test conditions were shown to each participant along with a hidden reference test sequence that did not contain any quality impairments. Four test conditions focused on increases in playback rate, whereby the playback speed was increased at one of four rates during the middle section of the clip. The other four test conditions were exactly the same as in the first study. Two containing a single rebuffering event, of 0.5s or 2s, and two where the video quality was reduced to a low or medium quality level, each occurring during the middle of the clip.

As in the first study, all six sourced clips were used, 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: Vegesaur, Football, and Forza, while set B contained

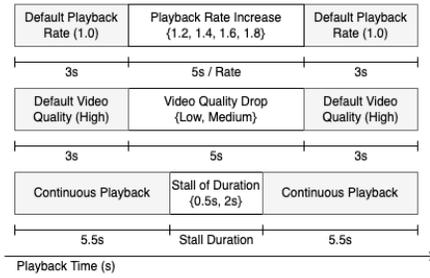


Fig. 4. Test conditions used in the second study, describing playback rate increases, video quality drops, and rebuffering events.

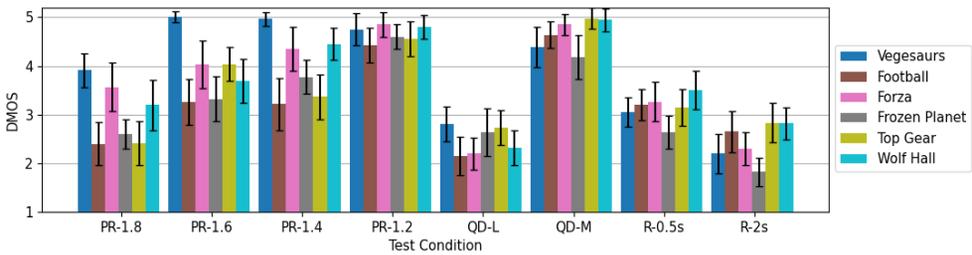


Fig. 5. Results of the second subjective study, with DMOS plotted for each test condition and content.

all test sequences based on: Frozen Planet, Top Gear, and Wolf Hall – following the same content allocation from the first study. This resulted in each participant rating 27 clips.

In total, 42 participants were recruited. Each set of test sequences was watched by 21 participants. 50% of participants were male, and 50% female. The age groups can be described as 18-28 for 14%, 29-38 for 28%, 39-48 for 24%, 49-58 for 24%, 59-68 for 7%, and 69 or older for 2% of participants. 93% of the participants reported having no previous experience in video quality assessment studies.

5.2 Results

The results of the second study are shown in Figure 5 and are presented in a manner similar to the results shown in the previous section. DMOS were plotted against the test conditions, organised by content, with confidence intervals of 95% plotted in the error bars. Two-point crushing function was applied (ITU Rec. P.910 [31]). Furthermore, we performed a paired T-test ($\alpha = 0.05$) to determine whether the differences in MOS between test conditions were statistically significant, with p-values shown in Table 3.

Impact of Individual Quality Impairments. In the results presented in Figure 5, we can observe the impact of each QoE factor individually, including increases in playback rate, drops in video quality, and rebuffering. We start by considering the increases in playback rate, where for all of the test sequences, the first and last 3s of the clip would play at a normal rate (1.0x) with the middle of the clip playing at the increased target playback rate (for 5s). These test sequences were encoded at the highest level and did not contain any rebuffering.

As the playback rate increased, the average DMOS across all content decreased. For the lowest tested playback rate increase of 1.2 (presented in test condition PR-1.2) the average DMOS equalled 4.7. The differences between these test sequences and the hidden reference were statistically

Table 3. Table showing the p-values for the paired T-test between all test sequences in the second study. Each cell contains the values for each content in the following order: Vegesaus, Football, Forza, Frozen Planet, Top Gear, and Wolf Hall.

	PR-1.8	PR-1.6	PR-1.4	PR-1.2	Ref	QD-L	QD-M	R-0.5s	R-2s
PR-1.8		0.0,0.0,0.0, 0.0,0.0,0.1	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.1,0.0, 0.9,0.2,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.2, 0.9,0.0,0.3	0.0,0.2,0.0, 0.0,0.0,0.1
PR-1.6	0.0,0.0,0.0, 0.0,0.0,0.1		0.7,0.8,0.0, 0.0,0.0,0.0	0.1,0.0,0.0, 0.0,0.1,0.0	0.2,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.1,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.6,0.0, 0.1,0.0,0.4	0.0,0.0,0.0, 0.0,0.0,0.0
PR-1.4	0.0,0.0,0.0, 0.0,0.0,0.0	0.7,0.8,0.0, 0.0,0.0,0.0		0.1,0.0,0.0, 0.0,0.0,0.1	0.3,0.0,0.1, 0.0,0.0,0.3	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.2,0.0,0.0	0.0,0.8,0.0, 0.0,0.5,0.0	0.0,0.0,0.0, 0.0,0.1,0.0
PR-1.2	0.0,0.0,0.0, 0.0,0.0,0.0	0.1,0.0,0.0, 0.0,0.1,0.0	0.1,0.0,0.0, 0.0,0.0,0.1		0.6,0.0,0.5, 0.0,1.0,0.8	0.0,0.0,0.0, 0.0,0.0,0.0	0.3,0.3,1.0, 0.1,0.0,0.1	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
Ref	0.0,0.0,0.0, 0.0,0.0,0.0	0.2,0.0,0.0, 0.0,0.0,0.0	0.3,0.0,0.1, 0.0,0.0,0.3	0.6,0.0,0.5, 0.0,1.0,0.8		0.0,0.0,0.0, 0.0,0.0,0.0	0.1,0.2,0.5, 0.0,1.0,1.1	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
QD-L	0.0,0.1,0.0, 0.9,0.2,0.0	0.0,0.0,0.0, 0.1,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0		0.0,0.0,0.0, 0.0,0.0,0.0	0.2,0.0,0.0, 1.0,0.0,0.0	0.0,0.0,0.5, 0.0,0.5,0.0
QD-M	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.2,0.0,0.0	0.3,0.3,1.0, 0.1,0.0,0.1	0.1,0.2,0.5, 0.0,0.1,0.1	0.0,0.0,0.0, 0.0,0.0,0.0		0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
R-0.5s	0.0,0.0,0.2, 0.9,0.0,0.3	0.0,0.6,0.0, 0.1,0.0,0.4	0.0,0.8,0.0, 0.0,0.5,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.2,0.0,0.0, 1.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0		0.0,0.0,0.0, 0.0,0.0,0.0
R-2s	0.0,0.2,0.0, 0.0,0.0,0.1	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.1,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.5, 0.0,0.5,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	

insignificant for half of the clips: Vegesaus, Forza, and Wolf Hall. At a playback rate of 1.4 (PR-1.4), the average DMOS decreased to 4.0, with differences being statistically insignificant for the same clips as for PR-1.2: Vegesaus, Forza, and Wolf Hall. These three clips experienced a significantly smaller decrease in DMOS of just 0.2 on average, when compared with the three other clips (Football, Frozen Planet, and Top Gear) where the average decrease was 1.1. At a playback rate of 1.6, represented in test condition PR-1.6, the average DMOS equalled 3.9, only a 0.1 decrease from PR-1.4, with the differences for clips Vegesaus and Football being statistically insignificant. When compared to PR-1.4, the DMOS increased by 0.1 and 0.6 for clips Football and Top Gear respectively, remained unchanged for Vegesaus, and decreased by 0.3, 0.5, and 0.7 for clips Forza, Frozen Planet and Wolf Hall respectively. At the highest tested playback rate increase of 1.8 (PR-1.8), the average DMOS decreased to 3.2, with all differences statistically significant, and the decrease present across all clips – even the Vegesaus clip for which all other playback rate increases were imperceptible.

When comparing the scores of test sequences featuring quality drops, we can observe the impact of this quality impairment alone – similarly to the first study. 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 4.7. All of the differences between these test sequences and the hidden reference were statistically insignificant – except the Frozen Planet clips which was scored the lowest. For the test sequences featuring a quality drop to the lowest level, QD-L, the average DMOS dropped significantly to 2.5 on average across all content with all differences being statistically significant. Both test conditions were scored lower than the equivalent conditions from the first study, presented in the previous section, by 0.1 and 0.4 for QD-M and QD-L, respectively.

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, as in the previous section. In these clips, there was a single rebuffering event in the middle of the video, of 0.5s or 2s duration. 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 3.1, 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 2.4, on average, across all content, with all differences being statistically significant again. Both rebuffering variants were scored lower than the equivalent conditions in the first study presented in the previous section, where drops in playback rate were evaluated.

FINDING 4. *There is a great disparity between how playback rate increases are perceived across different content genres.*

Playback Rate Increases vs. Quality Drops. When comparing the test sequences containing variations in playback rate increases and quality drops, the following observations can be made. The test condition QD-M, where the video quality dropped to a medium level in the middle of the clip, was imperceptible with an average DMOS of 4.7 across all content. Similarly, PR-1.2, where the playback rate of the middle section of the clip was increased to 1.2, was also imperceptible, with an average DMOS of 4.7. Differences in DMOS between these two test conditions were statistically insignificant for all content except the Top Gear clip.

PR-1.4 was scored 0.7 lower than QD-M on average across all content, with the differences for all clips except Frozen Planet being statistically significant. PR-1.6 still achieved 1.4 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 all clips except Frozen Planet.

The highest playback rate increase of 1.8 (PR-1.8) was scored higher than QD-L, with 0.5 higher average DMOS across all clips, with statistically insignificant differences for half of the clips: Football, Frozen Planet, and Top Gear – suggesting that playback rate increase of 1.8 was perceived as poorly as a quality drop to low level. PR-1.8 was scored higher than QD-L by 1.1, 1.4, and 0.9 for Vegesaur, Forza, and Wolf Hall respectively, indicating that for these clips, even the highest tested playback rate increase was significantly better perceived than a drop to low quality.

FINDING 5. *All variations of playback rate increases were perceived better than a drop to low quality for half of the clips.*

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

Test condition R-0.5s, which contained a single rebuffering event of 0.5s, was scored on average only 0.1 higher than the highest playback rate increase tested of 1.8, present in test condition PR-1.8. However, this increase was not uniform across all content. For the Vegesaur clip, PR-1.8 was scored 0.9 higher than R-0.5s. In the case of Football and Top Gear clips, PR-1.8 achieved lower DMOS than R-0.5s, by 0.8 and 0.7, respectively. Football's R-0.5s was scored 0.1 lower than its PR-1.6, which contained a playback rate increase of 1.6, with the difference being statistically insignificant. On the other hand, Top Gear's R-0.5s was scored 0.9 lower than its PR-1.6. For the three remaining clips Forza, Frozen Planet, and Wolf Hall, the differences between R-0.5s and PR-1.8 ranged between -0.3 and 0.3 – all of which were statistically insignificant.

A single rebuffering event of 2s, contained in test condition R-2s, achieved on average 0.6 lower DMOS than the highest playback rate increase of 1.8 (PR-1.8). The DMOS was 1.7, 1.3, 0.8, and 0.4 lower for Vegesaur, Forza, Frozen Planet, and Wolf Hall clips respectively – however, the DMOS difference for Wolf Hall was statistically insignificant. For the two remaining clips, the DMOS was higher for R-2s, specifically by 0.3 and 0.4 for Football and Top Gear respectively – with the former being statistically insignificant. This suggests that three out of six clips were significantly more sensitive to increases in playback rate in terms of its perceptibility.

FINDING 6. *Rebuffering of 2s was perceived to be worse than an increase of 1.8 in playback rate for half of the content.*

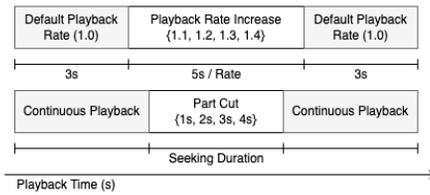


Fig. 6. Test conditions used in the third study, describing playback rate increases and seeking events.

6 Playback Rate Increases vs. Seeking Events

In this section, we present and describe the results of the third subjective study, designed to investigate the QoE impact of increases in playback rate when compared to seeking events. Both playback rate increases and seeking events can be employed to maintain the desired target latency in live adaptive streaming. This is the first study to evaluate the perceptibility of seeking events, which are currently under-represented in the literature despite having the potential to impact QoE.

6.1 Study Set-up

We created eight test conditions, as illustrated in Figure 6, which were shown to each participant along with a hidden reference test sequence that did not contain any quality impairments. The first four test conditions focused on increases in playback rate, whereby the playback speed was increased to one of four rates (1.1, 1.2, 1.3 and 1.4) during the middle section of the clip. These rates were chosen based on the findings from the second study, indicating that increases in playback rate would be imperceptible for most content somewhere within this range. The remaining four test conditions are designed to explore scenarios involving seeking, with a single seek event occurring in the middle of the clip, of 1, 2, 3, and 4 seconds. The seeking event in each test condition is instantaneous, which represents the best-case scenario, whereby the video player jumps ahead to a segment that is already fetched and available in the client buffer.

Similar to the two previous studies, all six sourced clips were used. This resulted in 54 test sequences which were divided into two sets with each containing all test sequences belonging to 3 out of 6 clips. As in the first two studies, Set A contained all test sequences based on clips: Vegesaur, Football, and Forza, whereas Set B contained all test sequences based on clips: Frozen Planet, Top Gear, and Wolf Hall. This resulted in each participant rating 27 clips.

We recruited 42 participants. Test sequences belonging to Set A were viewed by 20 participants, while Set B was viewed by 22 participants. 22 participants were male, and 20 were female. The age groups can be described as: 18-28 for 19%, 29-38 for 36%, 39-48 for 19%, 49-58 for 17%, 59-68 for 7%, and 69 or older for 2% of participants. 90% of the participants reported having no previous experience in video quality assessment studies.

6.2 Results

Figure 7 presents the comprehensive results of the third study, organised by content. Similar to the first two studies, DMOS were calculated using the hidden reference test sequence without quality impairments and plotted against the test conditions, with confidence intervals of 95% plotted in the error bars. Two-point crushing function was applied (ITU Rec. P.910 [31]) and a paired T-test ($\alpha = 0.05$) was used to determine whether differences were statistically significant, with p-values shown in Table 4.

Impact of Individual Quality Impairments. First, we analyse the QoE impact of each quality impairment alone, starting with increases in playback rate. In these test sequences, the first and

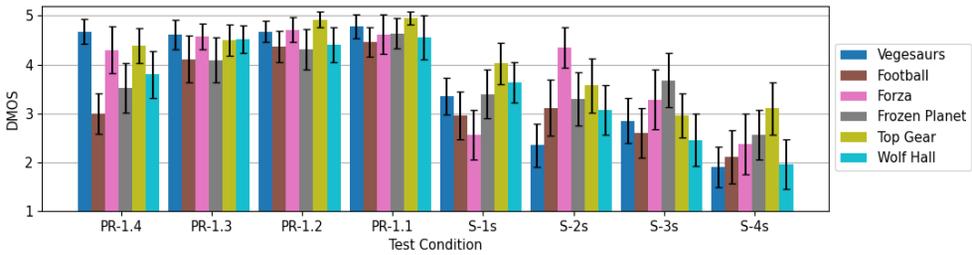


Fig. 7. Results of the third subjective study, with DMOS plotted for each test condition and content.

Table 4. Table showing the p-values for the paired T-test between all test sequences in the third study. Each cell contains the values for each content in the following order: Vegeosaurs, Football, Forza, Frozen Planet, Top Gear, and Wolf Hall.

	PR-1.4	PR-1.3	PR-1.2	PR-1.1	Ref	S-1s	S-2s	S-3s	S-4s
PR-1.4		0.5,0.0,0.1, 0.0,0.4,0.0	1.0,0.0,0.0, 0.0,0.0,0.0	0.4,0.0,0.1, 0.0,0.0,0.0	0.1,0.0,0.0, 0.0,0.0,0.0	0.0,0.8,0.0, 0.7,0.1,0.4	0.0,0.6,0.5, 0.4,0.0,0.0	0.0,0.1,0.0, 0.4,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
PR-1.3	0.5,0.0,0.1, 0.0,0.4,0.0		0.6,0.2,0.5, 0.5,0.0,0.7	0.3,0.2,1.0, 0.1,0.0,0.6	0.1,0.0,0.1, 0.0,0.1,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.6, 0.0,0.0,0.0	0.0,0.0,0.0, 0.2,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
PR-1.2	1.0,0.0,0.0, 0.0,0.0,0.0	0.6,0.2,0.5, 0.5,0.0,0.7		0.4,0.6,0.5, 0.0,1.0,0.4	0.1,0.0,0.1, 0.0,0.5,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.3, 0.0,0.0,0.0	0.0,0.0,0.0, 0.1,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
PR-1.1	0.4,0.0,0.1, 0.0,0.0,0.0	0.3,0.2,1.0, 0.1,0.0,0.6	0.4,0.6,0.5, 0.0,1.0,0.4		0.3,0.0,0.1, 0.3,0.5,0.3	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.7, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
Ref	0.1,0.0,0.0, 0.0,0.0,0.0	0.1,0.0,0.1, 0.0,0.1,0.0	0.1,0.0,0.1, 0.0,0.5,0.0	0.3,0.0,0.1, 0.3,0.5,0.3		0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.1, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0
S-1s	0.0,0.8,0.0, 0.7,0.1,0.4	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0		0.0,0.3,0.0, 0.7,0.1,0.1	0.0,0.0,0.0, 0.3,0.0,0.0	0.0,0.0,0.5, 0.0,0.0,0.0
S-2s	0.0,0.6,0.5, 0.4,0.0,0.0	0.0,0.0,0.6, 0.0,0.0,0.0	0.0,0.0,0.3, 0.0,0.0,0.0	0.0,0.0,0.7, 0.0,0.0,0.0	0.0,0.0,0.1, 0.0,0.0,0.0	0.0,0.3,0.0, 0.7,0.1,0.1		0.0,0.0,0.0, 0.1,0.0,0.0	0.0,0.0,0.0, 0.0,0.1,0.0
S-3s	0.0,0.1,0.0, 0.4,0.0,0.0	0.0,0.0,0.0, 0.2,0.0,0.0	0.0,0.0,0.0, 0.1,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.3,0.0,0.0	0.0,0.0,0.0, 0.1,0.0,0.0		0.0,0.0,0.0, 0.0,0.5,0.0
S-4s	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.0,0.0	0.0,0.0,0.5, 0.0,0.0,0.0	0.0,0.0,0.0, 0.0,0.1,0.0	0.0,0.0,0.0, 0.0,0.5,0.0	

last 3s would play at normal rate (1.0x), with the middle 5s played at the increased playback rate. These test sequences were encoded at the highest level and did not contain any seeking events.

For the lowest tested playback rate increase of 1.1, represented by test condition PR-1.1, the average DMOS across all content equalled to 4.7, with all differences statistically insignificant – except for the Football clip which was scored the lowest at 4.5. The next playback rate increases of 1.2 and 1.3 (PR-1.2 and PR-1.3) were scored only 0.1 and 0.3, respectively, lower, on average across all clips. The highest playback rate increase tested, 1.4 in test condition PR-1.4, resulted in average across all content DMOS of 3.9, with the Football clip being scored significantly lower than the rest of the content, at DMOS of 3.0. These results indicate that a playback rate increase of up to 1.3 had little to no impact of QoE for all content. This trend continued for playback rate of 1.4, however, only for half of the content: Vegeosaurs, Forza, and Top Gear. The sports content, presented in the Football clip, was the most sensitive to playback rate increases when it comes to perceptibility.

Next, we compare the QoE impact of seeking events alone for various durations. Even the shortest seeking event tested of 1s, presented in test condition S-1s, resulted in a significant reduction in DMOS – which equalled to 3.3 on average across all content. However, a high disparity between the scores across content can be observed, with Forza at the lowest of 2.6, and Top Gear at the highest of 4.0. For test condition S-2s, where a seeking event of 2s was present, the average DMOS remained the same. However, it changed drastically for two of the clips, with Forza now scored the

highest at 4.3 and Vegesaur scored the lowest at 2.4. For seeking events of 3s and 4s, the average DMOS further decreased to 3.0 and 2.3 respectively, with the latter indicating a significant negative impact on QoE. Furthermore, for all seeking durations tested, a high disparity between the scores across content can be observed. For example, at the highest duration of seeking, there was a 1.2 gap between the lowest and the highest score, of Vegesaur and Top Gear respectively.

FINDING 7. *There is a significant disparity between the perceptibility of seeking events across different content.*

Playback Rate Increases vs. Seeking Events.

We now compare the test conditions containing playback rate increases and seeking events.

In general, playback rate increases were perceived as significantly better than seeking events. For playback rate increases of 1.1-1.3, the average across all content DMOS was 1.3-1.1 higher than a seeking event of 1 second, the lowest duration tested. A notable exception can be observed between the scores of PR-1.3 and S-2s for the Forza clip, which was scored exceptionally high in the case of a seeking event of 2s, with the difference between scores equalling to only 0.3 and being statistically insignificant.

The highest tested playback rate increase of 1.4, presented in PR-1.4, was scored on average 0.6 higher than the seeking event of the lowest duration of 1 second, tested in S-1s. However, this varied significantly across content. For the Vegesaur and Forza clips, PR-1.4 was scored higher by 1.3 and 1.7 respectively. For the remaining clips, this difference decreased to 0, 0.1, 0.4, and 0.2 for Football, Frozen Planet, Top Gear, and Wolf Hall respectively – with all differences being statistically insignificant. This indicates that a seeking event of 1 second was perceived as poorly as a playback rate increase of 1.4 for four out of six clips tested.

FINDING 8. *In the vast majority of cases, playback rate increases were perceived significantly better than seeking events.*

7 Gradual vs. Instant Changes in Playback Rate

We turn our attention to the last subjective study which investigates the impact of gradual changes in playback rate. This is inspired by existing literature on the impact of changes in video *quality* in adaptive streaming, which suggests that gradual changes – i.e. one quality level at a time – can offer improved QoE in contrast to making a major change to video quality between two consecutive segments [33, 35]. Motivated by this related work, we set out to find out whether a given target playback rate can become more tolerable to the user if the video player changes the rate gradually towards the target, as opposed to instantly modifying the rate to the target setting. For example, when the video player needs to slow down the playback to 0.7, which has a significant impact on QoE, the video player could start to slow down gradually before reaching the target rate – aiming to minimise its negative QoE impact. The reason for slowing down remains the same as with instant change in playback rate (to avoid rebuffering).

7.1 Study Set-up

For this study, we created 4 sets of test conditions, grouped by a target playback rate of 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, and G-0.05 and G-0.1, where the playback rate gradually changes, by 0.05/s and 0.1/s, respectively, towards the target playback rate. Figure 8 illustrates the test conditions. For both instant and gradual playback rate variations, the test sequence always begins with playback at the default rate (1.0) and finishes at the target playback rate. For each content, the presented part remained the same between instant and gradual variations of each target playback rate: first 9s for 0.8, 1.2, and first 11s for 0.7, 1.3. In total 12 test conditions were created, assessing not only a drop in

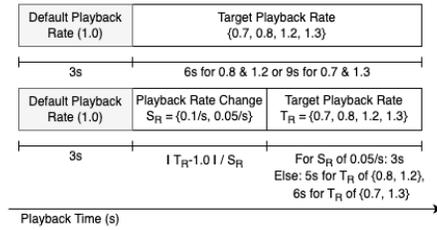


Fig. 8. Test conditions used in the fourth study, describing instant and gradual changes in playback rate.

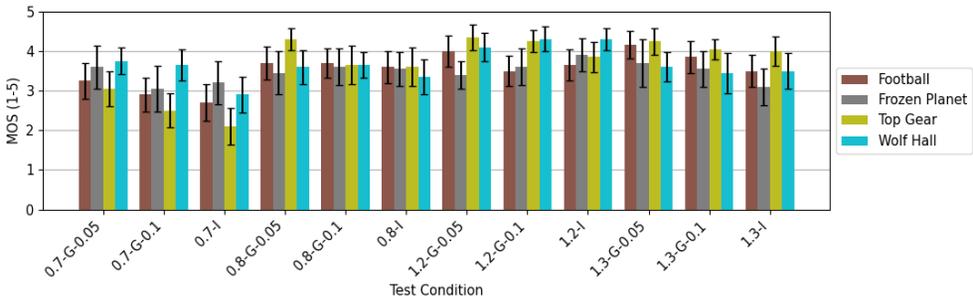


Fig. 9. Results of the fourth subjective study, with Mean Opinion Scores (MOS) plotted for each test condition and clip.

playback rate, but also an increase in playback rate. This study differs from the first three studies as it does not include a hidden reference - a test condition without any quality impairments which can be used to calculate DMOS. Instead, this study follows the standard Absolute Category Rating (ACR) methodology, where Mean Opinion Scores (MOS) of test conditions are analysed, enabling a comparison between instant and gradual changes in playback rate.

Four of the six pieces of content were used for this study. Based on the results of our first study, we selected the clips where changes in playback rate were the most perceptible: 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 four pieces of content, with each content presented at different target playback rate, including one instant and two gradual variations. This resulted in each participant rating 12 clips.

A total of 80 participants were recruited for this study. 20 participants watched each set of test sequences. The gender of the 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. Of the participants, 92.5% reported no previous experience with video quality assessment.

7.2 Results

Figure 9 shows the full results of the fourth 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, with p-values shown in Table 5. We 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.

Table 5. Table showing the p-values for the paired T-test between all test sequences in the fourth study. Each cell contains the values for each content in the following order: Vegesaur, Football, Forza, Frozen Planet, Top Gear, and Wolf Hall.

	0.7-G-0.05	0.7-G-0.1	0.7-I	0.8-G-0.05	0.8-G-0.1	0.8-I	1.2-G-0.05	1.2-G-0.1	1.2-I	1.3-G-0.05	1.3-G-0.1	1.3-I
0.7-G-0.05		0.1,0.0,0.0,0.6	0.0,0.0,0.0,0.0	0.2,0.7,0.0,0.7	0.1,1.0,0.1,0.7	0.2,0.9,0.1,0.2	0.0,0.6,0.0,0.2	0.3,1.0,0.0,0.0	0.1,0.4,0.0,0.0	0.0,0.8,0.0,0.5	0.1,0.9,0.0,0.4	0.4,0.2,0.0,0.3
0.7-G-0.1	0.1,0.0,0.0,0.6		0.2,0.3,0.0,0.0	0.0,0.3,0.0,0.9	0.0,0.2,0.0,1.0	0.0,0.2,0.0,0.4	0.0,0.3,0.0,0.1	0.1,0.2,0.0,0.0	0.0,0.0,0.0,0.0	0.0,0.2,0.0,0.9	0.0,0.2,0.0,0.6	0.1,0.9,0.0,0.6
0.7-I	0.0,0.0,0.0,0.0	0.2,0.3,0.0,0.0		0.0,0.6,0.0,0.1	0.0,0.3,0.0,0.1	0.0,0.3,0.0,0.3	0.0,0.5,0.0,0.0	0.0,0.3,0.0,0.0	0.0,0.1,0.0,0.0	0.0,0.3,0.0,0.0	0.0,0.4,0.0,0.1	0.0,0.8,0.0,0.1
0.8-G-0.05	0.2,0.7,0.0,0.7	0.0,0.3,0.0,0.9	0.0,0.6,0.0,0.1		1.0,0.6,0.0,0.8	0.6,0.7,0.0,0.3	0.4,0.9,0.8,0.1	0.5,0.7,0.8,0.0	0.9,0.2,0.1,0.0	0.1,0.5,0.8,1.0	0.6,0.8,0.3,0.6	0.5,0.3,0.1,0.8
0.8-G-0.1	0.1,1.0,0.1,0.7	0.0,0.2,0.0,1.0	0.0,0.3,0.0,0.1	1.0,0.6,0.0,0.8		0.6,0.8,0.8,0.1	0.3,0.5,0.0,0.1	0.5,1.0,0.0,0.0	0.9,0.4,0.5,0.0	0.1,0.8,0.1,0.8	0.6,0.8,0.2,0.5	0.4,0.1,0.3,0.6
0.8-I	0.2,0.9,0.1,0.2	0.0,0.2,0.0,0.4	0.0,0.3,0.0,0.3	0.6,0.7,0.0,0.3	0.6,0.8,0.8,0.1		0.2,0.6,0.0,0.0	0.7,0.9,0.0,0.0	0.9,0.2,0.3,0.0	0.1,0.7,0.0,0.5	0.4,1.0,0.1,0.8	0.7,0.1,0.2,0.7
1.2-G-0.05	0.0,0.6,0.0,0.2	0.0,0.3,0.0,0.1	0.0,0.5,0.0,0.0	0.4,0.9,0.8,0.1	0.3,0.5,0.0,0.1	0.2,0.6,0.0,0.0		0.0,0.3,0.6,0.4	0.0,0.0,0.0,0.4	0.6,0.4,0.7,0.1	0.5,0.6,0.1,0.1	0.1,0.2,0.2,0.1
1.2-G-0.1	0.3,1.0,0.0,0.0	0.1,0.2,0.0,0.0	0.0,0.3,0.0,0.0	0.5,0.7,0.8,0.0	0.5,1.0,0.0,0.0	0.7,0.9,0.0,0.0	0.0,0.3,0.6,0.4		0.5,0.2,0.0,1.0	0.0,0.8,1.0,0.0	0.1,0.9,0.3,0.0	1.0,0.1,0.4,0.0
1.2-I	0.1,0.4,0.0,0.0	0.0,0.0,0.0,0.1	0.0,0.1,0.0,0.0	0.9,0.2,0.1,0.0	0.9,0.4,0.5,0.0	0.9,0.2,0.3,0.0	0.0,0.0,0.0,0.4	0.5,0.2,0.1,1.0		0.1,0.6,0.2,0.0	0.5,0.3,0.4,0.0	0.6,0.0,0.6,0.0
1.3-G-0.05	0.0,0.8,0.0,0.5	0.0,0.2,0.0,0.9	0.0,0.3,0.0,0.0	0.1,0.5,0.8,1.0	0.1,0.8,0.1,0.8	0.1,0.7,0.0,0.5	0.6,0.4,0.7,0.1	0.0,0.8,1.0,0.0	0.1,0.6,0.2,0.0		0.3,0.6,0.3,0.6	0.0,0.0,0.2,0.7
1.3-G-0.1	0.1,0.9,0.0,0.4	0.0,0.2,0.0,0.6	0.0,0.4,0.0,0.1	0.6,0.8,0.3,0.6	0.6,0.8,0.2,0.5	0.4,1.0,0.1,0.8	0.5,0.6,0.1,0.1	0.1,0.9,0.3,0.0	0.5,0.3,0.4,0.0	0.3,0.6,0.3,0.6		0.1,0.0,0.8,0.8
1.3-I	0.4,0.2,0.0,0.3	0.1,0.9,0.0,0.6	0.0,0.8,0.0,0.1	0.5,0.3,0.1,0.8	0.4,0.1,0.3,0.6	0.7,0.1,0.2,0.7	0.1,0.2,0.2,0.1	1.0,0.1,0.4,0.0	0.6,0.0,0.6,0.0	0.0,0.0,0.2,0.7	0.1,0.0,0.8,0.8	

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.

FINDING 9. *Gradual changes were perceived better for playback rate drops to 0.7 only.*

When looking at increases in playback rate, starting with 1.2, gradual changes offered little to no improvement, with average MOS of 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 remained 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, but 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 a statistically significant difference. 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.

Table 6. A summary of the four studies in terms of the key findings.

Study	Finding
1	A medium quality drop as well as playback rate drops of 0.8-0.9 had little to no impact on perceived quality.
1	A playback rate drop of 0.6 was perceived as poorly as a low drop in quality.
1	Playback rate drops of 0.6 and 0.7 were rated as poorly as rebuffering events of 2s and 0.5s respectively.
2	There is a great disparity between how playback rate increases are perceived across different content genres.
2	All variations of playback rate increases were perceived better than a drop to low quality for half of the clips.
2	Rebuffering of 2s was perceived to be worse than an increase of 1.8 in playback rate for half of the content.
3	There is a significant disparity between the perceptibility of seeking events across different content.
3	In the vast majority of cases, playback rate increases were perceived significantly better than seeking events.
4	Gradual changes were perceived better for playback rate drops to 0.7 only.
4	Gradual changes offered some QoE improvement for increases in playback rate to 1.3 only.

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 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 remaining two clips, Top Gear and Wolf Hall, the differences were not statistically significant.

FINDING 10. *Gradual changes offered some QoE improvement for increases in playback rate to 1.3 only.*

Overall, the gentler gradual changes of 0.05/s were perceived to be better than 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 most notable for the lowest and highest playback rates tested, namely 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.

8 Discussion

Our four subjective studies were designed to provide comprehensive evaluation of playback rate changes and their impact on QoE, exploring the RQs set out in Section 1. The studies revealed several important patterns in how users perceive playback rate adjustments in adaptive streaming, with a summary of findings shown in Table 6. Most notably, we found an asymmetric perception between rate increases and decreases, with increases being generally more tolerable although this is also more content-dependent. In addition, we discovered clear thresholds for acceptable rate changes and identified important interactions between playback rate adjustments and content characteristics. Table 7 shows a summary of the tested playback rates in each study, as well as the emerging acceptable ranges.

In this section we discuss the implications of the results, first on ABR algorithm design, informed by the first and last studies, and second on catch-up mechanism design, informed primarily by the second and third studies. In addition, we discuss the constraints of our research methodology and scope.

Table 7. A summary of the four studies in terms of tested playback rate.

Study	Tested Playback Rates	Acceptable Playback Rates
1	0.6, 0.7, 0.8, 0.9	0.8-1.0
2	1.2, 1.4, 1.6, 1.8	1.0-1.2
3	1.1, 1.2, 1.3, 1.4	1.0-1.3
4	0.7, 0.8, 1.2, 1.3	0.8-1.3

8.1 Implications for ABR Algorithm Design

In our first study, designed to answer RQ1, we investigated the QoE impact of playback rate drops in comparison with two other important QoE factors in adaptive video streaming: video quality and rebuffering.

We found that a quality drop to the medium level (where bitrate was 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 the bitrate 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 rebuffering duration of 0.5s was perceived similarly to a playback rate drop of 0.7; however, both of these conditions still had significant negative impacts on 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.

In the fourth study, we investigated gradual changes in playback rate and whether they can absorb some of the negative QoE impact caused by the target playback rate, aiming to answer RQ4. This varied greatly across content, target playback rates, and the 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 instant variations. However, there were exceptions, with Wolf Hall clips 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 a minority of cases, as this improvement is not universal across all content and target playback rates. Additionally, the magnitude of the gradual playback rate change is an important factor. In terms of ABR algorithm design, lower playback rates of 0.7 and 0.8 can be more effective in a rebuffering avoidance strategy; however, even if adjusted gradually, they can still have severe negative impact on QoE and therefore should be avoided.

Dynamic playback rate can be employed to supplement quality bitrate adaptation in adaptive video streaming, potentially improving the QoE of steaming services by reducing the amount of rebuffering experienced by users. In particular, in low latency live streaming services, where the client buffer is restricted by the latency, creating challenging streaming conditions in which playback rate adaptation alone might not be sufficient to counteract any changes in network conditions. However, this requires a careful redesign of ABR algorithms that are built into video

players for the following reasons. First, playback rates of 0.7 and below can be counterproductive, resulting in similar QoE degradation to that of rebuffering itself. Second, playback rates of 0.8 and 0.9 can be imperceptible, offering a valuable rebuffering avoidance strategy in scenarios where further quality bitrate degradation would negatively impact QoE. Third, dynamic playback rate and quality bitrate adaptation might require careful coordination to prevent the video player from overreacting to changes in network conditions that can reduce the overall QoE.

8.2 Implications for Catch-up Mechanism Design

In the second study, we focused on evaluating the QoE impact of temporary increases in playback rate, in attempt to answer RQ2. Similar to the first study, we compared the playback rate test conditions against the same variations of video quality drops and rebuffering events, allowing for a further comparison between decreases and increases in playback rate using the overlapping test conditions. Based on related work, which indicated that increases are less perceptible than decreases in the playback rate, we evaluated a higher range of playback rates, up to 1.8.

A playback rate increase of 1.2 was on average rated as high as a medium quality drop across all content, indicating its negligible impact on QoE. Playback rates of 1.4 and 1.6 were on average rated higher than both variations of rebuffering tested, 0.5 and 2s. Compared to playback rate drops in the first study, where a playback rate of 0.7 was rated as poor as a rebuffering event of 0.5s, these results confirm that playback rate increases are less perceptible than the decreases. Nonetheless, the average scores across the tested content varied to a greater degree in the case of playback rate increases when compared to the playback rate drops. This suggests that the perceptibility of the playback rate increase is much more content-dependent.

In our third study, we evaluated the QoE impact of playback rate increases against seeking events of various duration. This study aimed to answer our third research question. For this study, we tested a smaller range of playback rates of 1.1, 1.2, 1.3, and 1.4, and compared them with seeking events of various durations: 1, 2, 3, and 4s.

In most cases, playback rate increases were perceived significantly better than the seeking events, with the latter test conditions scored 1.4 lower on average across all content. These results clearly demonstrate the significance of the negative impact of seeking events on the overall QoE, especially since the seeking events tested were instantaneous and therefore the best case scenario. In addition, the smallest seeking event of 1s was perceived as the same as the highest tested playback rate increase of 1.4 for four out of six clips. Furthermore, a great variance in scores across different pieces of content can be seen in the results, as in the second study. The only exceptions were playback rates of 1.1-1.3 as they had negligible impact on QoE. On the other hand, a playback rate of 1.4 was perceived differently across content, with the DMOS ranging from 3 to 4.7 for Football and Vegesaur clips, respectively. For seeking events this was even more prominent, with considerable variance across content for each tested seeking duration. In addition, the best perceived content for each seeking duration was not the same. This suggests that, in rare cases, heavily dependent on the content, a seeking event might not be as noticeable; however, it can still negatively impact QoE. Moreover, the perceptibility of seeking might be dependent on various content characteristics, with the timing rather than the duration of seeking being more important. For example, it might depend on user engagement or the importance of the missing content, for example a goal being scored during a football match.

Taking into consideration the results of these two studies, playback rate alteration appears to be a better way of reducing client latency in live streaming scenarios than the use of seeking events, in terms of the overall QoE, with playback rate increases of up to 1.3 having little to no impact on the overall QoE. Seeking events, even of just 1s, seem significantly more perceptible across all content and, therefore, should be avoided in the design of catch-up mechanisms where

possible. However, this might not apply to scenarios where the latency needs to be reduced by a large margin, as seeking would be vastly more effective, whereas a playback rate increase would need to be maintained for a prolonged period of time to achieve the same latency reduction, which could potentially become more perceptible to the user.

Smooth latency management can improve low latency live streaming services by reducing the average latency experienced by users. This is due to the fact that in low latency adaptive streaming, any rebuffering will cause the latency to increase, requiring a catch-up mechanism to bring it back down to the target setting of the video player. Both playback rate increase and seeking events can disrupt the overall QoE; therefore, a sophisticated catch-up mechanism needs to be designed as part of the video player. Such a catch-up mechanism needs to strike a balance between the QoE impact and the effectiveness of actions taken to reduce latency during a streaming session. Seeking events disrupt QoE to a greater degree but offer instant latency reduction, whereas playback rate increases are less perceptible but take longer to achieve equivalent latency reduction. Furthermore, video players need to consider the QoE importance of latency itself, as its positive impact should outweigh the potential negative impact of actions taken to reduce it, in order to maximise the overall Quality of Experience.

8.3 Study Constraints

Content genres representation. While we used six diverse content clips (four in the final study), this sample size limits our ability to draw definitive conclusions about genre-specific effects. The disparities we observed in scores across different content, particularly for playback rate increases and seeking events, suggest important content dependencies that require more extensive sampling within each genre for reliable conclusions on how content genres affect the perceptibility of changes in playback rate.

Content characteristics. Similarly to the content genres, our limited content sample size does not allow us to make general conclusions around how various content characteristics impact the perceptibility of playback rate changes. Such characteristics include Spatial Information (SI) and Temporal Information (TI), as described in ITU P.910 [31]. A larger sample size is needed to investigate a potential correlation between SI/TI and the perceptibility of playback rate.

Scope of test conditions. The experimental design of each study incorporated test conditions aligned to answer a specific research question. While the test conditions encompassed a wide range of quality impairments, testing all possible permutations was not feasible due to substantial resource requirements associated with evaluating additional test sequences. In the first three studies, the investigation of playback rate variations was constrained to examining single-duration alterations. This enabled us to evaluate how short temporary changes in playback rate are perceived by users, but precluded extrapolation to extended periods of adjusted playback rate. The final study explored a subset of gradual changes in playback rate, facilitating preliminary comparative analysis between gradual and instantaneous changes, while acknowledging that there still remains unexplored parameter space in possible gradual rate variations.

Participant data collection. Our use of the ethical crowdsourcing platform Prolific enabled the recruitment of a large number of participants. However, this introduced potential obstacles in contrast to in-lab studies as we could not directly observe the participants taking the survey to ensure the reliability of their responses. Moreover, we could not ensure a consistent environment across participants, such as the display configuration, screen size, noise levels, or the lighting intensity. In order to mitigate this, we incorporated metrics collection within the survey to report details of the clips watched and any pauses, restarts, playback rate adjustments or seeking initiated by the user. Finally, in order to confirm attention to the content, at the end of the survey we presented

participants with a specific question about the content they had viewed. In order to compensate for potential differences in the environment, we recruited a higher number of participants.

Objective metrics. For all four studies we have analysed the subjective scores given by participants. It would be useful to calculate an objective metric for each test sequence to complement the subjective data. However, currently there is no objective metric that captures playback rate or seeking. Our findings show the significance of these factors to overall QoE, highlighting the need for these factors to be included in the development of future objective metrics.

9 Conclusion

This paper presents a comprehensive evaluation of playback rate and its impact on QoE in adaptive video streaming. We conducted four subjective studies, covering 210 test sequences and involving 204 participants, each addressing one of the research questions outlined in Section 1. Based on our findings, we offer practical design guidelines for video players, emphasising potential QoE improvements in two major areas of adaptive streaming: ABR algorithms and catch-up mechanisms. As our final contribution, we make our results dataset available to the research community [1].

Playback rate adaptation, in combination with bitrate adaptation informed by an ABR algorithm, can reduce the rebuffering and improve the QoE of streaming services. However, this requires a careful balancing act: slowing down video playback excessively can be as detrimental to QoE as rebuffering itself. Our studies identified that playback rates of 0.7 and lower were perceived as equally disruptive as rebuffering, even when applied gradually. In contrast, playback rate reductions down to 0.8 were imperceptible and therefore could be used to prevent rebuffering, especially in scenarios where bitrate reductions would be more detrimental to QoE.

Furthermore, playback rate adaptation can be used to maintain a desired latency in live streaming when implemented as part of a catch-up mechanism. Such mechanisms typically employ two methods to reduce latency during a streaming session: seeking ahead and/or speeding up video playback. However, our findings show that seeking is highly detrimental to QoE, whereas playback rate increases of up to 1.3 had little to no impact. This highlights how playback rate adaptations can enable smooth latency management leading to improvements in QoE of low latency streaming services.

In summary, playback rate adaptation offers a promising means of balancing quality, rebuffering, and latency to improve QoE in adaptive video streaming. Future work will explore content-aware approaches, examining how factors such as scene transitions and camera movements influence the perceptibility of playback rate changes.

Acknowledgments

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

References

- [1] 2025. Playback Rate QoE Dataset. <https://github.com/tomlyko/Playback-Rate-QoE-Dataset>
- [2] Omer F Aladag, Deniz Ugur, Mehmet N Akcay, and Ali C Begen. 2021. Content-aware playback speed control for low-latency live streaming of sports. In *Proceedings of the 12th ACM Multimedia Systems Conference*. 344–349.
- [3] Brahim Allan, Mike Nilsson, and Ian Kegel. 2019. A Subjective Comparison of Broadcast and Unicast Transmission Impairments. *SMPTE Motion Imaging Journal* 128, 6 (2019), 1–15.
- [4] Apple. 2023. HTTP Live Streaming (HLS) Authoring Specification for Apple Devices. <https://developer.apple.com/documentation/http-live-streaming/hls-authoring-specification-for-apple-devices>.
- [5] Christos G Bampis, Zhi Li, Ioannis Katsavounidis, Te-Yuan Huang, Chaitanya Ekanadham, and Alan C Bovik. 2018. Towards perceptually optimized end-to-end adaptive video streaming. *arXiv preprint arXiv:1808.03898* (2018).

- [6] Christos George Bampis, Zhi Li, Anush Krishna Moorthy, Ioannis Katsavounidis, Anne Aaron, and Alan Conrad Bovik. 2017. Study of Temporal Effects on Subjective Video Quality of Experience. *IEEE Transactions on Image Processing* 26, 11 (2017), 5217–5231. doi:10.1109/TIP.2017.2729891
- [7] N. Barman and M. G. Martini. 2019. QoE Modeling for HTTP Adaptive Video Streaming—A Survey and Open Challenges. *IEEE Access* 7 (2019), 30831–30859.
- [8] Abdelhak Bentaleb, Mehmet N. Akcay, May Lim, Ali C. Begen, and Roger Zimmermann. 2021. Catching the Moment with LoL+ in Twitch-Like Low-Latency Live Streaming Platforms. *IEEE Transactions on Multimedia* (2021), 1–1. doi:10.1109/TMM.2021.3079288
- [9] Abdelhak Bentaleb, Bayan Taani, Ali C. Begen, Christian Timmerer, and Roger Zimmermann. 2019. A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP. *IEEE Communications Surveys & Tutorials* 21, 1 (2019), 562–585. doi:10.1109/COMST.2018.2862938
- [10] DASH-IF. 2022. DASH-IF publishes Low-Latency DASH extensions. (2022). <https://dashif.org/news/low-latency-dash>
- [11] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang. 2017. A Quality-of-Experience Index for Streaming Video. *IEEE Journal of Selected Topics in Signal Processing* 11, 1 (2017), 154–166. doi:10.1109/JSTSP.2016.2608329
- [12] Nagabhushan Esvara, K. Manasa, Avinash Kommineni, Soumen Chakraborty, Hemanth P. Sethuram, Kiran Kuchi, Abhinav Kumar, and Sumohana S. Channappayya. 2018. A Continuous QoE Evaluation Framework for Video Streaming Over HTTP. *IEEE Transactions on Circuits and Systems for Video Technology* 28, 11 (2018), 3236–3250. doi:10.1109/TCSVT.2017.2742601
- [13] M. N. Garcia, W. Robitzka, and A. Raake. 2015. On the accuracy of short-term quality models for long-term quality prediction. In *Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. 1–6. doi:10.1109/QoMEX.2015.7148123
- [14] Deepti Ghadiyaram, Janice Pan, and Alan C. Bovik. 2019. A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos. *IEEE Transactions on Circuits and Systems for Video Technology* 29, 1 (2019), 183–197. doi:10.1109/TCSVT.2017.2768542
- [15] Craig Gutterman, Brayn Fridman, Trey Gilliland, Yusheng Hu, and Gil Zussman. 2020. Stallion: Video Adaptation Algorithm for Low-Latency Video Streaming. In *Proceedings of the 11th ACM Multimedia Systems Conference (Istanbul, Turkey) (MMSys '20)*. Association for Computing Machinery, New York, NY, USA, 327–332. doi:10.1145/3339825.3397044
- [16] IETF. 2017. HTTP Live Streaming (RFC 8216). (2017). <https://tools.ietf.org/html/rfc8216>
- [17] IETF. 2024. HTTP Live Streaming 2nd Edition. (2024). <https://datatracker.ietf.org/doc/html/draft-pantos-hls-rfc8216bis>
- [18] ISO/IEC. 2019. ISO/IEC 23009-1: Information technology-Dynamic adaptive streaming over HTTP (DASH)-Part 1: Media presentation description and segment formats. (2019).
- [19] ISO/IEC. 2020. ISO/IEC 23000-19: Information technology - Multimedia application format (MPEG-A) - Part 19: Common media application format (CMAF) for segmented media. (2020).
- [20] Theo Karagkioulos, Rufael Mekuria, Dirk Griffioen, and Arjen Wagenaar. 2020. Online Learning for Low-Latency Adaptive Streaming. In *Proceedings of the 11th ACM Multimedia Systems Conference (Istanbul, Turkey) (MMSys '20)*. Association for Computing Machinery, New York, NY, USA, 315–320. doi:10.1145/3339825.3397042
- [21] Yunlong Li, Xinfeng Zhang, Chen Cui, Shanshe Wang, and Siwei Ma. 2023. Fleet: Improving Quality of Experience for Low-Latency Live Video Streaming. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 9 (2023), 5242–5256. doi:10.1109/TCSVT.2023.3243901
- [22] Yu Liang, Tomasz Lyko, Mike Nilsson, Paul Farrow, Steve Appleby, Matthew Broadbent, and Nicholas Race. 2024. Playing Catch-Up: Evaluating Playback Speed Control in Low-Latency Live Streaming. In *16th International Conference on Quality of Multimedia Experience (QoMEX)*. 270–273. doi:10.1109/QoMEX61742.2024.10598290
- [23] Tomasz Lyko, Matthew Broadbent, Nicholas Race, Mike Nilsson, Paul Farrow, and Steve Appleby. 2020. Llama - Low Latency Adaptive Media Algorithm. In *IEEE International Symposium on Multimedia (ISM)*. 113–121. doi:10.1109/ISM.2020.00027
- [24] Tomasz Lyko, Matthew Broadbent, Nicholas Race, Mike Nilsson, Paul Farrow, and Steve Appleby. 2023. Improving quality of experience in adaptive low latency live streaming. *Multimedia Tools and Applications* (2023), 1–27.
- [25] Tomasz Lyko, Yehia Elkhatib, Rajiv Ramdhany, and Nicholas Race. 2024. Drop or Stop: Investigating the Impact of Playback Rate on QoE in Adaptive Video Streaming. In *16th International Conference on Quality of Multimedia Experience (QoMEX)*. 111–117. doi:10.1109/QoMEX61742.2024.10598252
- [26] Tomasz Lyko, Yehia Elkhatib, Michael Sparks, Nicholas Race, and Rajiv Ramdhany. 2022. QoE Assessment for Multi-Video Object Based Media. In *International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. doi:10.1109/QoMEX55416.2022.9900905
- [27] R. K. P. Mok, E. W. Chan, and R. K. C. Chang. 2011. Measuring the quality of experience of HTTP video streaming. In *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*. 485–492. doi:10.1109/INM.2011.5990550

- [28] Mu Mu, Lyndon Fawcett, Jamie Bird, Jamie Jellicoe, Steven Simpson, Hans Stokking, and Nicholas Race. 2017. Closing the Gap: Human Factors in Cross-Device Media Synchronization. *IEEE Journal of Selected Topics in Signal Processing* 11, 1 (2017), 180–195. doi:10.1109/JSTSP.2016.2638358
- [29] Pablo Pérez, Narciso García, and Álvaro Villegas. 2019. Subjective Assessment of Adaptive Media Payout for Video Streaming. In *International Conference on Quality of Multimedia Experience (QoMEX)*. doi:10.1109/QoMEX.2019.8743320
- [30] Benjamin Rainer and Christian Timmerer. 2014. A Quality of Experience Model for Adaptive Media Payout. In *International Workshop on Quality of Multimedia Experience (QoMEX)*. 177–182. doi:10.1109/QoMEX.2014.6982316
- [31] Recommendation ITU-T P. 910. 2023. Subjective video quality assessment methods for multimedia applications. *International Telecommunication Union* (2023).
- [32] Werner Robitza, Marie-Neige Garcia, and Alexander Raake. 2017. A modular HTTP adaptive streaming QoE model—Candidate for ITU-T P. 1203 (“P. NATS”). In *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.
- [33] Nicolas Staelens, Jonas De Meulenaere, Maxim Claeys, Glenn Van Wallendael, Wendy Van den Broeck, Jan De Cock, Rik Van de Walle, Piet Demeester, and Filip De Turck. 2014. Subjective Quality Assessment of Longer Duration Video Sequences Delivered Over HTTP Adaptive Streaming to Tablet Devices. *IEEE Transactions on Broadcasting* 60, 4 (2014), 707–714. doi:10.1109/TBC.2014.2359255
- [34] Liyang Sun, Tongyu Zong, Siqian Wang, Yong Liu, and Yao Wang. 2021. 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 (Istanbul, Turkey) (MMSys '21)*. Association for Computing Machinery, New York, NY, USA, 200–213. doi:10.1145/3458305.3463382
- [35] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. Garcia. 2016. Perceptual Quality of HTTP Adaptive Streaming Strategies: Cross-Experimental Analysis of Multi-Laboratory and Crowdsourced Subjective Studies. *IEEE Journal on Selected Areas in Communications* 34, 8 (2016), 2141–2153. doi:10.1109/JSAC.2016.2577361
- [36] H. T. T. Tran, N. P. Ngoc, A. T. Pham, and T. C. Thang. 2016. A Multi-Factor QoE Model for Adaptive Streaming over Mobile Networks. In *IEEE Globecom Workshops (GC Wkshps)*. 1–6. doi:10.1109/GLOCOMW.2016.7848818
- [37] H. T. T. Tran, T. Vu, N. P. Ngoc, and T. C. Thang. 2016. A novel quality model for HTTP adaptive streaming. In *IEEE Sixth International Conference on Communications and Electronics (ICCE)*. 423–428. doi:10.1109/CCE.2016.7562674
- [38] K. Yamagishi and T. Hayashi. 2017. Parametric Quality-Estimation Model for Adaptive-Bitrate-Streaming Services. *IEEE Transactions on Multimedia* 19, 7 (2017), 1545–1557. doi:10.1109/TMM.2017.2669859
- [39] D. Zegarra Rodríguez, R. Lopes Rosa, E. Costa Alfaia, J. Issy Abrahão, and G. Bressan. 2016. Video Quality Metric for Streaming Service Using DASH Standard. *IEEE Transactions on Broadcasting* 62, 3 (2016), 628–639. doi:10.1109/TBC.2016.2570012

Received 22 August 2025; revised 5 December 2025; accepted 9 January 2026