Multimedia Tools and Applications (2022) 81:9771–9802
https://doi.org/10.1007/s11042-022-12065-1

Quality of experience of 360 video – subjective and eye-tracking assessment of encoding and freezing distortions

Anouk van Kasteren[1,2] · Kjell Brunnström[1,3] · John Hedlund[1] · Chris Snijders[2]

Received: 9 October 2020 / Revised: 16 July 2021 / Accepted: 4 January 2022 /
Published online: 14 February 2022

Abstract

The research domain on the Quality of Experience (QoE) of 2D video streaming has been well established. However, a new video format is emerging and gaining popularity and availability: VR 360-degree video. The processing and transmission of 360-degree videos brings along new challenges such as large bandwidth requirements and the occurrence of different distortions. The viewing experience is also substantially different from 2D video, it offers more interactive freedom on the viewing angle but can also be more demanding and cause cybersickness. The first goal of this article is to complement earlier research by Tran, et al. (2017) [39] testing the effects of quality degradation, freezing, and content on the QoE of 360-videos. The second goal is to test the contribution of visual attention as an influence factor in the QoE assessment. Data was gathered through subjective tests where participants watched degraded versions of 360-videos through a Head-Mounted Display with integrated eye-tracking sensors. After each video they answered questions regarding their quality perception, experience, perceptual load, and cybersickness. Our results showed that the participants rated the overall QoE rather low, and the ratings decreased with added degradations and freezing events. Cyber sickness was found not to be an issue. The effects of the manipulations on visual attention were minimal. Attention was mainly directed by content, but also by surprising elements. The addition of eye-tracking metrics did not further explain individual differences in subjective ratings. Nevertheless, it was found that looking at moving objects increased the negative effect of freezing events and made participants less sensitive to quality distortions. More research is needed to conclude whether visual attention is an influence factor on the QoE in 360-video.

Keywords QoE · Quality of experience · 360-video · Visual attention · Quality perception · Eye-tracing · Perceptual load · Selective attention · Cybersickness

✉ Kjell Brunnström
kjell.brunnstrom@ri.se

Extended author information available on the last page of the article

1 Introduction

Multimedia streaming has been gaining popularity amongst consumers everywhere. The number of streaming services (e.g., Netflix, Amazon prime, and HBO) has been growing, and the available content even more so. Most of the material offered is 2D video streaming such as movies and TV shows. However, a new format, "VR 360-degree video" - omnidirectional content that can be viewed through a Head-Mounted Display (HMD) - is growing in popularity. This format offers a much more immersive viewing experience compared to 2D video. Popular streaming platforms such as YouTube, Vimeo, and Fox Sports are increasing their offer of 360-degree content. Additionally, HMDs are becoming more affordable and available for the public, allowing for a larger audience to access these omnidirectional videos. For acceptance and application of this media format in people's everyday life, an understanding of how to provide a pleasant viewing experience while efficiently utilizing network resources is needed.

In multimedia research the Quality of Experience is an important measure, which is defined by the International Telecommunication Union (ITU) as follows: "Quality of Experience (QoE) refers to the overall acceptability of an application or service, as perceived subjectively by the end-user." (ITU-T (2017) [16, 20]). There are many factors that can influence the QoE. An influencing factor (IF) is any characteristic of a user, system, application, or context whose actual state or setting influences the QoE, Reiter, et al. (2014) [28]. Important system IFs on the QoE of 2D video streaming are among others: viewing distance, display size, resolution, bitrates, and network performance, Kuipers, et al. (2010) [18]. Distortions and artifacts that occur during the different processing steps have a (negative) influence on the QoE, Möller, et al. (2014) [21], even more so when the distortions occur in salient regions of the video Engelke, et al. (2010) [5, 6].

As with 2D video, studying the QoE is important in the development and improvement of 360-video technology. The processing and transmission of 360-degree format brings along new challenges such as large bandwidth requirements and a more likely occurrence of different distortions. The viewing experience is substantially different from 2D video; it offers more interactive freedom with respect to the viewing angle but can also be more demanding and cause cybersickness. Simply applying the theory and methods of 2D video to the 360-video domain is not trivial and requires more specific research into how new challenges and the different viewing experience that come with 360-video relate to the QoE.

360-videos are recorded with multiple dioptric cameras, each capturing a different angle. The input from these cameras is then stitched together by a mosaicking algorithm. Artifacts may occur due to inconsistency between the cameras. For example, the illumination could be different in the different camera directions. As the material is stitched together, other issues could arise that cause artifacts and distortions such as blurring, visible seams, ghosting, broken edges, missing information, and geometrical distortions, Azevedo, et al. (2019) [1]. How these new artifacts affect the QoE and users' viewing behaviour has yet to be determined. Additionally, the transmission of 360-videos is a challenge because of the high bandwidth requirements. Today, 4 K resolution is accepted as a minimum functional resolution but requirements are increasing to 8 K, and even 16 K resolutions [1]. Therefore, the video material must be compressed to lower qualities which causes distortions such as blurring, blocking, ringing, and the staircase effect, which may negatively affect the experience [1]. The right balance is still to be found.

Watching VR 360-videos through an HMD offers an immersive experience. The higher level of immersion and presence could influence the QoE [38]. A realistic environment will have a positive effect on presence, Cummings, et al. (2016) [4]. Additionally, Salomoni, et al. (2017) [32] have found that more diegetic VR interfaces, scenes where all of the objects seen by a user belong to the virtual world, result in better user experiences. Delays, rebuffering, and quality fluctuations due to network limitations, could cause confusion and cybersickness by disrupting the course of movements which also negatively influences the viewing experience. The HMD is much closer to the eye compared to conventional 2D displays, which may cause distortions to stand out more and induce more eye strain and fatigue. Space between pixels may also be visible due to the closeness of the screen to the eyes. These effects could increase the perceptual and cognitive load (PCL) which can lead to stress Sweller, et al. (2011) [35].

The 360-degree material allows for a new way of interacting with media, with more freedom in deciding from what angle to watch the content. The total image is larger than the users' field of view, so different users may view different parts of the content and explore it in different ways. Visual attention is a likely IF in 360-video and its relation to the QoE should be studied [33, 38] [25, 34, 40]. This article focusses on the validation of previously researched Ifs mainly for 2D-video but very little for 360-video, such as compression, freezing events, and contents. In addition, it contributes by evaluating whether eye-tracking and visual attention are valuable additions to 360-video assessments. The combination of factors has not been investigated before for 360-video to our knowledge.

1.1 Related work

Thus far not many studies on the topic have been conducted. Some studies have been performed adapting methodologies from 2D video quality assessments. An elaborate study by Tran, et al. (2017) [38] tested the effects of several IFs such as resolution, bitrate, content, camera motion, and viewing device on the perceptual quality, presence, cybersickness, and acceptability of 360-video. They found that for videos with QP values of 40 or higher the acceptance level drops below 30%, but that QP values from 22 to 28 did not significantly differ in acceptance. Additionally, the acceptable bitrate was found to be content-dependent as more motion required higher bitrates, indicating that motion activity should be further investigated. Recently, Gutierrez, et al. (2021) [9] published a cross lab investigation based on the work of the Video Quality Experts Group (VQEG), that has greatly influenced the recent Recommendation from the ITU [17]. The study in [9] was conducted time wise in parallel with this study and studied audio visual quality, simulator sickness symptoms, and exploration behaviour in short 360-video sequences. It was targeted to study methodological questions on how to conduct subjective experiment with 360-video. Unlike the current study it did not consider freezing and did not involve eye-tracking. Another study looked at the effects of stalling or freezing events on the QoE and annoyance [33]. Their results show that even a single stalling event leads to a significant increase in annoyance and should thus be avoided. However, different freezing frequency patterns did not necessarily result in different annoyance scores. Freezing studied for 2D video has shown that in addition to the length the number of occurrences is also important. Previous studies have found that adaptation of the 2D video methods to 360-video is not trivial. This article will complement existing results by gathering additional kinds of subjective data and by including eye-tracking data to evaluate visual attention as an influence factor to the QoE in 360-videos.

For 2D videos it has been found that global distortions, do not significantly alter our gaze patterns, Ninassi, et al. (2007) [23]. Local distortions, have been shown to draw significant attention, and alter viewing behaviour, Engelke, et al. (2017) [7]. This has in turn been shown to have an impact on the overall QoE [7]. It has been found that an overload of PCL lead to selective attention [19, 24]. Furthermore, in situations with a higher cognitive load, the average fixation duration increases [12, 43]. Whether visual attention could improve objective 360-video quality metrics is still unknown. This article will provide first explorations into the relation between visual attention and QoE.

1.2 Research questions and hypotheses

The first of research questions for this study is:

1. Which factors influence the QoE in 360-degree video and how?

a. How do video quality degradations, freezing events, and content relate to the QoE in 360-degree video?
b. What is the threshold for an acceptable QoE (MOS $>$ 3.5), given the effects of different influence factors?

The following are hypothesized about the first research question:

> **H1a:** Degrading the video quality will have a stronger than linear negative effect on the perceived quality and overall experience [38]; additionally, it will lead to an increase the PCL.
> **H1b:** Adding freezing events and increasing the frequency of these events will have a negative effect on the perceived quality and the overall experience; additionally, it is expected to increase the perceived PCL and cybersickness.
> **H1c:** The content of the video will moderate the effect of video quality. It is expected that the effect of degrading the video quality on the QoE will have a weaker effect on video content with higher motion activity.

The second research question is:

2. How can eye tracking be used to gain further insights into factors influencing the QoE of 360-videos?

a. How do video quality degradations, freezing events, and content influence the viewer's eye movements?
b. How are eye movements and the user's focus related to the QoE?

The following are hypothesized about the second research question:

> **H2a:** Degrading the video quality, adding freezing events, and increasing the frequency increases the average fixation duration, decreases fixation count, and attention will be more selective.
> **H2b:** Where a person looks is expected to be affected by the content of the video. Additionally, content with high motion activity is expected to result in a higher average fixation duration and lower fixation count.

H2c: The proportion of the 360-degree field that is looked at is related to the QoE.

H2d: Due to a masking effect the extent to which a viewer looks at moving objects has a moderating effect on the effect o of degrading video quality on the QoE. It is expected that when participants look more at moving objects the effect of quality degradation is weaker.

H2e: The extent to which a viewer looks at moving objects moderates the effect of freezing events. It is expected that the effect of freezing events on the QoE will be stronger if a viewer looks more at moving objects, as the course of movements in the viewers' attention are disrupted more.

H2f: Quality degradations and freezing events may increase the PCL, which in turn negatively affects the QoE. Thus, eye tracking metrics on selective attention and the average fixation duration are expected to mediate the effect of video quality and freezing.

The different hypotheses and their interrelations are illustrated in Fig. 1.

2 Method

In QoE assessment both objective and subjective measurements play a role, as it involves the objective quality of the video/audio content, the human perception, and the Quality of Service (QoS) [18]. Several standardized methodologies for subjective testing have been defined by the [13, 15, 17]. In this article, the single stimulus method will be used. In this method, each video is shown independently to an observer who is then asked to score it, Nidhi, et al. (2014) [22]. Results are often expressed through a Mean Opinion Score (MOS) on a 5-point scale, where higher values correspond to higher quality, calculated as the average of the ratings given by the subjects. Because people have a tendency to avoid perfect ratings, scores from 4.3 and

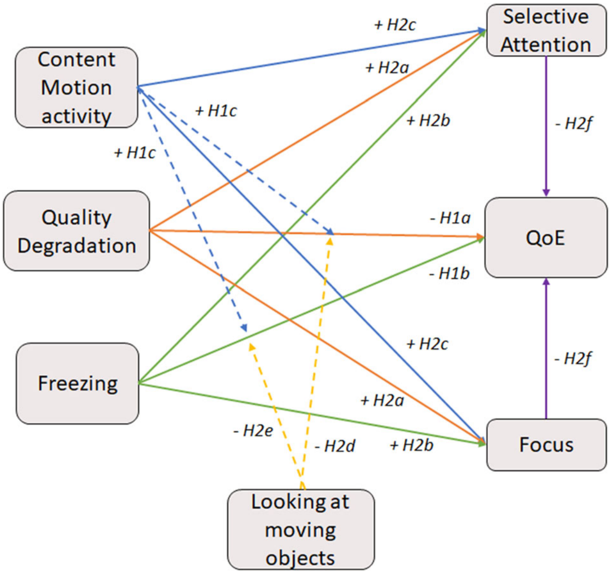

Fig. 1 Illustration of the different hypotheses. Solid lines represent direct effects and dashed lines represent interaction effects. Plus, and minus symbols indicate a positive or negative effect

up are generally considered as excellent quality, Winkler (2005) [42]. Additionally, a MOS of 3.5 is considered acceptable quality by most users and is often used as standard minimal threshold.

2.1 Design

The main dependent variable of the study is the QoE of 360-videos, which is measured through four subjective post-hoc evaluations. After viewing a video, participants were asked to rate their perceived video quality, cybersickness, overall experience, and perceptual load. While viewing the videos the eye-movements of the participants were tracked. To test the hypotheses, the manipulated independent variables are:

- video content: three different videos (Dojo, Flamenco, and intersection).
- video quality: degradation of the videos to 4 levels (CRF 20, 28 and 36).
- freezing events in no, low and high frequency.

This results in a 4 (video quality) × 3 (freezing events) × 3 (content) design with multiple dependent variables. Participants view all the versions of the videos in a single stimulus way [15]. The eye-tracking resulted in several variables to be considered in the analysis which will be explained in the measurement section. Demographic variables collected are age, gender, education/profession, visual acuity, possible colour vision deficiency, and whether participants wear glasses or contacts. Additionally, questions were asked regarding participants' VR experience and motion sickness.

2.2 Participants

Based on Brunnström, et al. (2018) [3, 31] on sample size estimation, the number of minimum required participants was calculated. We set the required power to 90%, expected standard deviation to 0.7, desired MOS difference to be resolved to 0.6 and from that arrived at 33 participants, which were recruited at the RISE Kista office in Sweden or via known affiliates and signed up on a voluntary basis. Participants were required to have good vision (through aids that do not interfere with the HMD), have no other severe eye or vision impairments, and do not have progressive glasses or contact lenses, which was screened before the experiment. Due to missing data, one participant was excluded from the dataset. 22 participants were male and 10 were female, aged 21 to 44 (Mean = 28.8, SD = 6.3), 9 of them used glasses and 3 contact lenses and one suffered from minor colour vision deficiency. Visual acuity was measured before the experiment and results were between 0.65 and 1.5 (Mean = 1.1, SD = 0.2).

2.3 Setup and stimulus materials

The experiment took place at the RISE office in Kista, Sweden, in one of the labs designated for video quality experiments. Hardware used in the experiments was an ASUS ROG ZEPHYRUS (GX501) laptop running on Windows 10 and an HTC VIVE VR headset with integrated Tobii eye trackers. For this headset, the maximum recommended resolution of 360-video material is 4096 × 2048 @ 30 fps.

The software used to conduct the experiments was Tobii pro lab v1.111 [36]. In this software, video or image material can be uploaded and prepared in timelines. Seven separate timelines were created for the current study. One timeline for the practice trial, and for each video a reference and degradations timeline. In addition to the videos, images with the questions were added to the timeline after each video.

In the Tobii software, one can define areas of interest (AOIs), an example can be seen in Fig. 2, on which metrics can be calculated such as the fixation duration and count. AOI's allow to divide the stimulus space and gather more detailed data for areas of special interest Holmqvist, et al. (2017) [11]. The collected dataset was downloaded with the Tobii I-VT fixation pre-set [37]. AOIs were drawn on the total video stimulus and on moving objects, buffer spinners, and bottom logos (if applicable) as these were found to attract attention. Besides exporting the collected metrics data, a raw data export was used for analysis. During the experiments, the software shows the eye-movements live during the experiment as can be seen in Fig. 3 a recording can also be viewed after completion of the experiment. Lastly, the program can create heatmap or scan path visualizations.

2.3.1 Overview of the video stimuli

Three different videos were selected that had a very high-quality version or an uncompressed original, see Table 1 and Fig. 4 . The motion vectors were calculated as well as the spatial and temporal index (SI and TI), as shown in Fig. 5 and Table 2 [14]. The motion vector example script of FFmpeg [8] was used, and for the SI and TI the program "siti" [29] was used. These outcomes were used to classify the videos based on their motion activity. Looking at the SI, TI and motion vectors it can be seen that there is less movement in the intersection video. The Dojo and Flamenco videos are more similar, however, the movements in the dojo video are more intense, hence, there are larger spikes in the motion vector graphs (Fig. 5). The spatial index is slightly higher for the flamenco video as there are more structure (people in this case). Another distinction between videos is the lighting. The intersection video is recorded outside with natural lighting, the flamenco video is inside but with a reasonable number of windows and the Dojo video is inside in a space with mainly artificial lighting. Looking at diegetics of the interface, the videos represent a realistic environment that can be viewed freely in all directions. However, the user is not able to interact with the environment in any other way (e.g., moving around or interacting with objects).

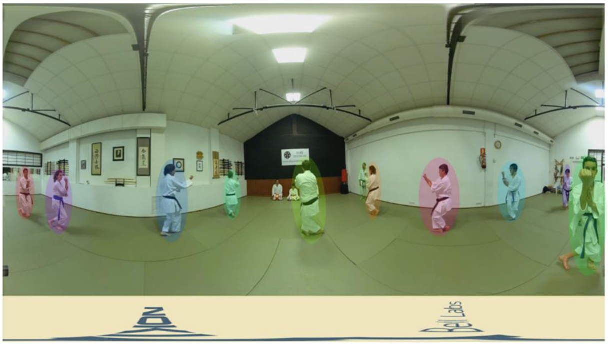

Fig. 2 Screenshot from the Tobii pro lab, the picture shows AOIs drawn on a still image

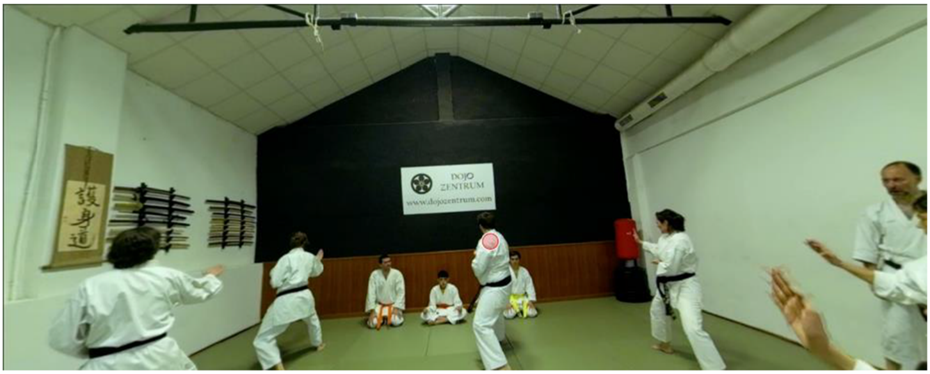

Fig. 3 Example of the view while observing the participants. The participant's fixation-point is indicated by the red circle

These videos were degraded to different quality levels using FFmpeg, the script can be found in [39]. The videos were cut, encoded with H.264 with resolution of 3840 × 2160 format and degraded to lower qualities by changing the Constant Rate Factor (CRF). CRF is a quality setting with values ranging between 0 and 51. Higher values result in more compression, while trying to keep the video at a constant quality at each level. This implies that the bitrate will vary depending on the content. To add the freezing events, the program "bufferer" [30] was used. The bufferer program will freeze the video and add a spinner at desired moments for a set time (script in [39]). In total, this resulted in 36 videos to be used in the experiment.

2.4 Measurements

Variables used in the study can be divided into four categories: subjective measurements (the dependent variables), manipulation variables (independent variables), eye-tracking data (both dependent and independent variables) and control variables.

2.4.1 Subjective measures

Four questions were shown after each video through a non-diegetic interface. These questions were accompanied with a 5-point scale and participants verbally answered by stating the number corresponding to their answer. These four questions were:

Table 1 Properties of the original videos used in the experiment

Name	Original resolution	Original encoder	fps	Bitrate (Mbit/s)	Size (GB)	Duration (min)	Source	Setting	AOI	Bottom logo
Dojo	3840× 2160	H.264	30	40	1.2	4:19	Nantes/ Nokia	inside	People	Yes
Flamenco	3840× 2160	H.264	30	40	0.92	3:10	Nantes/ Nokia	inside	People	Yes
Inter-section	7680× 3840	H.265	30	157	5.5	5	Inter-digital	outside	Cars	No

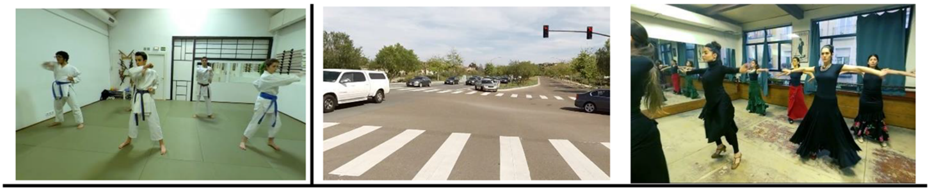

Fig. 4 Still images of the video stimuli. From left to right: Dojo, Intersection, Flamenco

- “Did you experience nausea or dizziness?” 1 (not at all) – 5 (a lot).
- “How would you rate the video quality?” 1 (very low) – 5 (very high).
- “How much cognitive and perceptual load did you experience?” 1 (none) – 5 (a lot).
- “How would you rate your overall experience?” 1 (unpleasant) – 5 (pleasant).

These four questions resulted in the four variables in the dataset perceived Video Quality (VQ), Overall Experience, Cybersickness, and Perceptual and Cognitive Load (PCL).

2.4.2 Manipulations

The manipulations of videos served as independent variables. Videos were generated with CRF values of 15 (visible lossless), 20, 28, and 36. CRF was added as a categorical variable. The second manipulation variable was the freezing event frequency. Freezing events of three seconds were added in different frequencies to the videos. Videos either had no freezing (none), two freezing events (low), or four freezing events (high). The third manipulation was the content of the video. There were three different videos: Dojo, Intersection, and Flamenco, see Section 2.3.1.

2.4.3 Eye-tracking data

Eye-tracking data can be described through quantitative metrics (based on fixations, saccades, or smooth pursuit eye movements) or through visualizations such as heatmaps and scan paths. The quantitative metrics used in the current study are as follows:

- **Total Fixation duration:** The total cumulative duration of all fixations in a certain area. It can provide insights when applied to AOI’s into how long a person looked at a certain area within a certain timespan.
- **Average Fixation duration** (one fixation): Longer fixations can be an indication of a person having more trouble finding or processing information.

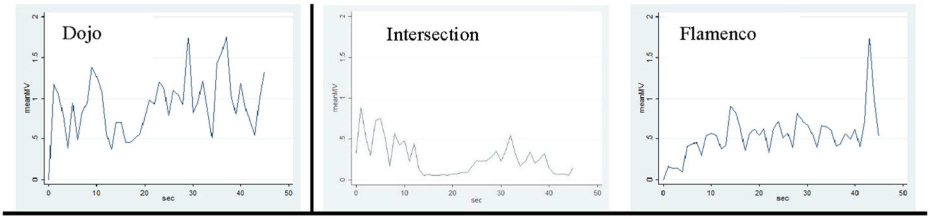

Fig. 5 Motion vector graphs over time for the three videos. Left to right: Dojo, Intersection, Flamenco. Where Dojo has the most motion and Intersection the least

Table 2 the mean Spatial (SI) and Temporal (TI) index of the three videos

video	mean SI	mean TI
Dojo	80.4	90.2
Flamenco	83.5	89.4
Intersection	76.8	80.1

- **Fixation count:** The number of fixations in a certain area. A higher fixation count could indicate lower search efficiency. More AOI hits could indicate the more importance of that area.
- **Fixation Rate:** Refers to the number of fixations per second and is closely related to the fixation duration. A higher fixation rate could indicate less focus, and more difficulty searching or processing a stimulus.

2.4.4 Control variables

Most of the control variables were collected through participants filling out a short survey prior to the experiment [39].

2.4.5 Qualitative measures

There were also measures of a more qualitative nature. During the experiment, the researcher wrote down based on live view what the participant was looking at. Comments included what people looked at, how they moved their head and eyes and how they explored the 360-environment. Attention maps were generated by the Tobii software, visualized as heatmaps, and analysed. They represent the spatial distribution of the eye-movement data. It can provide an intuitive overview of the overall focus areas.

2.5 Procedure

Participants were received and welcomed at the lab after which a short introduction of the experiment was given. Participants were sent instructions beforehand; these were discussed to make sure the participants have understood the procedure. A consent form was signed after which the participants' visual acuity were measured (Snellen) and colour vision were checked (Ishihara). A survey was filled out on demographics: age (years), gender(male/female/other), education or profession, and vision (whether they wore glasses or contact, if they had any other severe vision disability, their visual acuity and possible colour vision deficiency). The participants were seated and instructed on how to adjust the headset. First, a training sequence was done for the participant to get acquainted with the VR environment, the question format, to adjust the lenses and calibrate the eye tracking. The experiment consisted of a total of three video sets with each 11 degraded videos and one reference video. The order of the video sets was determined by rolling a dice. The order of the videos within a video set was randomized within the software, except for the reference, which was shown first. After each video, the four rating questions was shown one by one inside the headset and the participant was asked to answer verbally on a 1–5 scale. After each video set, the participant could take off the headset for a short break. This was repeated for all three video sets. The participants were compensated with a movie ticket. The experiments followed the script in [39].

2.6 Analysis

Analysis of the results consisted of two parts, a qualitative analysis of observations and a statistical data analysis.

2.6.1 Qualitative analysis

Observations of the visual behaviour were analysed and summarized as well as the comments made by participants during or after the experiment. Additionally, attention maps in the form of heat maps were generated of each video to visualize the attentional behaviour and compliment the observations. To Generate the heatmaps the default fixation gaze filter of the Tobii program was used, and the visualization is based on the absolute fixation count.

2.6.2 Quantitative analysis

The dataset contained 1148 observations gathered from 32 participants with 36 measurements each. For two participants two measurements were missing due to technical issues, but only one was excluded.

Data preparation The final dataset on which the data analysis was performed was a combination of hand-written subjective answers, the metric-based data, and the raw data export from the Tobii software. Data preparation was done both using Python 3.7 and STATA/IC 14. The metric-based export contains data based on the marked AOI's. Each line in the data represents a video. The raw data was millisecond interval-based and contained a large bulk of data. Interesting variables from this dataset were selected and grouped on video level such that one line corresponds to one video again. These two sets were merged, and the subjective ratings were added by hand. To give a more global representation of the eye-tracking data, four summarizing variables were created by principal component factor analysis Table 3. To capture the proportion of the total viewing area that participants look at AreaX and AreaY were created as a measure of selective attention, RvsI was created and as a measure of focus-related fixation data, PCL-fix was created (See Table 3 for how these variables were formed). Additionally, the ratings of perceived video quality and overall experience were quite similar, therefore, to create a better representation of the quality of experience, a scale variable, QoE (alpha = 0.84), was created as well.

Variable description As mentioned before, there are four categories of variables: manipulations, subjective measures, eye-tracking data, and other descriptive variables, see Table 3.

Analysis The analysis was done in three parts:

1. The effects of the manipulations on the subjective evaluation.
2. The effects of the manipulations on the eye-tracking data.
3. The relation between the eye-tracking data and the subjective evaluations.

As there are multiple measurements per person (36 videos nested within persons), the measurements per video are not independent and therefore multi-level regression was used to analyse the data. Multi-level regression (also called hierarchical linear regression, mixed

Table 3 Taxonomy of the variables used in the quantitative data analysis

Name	Category	Description	Values
CRF	Manipulation/ Independent	Quality Parameter. Larger means worse quality	15 (Reference), 20 (Good), 28 (Medium), 36 (Bad)
Freeze	Manipulation/ Independent	Freezing events	None, low, high frequency
Content	Manipulation/ Independent	The video content	Dojo, Flamenco, Intersection
VQ	Subjective measure/ Dependent	The perceived video quality	Scale: 1 (very low) - 5 (Very high)
Experience	Subjective measure/ Dependent	The perceived overall experience	Scale: 1 (Very unpleasant)– 5 (Very Pleasant)
QoE (Alpha = 0.84)	Subjective scale variable/ Dependent	The Quality of Experience as scale variable of VQ and experience	Scale: 1–5
PCL	Subjective measure/ Dependent	The perceived perceptual and cognitive load	Scale: 1 (none) – 5 (a lot)
Cybersickness	Subjective measure/ Dependent	The perceived dizziness and or nausea	Scale: 1 (Not at all) – 5 (A lot)
AreaY (Alpha = 0.90)	Eye tracking/ Dependent and independent	The area looked at in the x dimension as scale variable of: Pixels looked at in the x dimension, the standard deviation of pixel looked at in the x dimension and the distance between fixation points.	Standardized values between −1.69 and 4.14
AreaX (Alpha = 0.72)	Eye tracking/ Dependent and independent	The area looked at in the y dimension as scale variable of: Pixels looked at in the x dimension, the standard deviation of pixels looked at in the y dimension.	Standardized values between −1.48 and 5.41
RvsI (Alpha = 0.94)	Eye tracking/ Dependent and independent	How much a participant looks at relevant vs. irrelevant areas as scale of: total fixation duration and the fixation count on both relevant and irrelevant areas.	Standardized values between −2.60 and 1.51
PCLfix (Alpha = 0.79)	Eye tracking/ Dependent and independent	Fixation data related to perceptual and cognitive load as scale of: average fixation duration, total fixation count and total saccade count	Standardized values between −1.81 and 5.66

modelling, or nested data modelling) takes the interdependency of evaluations within persons into account and is a standard method in for instance psychology or sociology, although it is less common in the QoE community, Raudenbush, et al. (2001) [27]. In the first step (H1a, H1b, and H1c) the subjective evaluations of perceived quality, overall experience, perceptual and cognitive load, and cybersickness as the dependent variables and the manipulations CRF, Freezing, Content and their interactions were added as independent variables. The video order was added as an additional covariate. The final models were selected through backward elimination based on AIC, BIC and likelihood ratio tests.

For the second part (H2a, and H2b), the eye-tracking data was first summarized as four scale variables (see Table 4 for description) using principal component factoring and Cronbach's alpha. These four variables are used as dependent variables, and the manipulations and their interactions as independent variables. Additionally, the video order was added as a covariate. The final models are selected through backward elimination based on AIC, BIC and likelihood ratio tests.

In the final step of the analysis (H2c, H2d, H2e, and H2f), the four eye-tracking scale variables and their interactions with the manipulations are added to the multi-level regressions of the first part to test if they have a relation to the subjective evaluations. The final models were selected through backward elimination based on AIC, BIC and likelihood ratio tests. Additionally, significant eye tracking variables were tested for mediating effects.

3 Results

Results of the experiment were analysed both in a qualitative as well as a quantitative manner. Observation notes and gaze plots served as qualitative material; the eye-tracking data combined with the subjective questions data were used for the quantitative statistical analysis.

Table 4 Regression coefficients of the multi-level regressions on the subjective metrics

Predictors	QoE	Experience	PCL	Cybersick
CRF (15)				
20	−0.156	0.024	−0.039	0.058
28	−0.282**	−0.088	−0.025	0.036
36	−0.891***	−0.771***	0.125*	0.117***
Freeze	−0.398***	−0.398***	0.124***	0.046***
Content (Intersection)				
Dojo	−0.086	0.018	0.147***	0.039
Flamenco	0.062	0.181	0.014	−0.032
Video Order	−0.004*	−0.003*	0.004*	0.003**
PCLfix		−0.028		
RvsI		−0.018		
CRF*Freeze (15)				
20	0.109	0.125*		
28	0.134*	0.149*		
36	0.302***	0.309***		
Content*CRF (Intersection | 15)				
Dojo | 20	−0.101	−0.326*		
Dojo | 28	−0.147	−0.403**		
Dojo | 36	−0.419***	−0.591***		
Flamenco | 20	−0.002	−0.311		
Flamenco | 28	−0.123	−0.458**		
Flamenco | 36	−0.581***	−0.788***		
CRF*RvsI (15)				
20		0.181*		
28		0.201*		
36		0.127		
Freeze*RvsI		−0.054*		
CRF*PCLfix (15)				
20		−0.101		
28		−0.042		
36		0.052		
_cons	3.860***	3.759***	1.703***	1.067***

Note: *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

3.1 Qualitative analysis

3.1.1 Observations

While observing participants viewing of the 360-videos it became clear in which way participant behaviour tended to differ. It was a clear difference in answers between participants. Some rated all the videos on average high, others on average low. Some had a large variation in their answers to different conditions whereas others hardly noticed any difference. There were large variations among participants in the way they moved their head. However, after watching the same video a few times they started moving their head less and focus longer on one point before moving on. How fast this effect occurred would differ between participants. To a certain degree all participants' attention was at some point drawn to faces, signs, other text, hands, feet, and other moving objects. "Leader" persons such as the teachers in the video also drew more attention. Some participants would only look at these "relevant" stimuli and ignore the surroundings partly or completely. In contrast, other participants would mainly focus on surroundings and would only briefly pay attention to the aspects mentioned above. A similar difference was observed on whether participants looked at the spinner during freezing events or not. Some would immediately be drawn to look at it even when they were facing the opposite direction and others did not even look at the spinner even if it appeared right next to their current focus point. Furthermore, distortions and artifacts did also draw attention to different degrees. Even though asked to ignore it in their judgment as it was not possible to control for, participants were still distracted by for example stitching artifacts or focused on them for a while. In conclusion, there were substantial individual differences in the viewing behaviour among participants.

Apart from these observed individual differences, there are some other comments and observations worth mentioning. First, quite a few comments on the buffering and freezing were made. One participant stated that without audio, the buffering had a weaker effect as it seems less of an interruption. Another participant mentioned that when buffering occurred while looking at moving objects it was perceived as more annoying. Furthermore, it was observed that after more videos were played, some participants that initially were drawn to the spinner seemed to have gotten less sensitive to it. Some comments regarding realism of the interface included the fact that the camera appears to be higher than the participant's natural height which made them feel uncomfortable; and not being able to see details or read text was also mentioned as an annoying feature. These non-diegetic elements could negatively influence the viewing experience Salomoni*,* et al. (2017) [32]. One participant said he got tired after a few videos and that moving his head would make him dizzier. Finally, some participants indicated that they were too busy comparing videos and that they were influenced by the previous video when rating the current one. It was, for example, observed that after seeing one of the worst quality videos, participants would rate the videos of CRF = 20 higher than the reference video.

3.1.2 Attention map analysis

The fixation heat maps show a summary of where participants fixated on in a particular video, see Figs. 6, 7 and 8 as well as [39]. The horizontal dimension was more explored compared to the vertical dimension. The participants attended to areas that help them orient. In the Dojo and Flamenco videos this was by looking at people's faces or limbs. In the intersection video

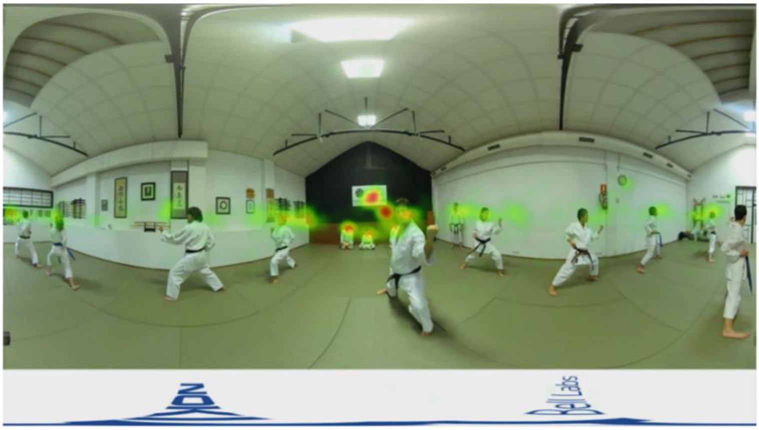

Fig. 6 Attention map pasted on a video still from the Dojo video. Focus on the middle of the video and on faces can be observed

people viewed the streets, at traffic lights and signs. Freezing events drew more attention towards the centre. In the intersection video the horizontal spread were more focused towards the centre areas.

3.2 Quantitative analysis

As described above, the statistical analysis consisted of three parts which will be covered in the following section.

3.2.1 Effects of the manipulation on the subjective ratings

The effects of the manipulation on the subjective ratings were analysed based on the hypotheses H1a, H1b, and H1c. QoE was measured through the perceived quality and the overall experience and expressed in a MOS. Values range from 1 to 5 (mean = 3.10; SD = 0.97). Figure 9 shows the QoE MOS as an interaction plot of the different manipulations. The highest

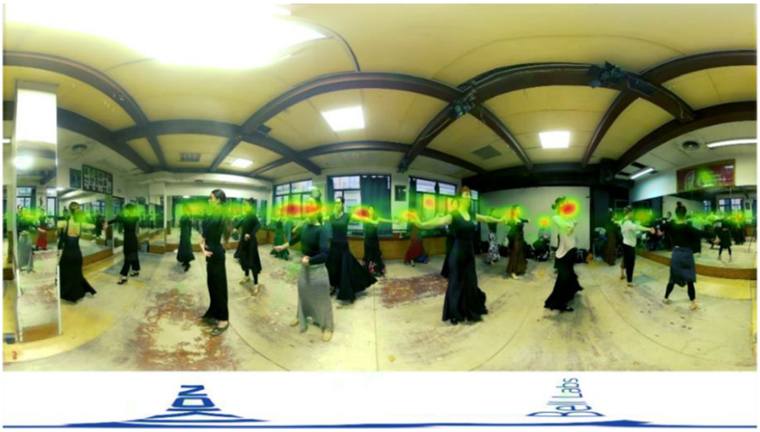

Fig. 7 Attention map pasted on a video still from the Flamenco video. A focus on faces can be observed

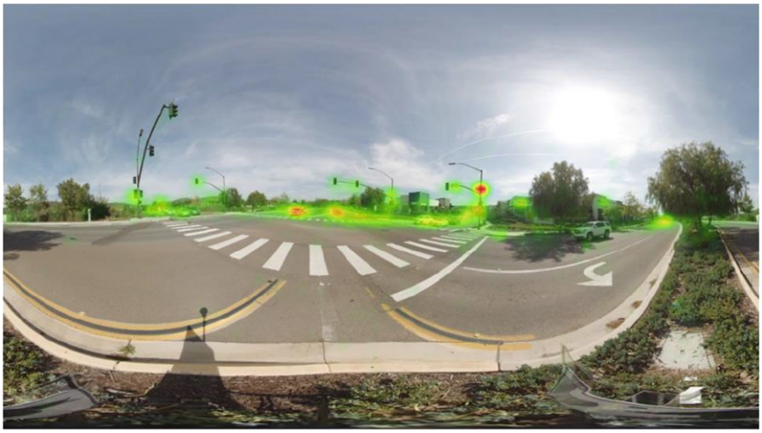

Fig. 8 Attention map pasted on a video still from the Intersection video. A focus on the end of the road, traffic lights and signs can be seen

MOS was received by the flamenco reference video without freezing (MOS = 3.92). For the none freezing condition all except videos with CRF = 28 or higher had a score above the acceptable level of 3.5. Any of the freezing conditions fall below 3.5. The lowest MOS was received by the flamenco video with CRF = 36 and high freezing frequency.

Looking at the individual responses (Fig. 10), there were differences in baseline between participants. Some rated everything high as for example participant 31, while others rated everything low as for example participant 1. Therefore, to test the significance of these results a multilevel regression was performed, with video condition as the first level and participant as the second. Estimating an empty multilevel model showed that 45% of the variance was on the participant level, confirming the need for multi-level analysis.

The resulting model on QoE after backward elimination can be seen in Table 4 column "QoE part 1" (within-R2 = 0.42; between-R2 = 0.02; rho = 0.57). Included in the model were CRF, Freeze, Content, and the interactions between CRF and both freeze and content. The results showed a negative effect of CRF in the none freezing condition where higher CRF values resulted in lower QoE. This effect increased as the CRF value increased showing a larger than linear effect. As can be seen in Fig. 11, the effect of CRF in the low and high freezing frequency conditions was found to be smaller compared to the none freezing condition. The effect of freezing was larger for lower CRF values compared to higher ones.

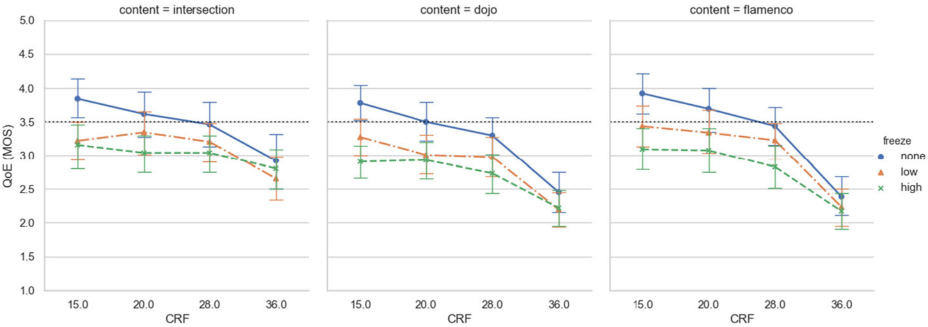

Fig. 9 MOS Quality of Experience (y-axis) interaction plot for the different manipulation conditions. CRF on the X-axis, and a different line per freezing frequency. Reference line of the acceptable level of 3.5 included

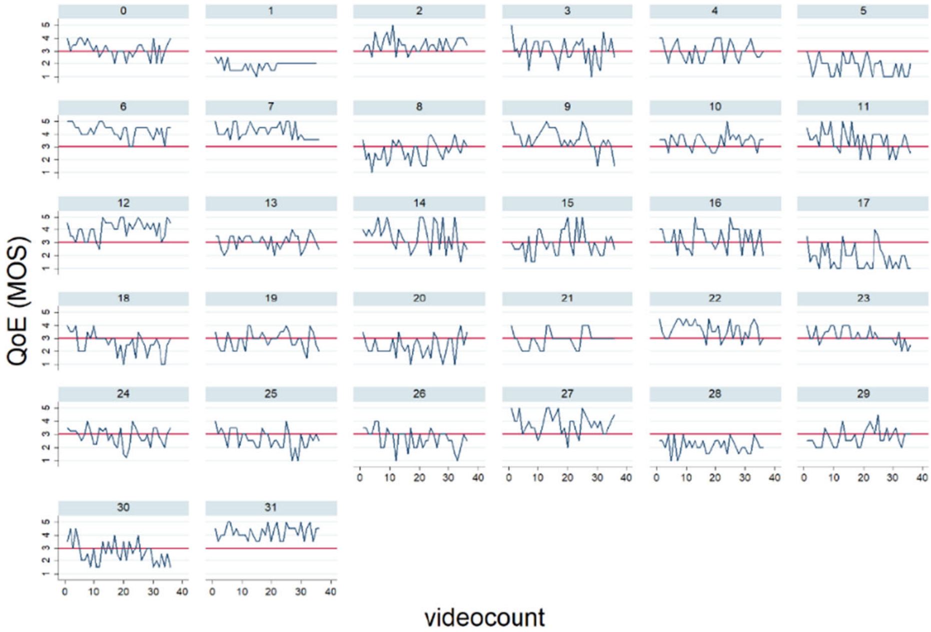

Fig. 10 Quality of Experience rating (y-axis) per participant for consecutive randomized videos

Furthermore, it was found that the effect of CRF 36 on the QoE was stronger in the Dojo and Flamenco video compared to the Intersection video. Additionally, a small but significant effect of the video order was found.

Secondly, the effects of the manipulations on the PCL were evaluated. Answers were expressed in a MOS, and the answers ranged from 1 to 4 (mean = 1.97; SD = 0.91). Figure 11 shows the interaction plot of the effects in the different conditions. PCL scores were rather low in all conditions, never surpassing MOS = 2.5. The resulting model of the multi-level regression on PLC after backward elimination can be seen in Table 4 (within-R^2 = 0.06; between-R^2 = 0.02; rho = 0.58). Included in the model are CRF, Freeze, Content and the video order. Here the fit of the model is not good enough to draw any conclusions from it.

Third, the effect of the manipulations on cybersickness was tested. Answers are again expressed through a MOS, ranging between 1 and 5 (mean = 1.25; SD = 0.52). The MOS

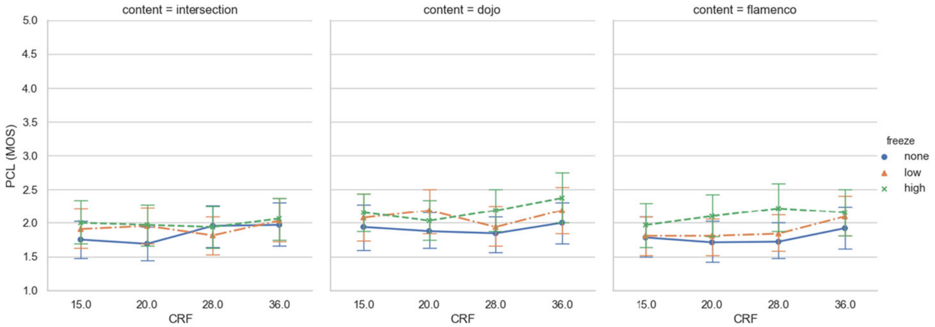

Fig. 11 MOS Perceptual and Cognitive load (y-axis) interaction plot for the different manipulation conditions. CRF on the X-axis, and a different line per freezing frequency

never exceeded 1.5 indicating low levels of cybersickness. Figure 12 displays the interaction plots for the different conditions and shows no clear trends. As shown in Fig. 13, it is quite dependent on the participant whether cybersickness occurred at all. The resulting model of the multi-level regression on PLC after backward elimination can be seen in Table 4 (within-R^2 = 0.04; between-R^2 = 0.00; rho = 0.50). Here as well the fit of the model is not good enough to draw any conclusions from it.

Intermediate discussion It was hypothesized (H1a, H1b, and H1c) that increasing the quality parameter (CRF) and adding freezing events negatively affect the QoE and increase the PCL and cybersickness. Furthermore, it was expected that the content of the video would negatively moderate the effect of quality degradations. As expected in H1a, results showed a stronger than linear effect of quality degradations on the QoE. The MOS drops strongly below the acceptable level in the CRF 36 conditions. CRF 28 was found to be on the border of acceptable quality. These results are in line with the findings of Tran, et al. (2017) [38] in terms of the trend. However, they reported on slightly higher scores on perceived quality. Against the expectation of H1c, the effect of quality degradations on the QoE is stronger for the more active videos (Dojo, Flamenco). It was also found that adding even a single freezing event would drop the QoE below the acceptable level, as expected, based on findings by Schatz, et al. (2017) [33]. The effect of quality degradation was found to be smaller when freezing events were added as the QoE is rated lower in higher quality conditions already. The PCL, however, was found to be low in all conditions and only slightly elevated in the lowest quality condition and after adding freezing events. Contrary to the expectations, increased PCL was not a concern in the conditions tested in this article. Overall, cybersickness was rated very low and the effects, even though significant, were so small one should wonder whether the difference can be consciously perceived. This is not in line with the expectations and is contradicting findings by e.g. [38] who found cybersickness to be a serious problem in 360-videos.

3.2.2 Effects of manipulations on visual attention

In the second step of the data analysis, it was to test the hypothesis H2a, H2b, and H2c. Analysing at the absolute values of area looked at in the x dimension, people on average looked at 2654 pixels of the total area of 4086 pixels, with an average standard deviation of

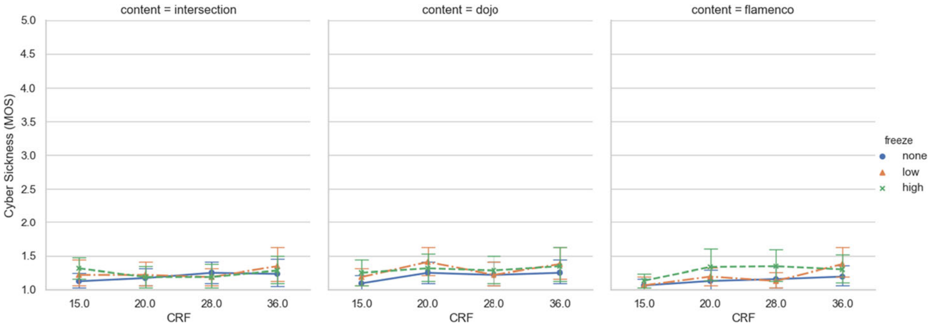

Fig. 12 MOS Cybersickness (y-axis) interaction plot for the different manipulation conditions. CRF on the X-axis, and a different line per freezing frequency

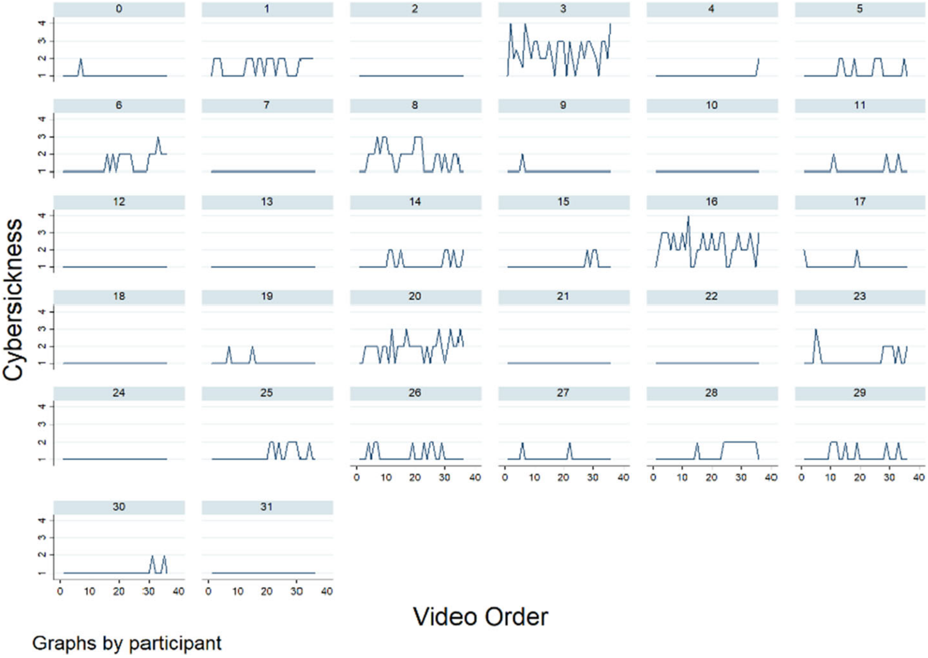

Fig. 13 Cybersickness MOS (y-axis) per participant. Over the course of the videos

645.5. Further analysis was done on the scale variable AreaY, which interaction plot can be seen in Fig. 14. The resulting multi-level regression model after backward elimination can be seen in Table 5 (within-R^2 = 0.13; between-R^2 = 0.04; rho = 0.25). in the model were CRF, Freeze, Content and total fixation duration on the spinner. Results showed a significant effect of CRF. The videos with CRF 20, 28, and 36 decreased the AreaX compared to the CRF 15 conditions; however, they do not differ significantly among each other. Furthermore, the AreaY is larger for the Dojo video compared to the Intersection and Flamenco video. Initially it seemed Freezing had a significant effect on the AreaX, however after further analysis it was found that this effect was mediated by the total fixation duration on the spinner (which appears during freezing events). No significant interaction effects were found.

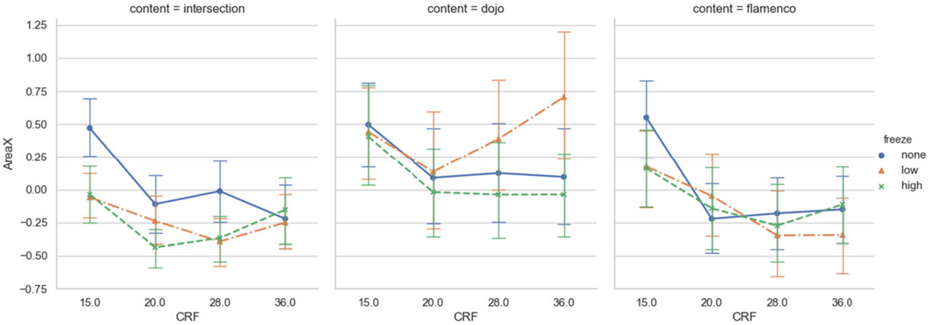

Fig. 14 Interaction plot of the manipulation effects on Area X (y-axis). CRF on the X-axis, and a different line per freezing frequency

Table 5 Regression coefficients of the multi-level regressions on the eye tracking scale variables

Predictors	AreaX	AreaY	RvsI	PCLfix
CRF (15)				
20	−0.415***			0.641***
28	−0.416***			0.699***
36	−0.351***			0.742***
Freeze	−0.005	−0.102*	0.088***	0.323***
Content (Intersection)				
Dojo	0.419***	−0.085	1.269***	0.195***
Flamenco	0.063	−0.134	1.633***	0.06
Total Fixation Duration Spinner	-1.78×10^{-4} ***	0.841×10^{-4} **		
Content*Freeze (Intersection)				
Dojo		0.167*		
Flamenco		0.135*		
Video Order			0.237***	
CRF*Freeze (15)				
20				−0.246***
28				−0.250***
36				−0.292***
_cons	0.237*	0.119	−0.944***	−0.736***

Note: *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

Next, the effects on the area looked at in the y-direction were analysed. Looking at the absolute value of the area looked at in the y-dimension, of the total area of 1970 pixels, people on average looked at 583 pixels with a standard deviation of 90. This shows people explored a smaller proportion of the y-dimension compared to the x-dimension. Further analysis was done on the scale variable AreaY. The resulting multi-level regression model after backward elimination can be seen in Table 5 (within-R^2 = 0.02; between-R^2 = 0.04; rho = 0.19). The model explains such a small proportion of the variance in the data that further analysis was discarded.

The third eye-tracking variable refers to how much participants looked at relevant or moving objects compared to the surroundings as a measure of selective attention. The mean of the absolute ratio is 0.64 with standard deviation 0.24, meaning on average participants fixated on relevant areas 64% of their total fixations. Further analysis was done on the scale variable RvsI, see Fig. 15 for the interaction plot. The proportion areas marked as relevant differ between the videos, this difference can be seen in Fig. 16, as the scores were lower in the

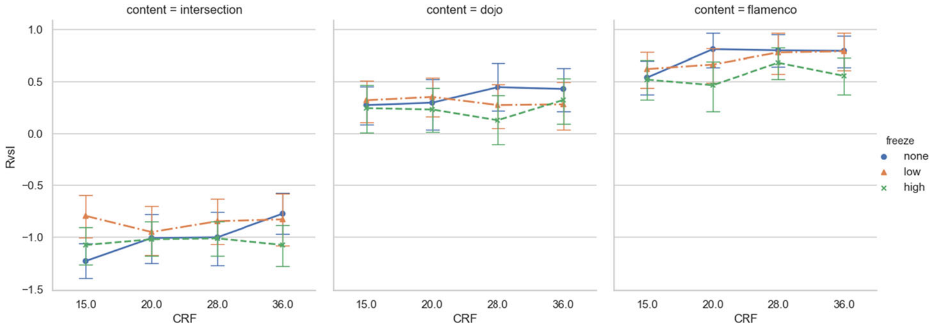

Fig. 15 Interaction plot of the manipulation effects on RvsI (y-axis). CRF on the X-axis, and a different line per freezing frequency

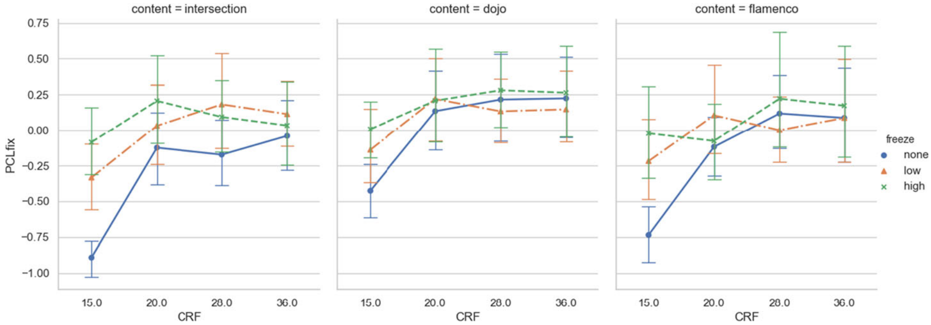

Fig. 16 Interaction plot of the manipulation effects on PCLfix (Y-axis). CRF on the X-axis, and a different line per freezing frequency

intersection video due to fewer areas marked relevant. The resulting model after backward elimination can be seen in Table 4 (within-R^2 = 0.67; between-R^2 = 0.02; rho = 0.27). Included in the model were: Freeze, Content, video order (As the difference between the first and the rest of the videos). Results showed a negative RvsI values for the Intersection video which significantly increased for the Dojo and Flamenco videos which for both resulted in positive values. In the Flamenco video the values were also significantly larger compared to the Dojo video. Furthermore, adding freezing events had a small effect. A difference between the first and the other videos was observed through the video order variable.

The last eye-tracking variable evaluated is PCLfix which refers to fixation data related to the perceptual and cognitive load. The mean fixation duration was 285 ms with standard deviation of 104 ms. The mean fixation count was 2.48 with standard deviation 0.46. Further analysis was done on the scale variable PCLfix, which interaction plots are visible in Fig. 16. The resulting model after backward elimination is displayed in Table 5 (within-R^2 = 0.14; between-R^2 = 0.14; rho = 0.37). Results showed a significant positive effect of CRF values 20, 28, and 36 compared to the CRF 15 conditions, however, they do not differ significantly among each other. Freezing also was found to have a significant effect, it was found that increasing the freezing frequency the PCLfix would increase likewise. Significant interaction effects between Freezing and CRF showed that for CRF 20, 28, and 36 the effect of Freezing decreased to almost nothing as the difference between the CRF 15 and other values was smaller for the low and high freezing condition. Furthermore, the PCLfix was significantly higher in the Dojo video compared to the Intersection and Flamenco video.

Intermediate discussion In the second step, the effects of the manipulations on participants eye movements and visual attention was evaluated to test whether they induced selective attention [19] and/or influenced the fixation duration [19, 43]. It was hypothesized (H2a and H2b) that degrading the video quality and adding freezing events would increase the PCL and thus increase the selective attention and the average fixation duration. Furthermore, it was expected that visual attention was affected by the natural content of the video and that videos with higher motion activity resulted in an increase in selective attention and the average fixation duration.

Results show participants on average looked at 50% of the total 360-degree horizontal field. As expected in H2a, for the degraded videos it was found that participants looked at a smaller area compared to the reference video. However, no difference between the degraded videos themselves was observed. As expected in H2b, it was found that content influences where people look and at how much of the total field they look. Looked at. For example, In the Dojo video, there was action covering more of the 360-degrees which can be observed in the data as well. On average, 64% of the total fixations were on relevant areas. Against the expectation of H2a, quality degradations do not have an effect on the selectiveness of attention. More selective attention was, however observed after repeating the same video. Additionally, the average fixation duration was also found to increase in the degraded videos compared to the reference but did not differ among the degradations self (H2a). Freezing events affected the visual attention, as expected in H2a, except for the AreaX for which the effect of freezing was due to the spinner.

3.2.3 Effects of visual attention on quality perception

In the third part of the analysis, these were combined to test how eye movements were related to the quality perception. In this part of the analysis hypothesis H2d and H2e were tested.

The resulting model after backward elimination can be seen in Table 4 column "QoE part 3", (within-$R^2 = 0.44$; between-$R^2 = 0.01$; rho = 0.65). Only PCLfix, RvsI, the interaction between PCLfix and CRF, and the interactions between RvsI and CRF and Freeze were additionally included in the final model. Of these only the interaction effects between RvsI & CRF and RvsI & Freeze were found to be significant. The contribution of the other eye-tracking variables (AreaY, and AreaY) to the model and their effects were too small to be included. Comparing results to the model without eye tracking data a negative effect of CRF, which is especially strong for CRF 36 is still observed. The effect of CRF is larger and more exponential for the Dojo and Flamenco video. Freezing frequency also still have a negative effect on the QoE, this effect is smaller in conditions with higher CRF values. The video order was also found to have small negative effect. A significant interaction effect between RvsI and CRF was found. In the CRF 20 (B = 0.181) and 28 (B = 0.201) conditions, RvsI had a positive moderating effect on the QoE which shifts to a negative moderating effect for negative RvsI values (Fig. 17). For higher values of RvsI this resulted thus in videos in the CRF 20 and 28 conditions to be rated higher than the reference (CRF 15). The interaction effect between RvsI and CRF 36 was not significant, keeping it constant for all RvsI values. Due to this the difference between CRF 36 and CRF 28 and 20 increases as the RvsI increases as well. Finally, there was a significant negative interaction effect between RvsI and Freeze, resulting the effect of RvsI to be smaller in the low and high freezing conditions (Fig. 17, right). RvsI was the only eye-tracking variable with effects on the QoE, however, as the manipulations were not found to affect RvsI (part 2), mediation was ruled out. Furthermore, a large rho of 0.65 shows that in the final model still 65% of the variance is still explained on participant level. Including eye tracking data in the model was not able to explain part of this variance.

Intermediate discussion In the third part of the analysis, the relation visual attention to the QoE is tested by including the eye tracking variables in the analysis of part 1. It was hypothesized that looking at moving objects moderates the effect of quality degradations and freezing events (H2d and H2e). Furthermore, it was expected that selective attention and focus as measures of PCL would (partially) mediate the effects of the manipulations (H2f). Additionally, individual differences in viewing behaviour were observed and it was therefore tested whether that could (partially) explain

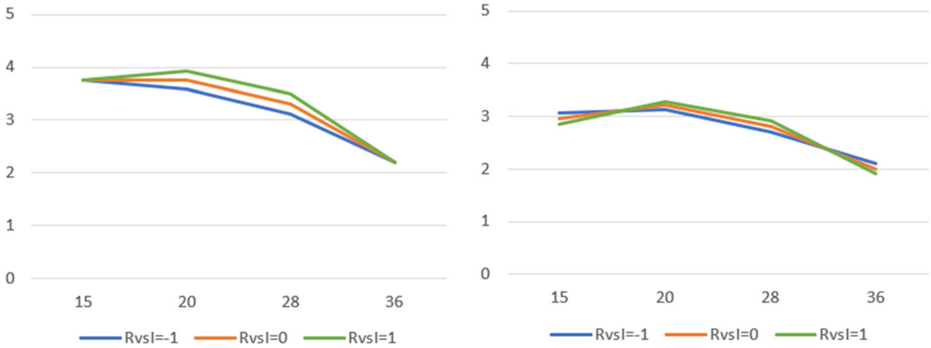

Fig. 17 Illustration of the interaction effect between CRF (x-axis) and RvsI(colours) on QoE(y-axis) for the Flamenco video with freezing event frequency none (left) or high (right). Shows the moderating effect of RvsI

the individual differences in subjective ratings. In H2c it was expected that the proportion of the total viewing space that was looked at is related to the QoE.

Against the expectation of H2c, AreaY was found to have no relation to the QoE, meaning that the proportion of the total viewing space looked at does not influence the quality perception or experience. Comparing the model of the first and the third part, RvsI had a significant interaction effect with both CRF and Freeze. In other words, the amount people look at relevant areas influences how the IFs influence the QoE. These results are partly in line with H2d as looking at moving objects was found to only have a small but positive moderating the effect on the effect of low and medium compression. The effects of freezing events were moderated by looking at moving objects as was expected in H2e; freezing is experienced as more intense when it disrupts the course of movements in the user's field of attention. Additionally, as the manipulations had some effects on the eye-tracking data, it was tested whether the eye tracking variables mediated any of the manipulation effects. However, this was found not to be the case, neither when removing outliers from the data or only testing data of later videos or of bad quality videos. These findings are thus not in line with H2f. Further comparing the two models, part 3 does not explain much more of the variance as the variance on participant level remains large. Adding visual attention variables does thus not further explain individual differences in QoE.

4 Discussion

This article tries to provide insights into users' visual behaviour and QoE in VR 360-videos. Observations and heatmaps showed that participants' attention was drawn to faces, moving objects, and surprising/unexpected events. It is known that people have an attentional bias towards faces [2]. Looking at signs and text can be explained by the desire to orient in one's environment. Thus, these human attentional behaviours as they have been observed in reality, were also observed in the VR environment. However, differences in visual behaviour between participants in a free viewing quality assessment were observed too. People's behaviour differs in terms of how much they looked around, how much they would look to the surroundings and whether they would be distracted by the spinner. Similar for most people was that they would look around less and focus more on one area after watching the same video several times.

Furthermore, the attention maps show indications of selective attention in videos with lower quality and/or freezing events which were further tested in the statistical analysis.

4.1 Manipulations and the subjective experience

The first part of the analysis provides answers to research question 1a: "How do video quality degradations, freezing events, and content relate to the QoE?" and research question 1b: "What is the threshold for an acceptable QoE (MOS > 3.5) regarding these influence factors?"

The overall QoE of the 360-videos was found to be rather low, even the visually lossless reference video would not reach a MOS of 4 (good). This shows that 4 K resolution is alright if an acceptable QoE is sufficient. However, to provide users with good quality and a more pleasant experience, higher resolutions may be required which in turn has higher bandwidth demands. Furthermore, as expected, it was found that quality degradations negatively influence the QoE. Overall, it was found that for 360-videos encoded in 4 K resolution and CRF 28 were rated on the border of acceptance. Adding a single freezing event, however, immediately drops the QoE below the acceptable level of MOS = 3.5; even in the high-quality videos. These findings support the results of [33, 38]. Thus, based on results from prior studies and the current article, one could thus argue that in the trade-off between bandwidth requirements and user satisfaction, increasing the compression and in this way avoiding freezing events would result in a better experience compared to maintaining a high-quality video with more risk of the occurrence of freezing. It should be noted that the effect of quality degradation was found to be stronger for videos with higher motion activity. Therefore, it would be recommended to compress these videos less if acceptable quality is desired.

In contrast to expectations, the subjective PCL was overall rated low, indicating that at least in the situation of the current study, watching 360-videos did not put much load on the participants. Merely in the lowest quality condition and with added freezing events, a minor increase was observed. As there are many aspects that can influence a person's PCL, a large effect was not expected. However, the size of the found effects and R^2 of the model are so small that the impact on the experience appears to be minimal. It could be concluded that under the conditions tested in this article, PCL does not cause any serious issues for QoE of 360-videos.

Cybersickness was also reported to be very low in all conditions, as the group means would never surpass MOS = 1.4. This is interesting, since other studies such as, for example, Tran, et al. (2017) [38], found cybersickness to be prominent problem in 360-videos. 93% of the participants experienced symptoms. It was also expected that the disruptions due to freezing events could cause discrepancies in movements, which in turn cause confusion and cybersickness symptoms [26]. It must be noted that all videos in the current study had a static camera position which could play a role here. Freezing events were found to influence the cybersickness, however, this effect was very small. Some participants did mention to be more annoyed by freezing events when there is more movement in the video. The occurrence of cybersickness also appeared to be more participant dependent. In conclusion, the level of cybersickness remains low across all conditions, showing that cybersickness is not an issue and is not seriously affected by quality degradations or freezing events, at least not for static camera positions.

4.2 Manipulations and visual attention

Effects on the area looked at in the vertical dimension were discarded due to a low R^2 and based on that heatmap showed that there was not much exploration in the vertical dimension.

The remaining data of the second part of the analysis was used to answer research question 2a: "How do video quality degradations, freezing events and content influence the viewer's eye movements?". On average 50% of the total horizontal area was looked at. Half the available viewing space was left unobserved. Furthermore, the area looked at was smaller in the degraded videos compared to the reference video. As the reference video was shown first, an order effect between the first and the remainder of videos could be a likely explanation. Repetition of the videos could therefore also explain the low average proportion that was looked at, as after the first time, less exploration and orientation were needed. The effect of freezing on the area looked at in the x-dimension was mediated by the total fixation duration on the spinner. Video content also matters. In the Dojo video participants looked at a larger area compared to the Intersection and Flamenco video.

Quality degradations did not influence selective attention. This is not in line with prior expectations. Following the theory by [19] global quality distortions did not cause more selective attention. This could be a sign that the distortions did not cause substantial annoyance such that perceptual and cognitive overload occurs. However, an increase in selective attention was observed after watching a video once. This could be interpreted as a form of late selective attention as the scene had already been evaluated during the first time the video was watched. Additionally, results show an increase in fixation duration and a decrease in fixation count (more focus) between the reference video and the quality degradations. However, these degraded videos did again not differ from each other, and an order effect is a potential explanation. The first time watching a video invites exploration, which would, following the theory, be the optimal level of load. Once the video is known, watching the same video could quickly become boring and unchallenging, resulting in under load and thus lower fixation counts and longer average fixation durations Wang, et al. (2014) [41]. Furthermore, freezing events increased the focus. Considering the study by [33] showing freezing does cause annoyance which is associated with cognitive load [10].

4.3 Relation between visual attention and the QoE

In the final step of the analysis, the eye-tracking data was added to the model on the QoE to test whether the gaze position influences the relationship between our independent variables and QoE. In 360-videos, different people may look at different parts or proportions of the available omnidirectional space; people are not exposed to the same view and might have different experiences. However, the area looked at was excluded from the model as it did not contribute significantly to the QoE. This means that even though people look at different parts of their 360-degree surroundings, this does not result in different quality perceptions or experiences. The resulting model is visualized in Fig. 18.

As no (partially) mediating effects were found of selective attention and/or focus, perceptual and cognitive overload cannot further explain why quality degradations and freezing events have a negative effect on the QoE. Neither subjective evaluations nor the eye-tracking data showed a clear relation between the manipulations, PCL and the QoE. The manipulations did not influence the PCL, such that selective attention and more focus occurs or have a significant influence on the users' experience.

What the model does show is how the ratio of the extent to which participants look at relevant versus irrelevant areas moderates the effect of CRF 20 and 28. In good and medium quality conditions, if participants look more at relevant areas (which are in general the moving objects), they would rate the QoE higher. In line with the theory on masking and contrast

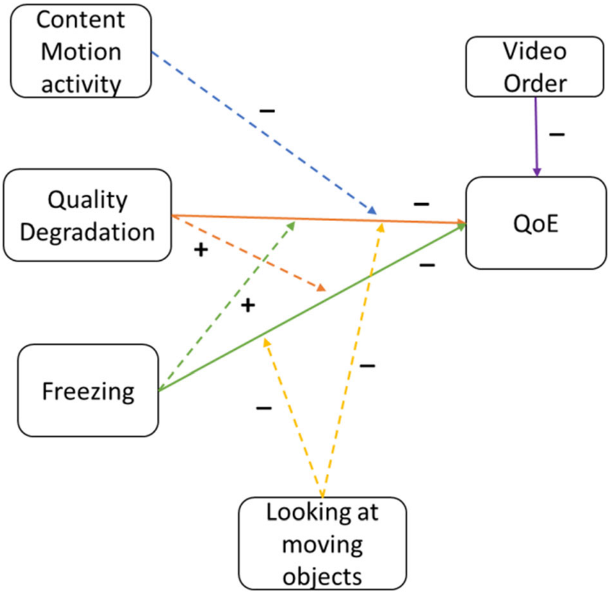

Fig. 18 Graphical illustration of the statistically significant effects. Solid lines are direct effects, dashed lines are interaction effects. Plus, and minus signs indicate positive and negative effects

sensitivity [21], attending to movement would make people less sensitive for other details, such as quality distortions which could, therefore, be noticed less resulting in the higher QoE rating. Depending on the condition and the effects of other factors, this can result in the degraded videos being rated higher than the reference video. For the Intersection video, for example, it occurred more often that a degraded video was rated higher than the reference compared to the other two videos. The effect of CRF was already found to be smaller in the Intersection video, thus differences were already less visible, and thus combined with the masking effect even less. Additionally, previously watched videos and memory could play a role as well. After seeing the worst quality, videos with CRF 20 could be perceived of higher quality due to the contrast. If the memory of the reference quality and rating is ceasing it can occur that a degraded video is rated higher than the reference. The effect of RvsI was absent in the CRF 36 conditions, indicating that the quality degradation is prominent enough to overcome temporal masking, the negative effect on the QoE is equally present regardless of how much one looks at moving objects. The effect of CRF 36 being constant means that as participants look more at moving objects, the difference in perceived quality and experience between CRF 36 and CRF 20 and 28 increases.

How much one looks at relevant areas was also found to have a moderating effect on the effect of freezing. The effect becomes more negative when people look more at moving objects. Movements are being disrupted as the images freezes; when looking at these movements this disruption is more salient and can thus cause more annoyance to the person.

4.4 Application and further research

Results from the subjective ratings may be used in the development of 360-video technology and applications regarding the trade-off between bandwidth requirements and user satisfaction.

In case of network resources reaching their limits, delays and freezing should be avoided. To save bandwidth quality could be degraded quite a bit before its negative effects are worse than those of freezing. Additionally, videos with more motion activity should be compresses less to remain an acceptable quality perception.

This article showed that visual attention on movement has a small effect on quality perception and experience. Accounting for effects of motion in salient regions alone might not be enough to successfully improve objective metrics by implementing visual attention. Therefore, if we want to include human perception in objective metrics, more research on possible other relations between visual attention and the QoE would be needed. For example, the effects discussed in this article are related to the visual attention on natural content (movements), a next step would, for example, be to study the relation of visual attention on (local) distortions and the QoE.

On a final note, based on the findings of the current study, the quality of VR 360-videos is not optimal yet. Where 4 K is considered a good resolution in 2D video, almost all video conditions in this study were rated only just above or below an acceptable level. Furthermore, the HMD screen is perceived as grainy, and it is uncomfortable to watch for a longer period. Some participants, for example, commented that it is quite frustrating they cannot see certain details they would be able to see in real life. Possibly, due to the more immersive experience of the VR environment, they expect a certain level of quality as they would expect in the natural world. As a more immersive environment has been linked to better experiences Salomoni, et al. (2017) [32], it would be interesting to study how the immersiveness is related to quality expectations and thus perception. Overall, more work and improvement are required to make streaming 360-video a successful, and pleasant experience. It is important that streaming in higher resolutions videos become more feasible and available. Additionally, HMDs could be further improved to be able to handle these higher resolutions.

4.5 Limitations

4.5.1 Technical limitations

The eye-tracking software could not show uncompressed material and took up too much CPU causing the program to freeze. The reference videos were compressed slightly, but still on a visually lossless level. Additionally, the reference had to be shown separately, causing a possible order effect as the reference was always shown first.

The AOI’s on which large part of the analysis is based had to be defined manually. This was done with most care and precision based on theory and a pilot. However, it remained a subjective task.

Furthermore, the spinner was only placed in one position during the freezing events which would significantly draw participants attention to the centre. In future studies and applications making use of spinners, it should be carefully considered where to place them.

Finally, there was no audio included in the current study to isolate the effects which were studied.

4.5.2 Experience related limitations

Some participants had indicated they did not see much difference between the videos, which was also visible in the data. This, on the one hand, shows that there are individual differences between people regarding quality perception. However, videos could have been degraded even further to have larger differences.

Participants also indicated they were quite focused on comparing videos and their own answers. This behavior would distract them from observing and experiencing video material.

Finally, it could get boring seeing the same video repeatedly. Randomization should account for this.

5 Conclusion

360-video is a video format that is currently growing in popularity and even though QoE research on 2D video streaming is well established, little is done yet on 360-video. 360-video brings new challenges and experiences along. Applying QoE theory and methods from 2D video streaming is not trivial and more research is one 360-video specifically is needed. QoE assessment metrics are still in development and subjective methods have yet to be standardized. The key contributions of this article are to provide subjective data on the effects of quality degradations, freezing, and content on the QoE; and to evaluate visual attention as influence factor in the QoE assessment.

The overall Quality and Experience was rated rather low, and answers varied a lot among participants. Degrading the quality does not have much impact on the QoE up until the threshold (here CRF 28), after which the QoE dropped. Freezing events drop the QoE below acceptable level in any condition. In general, higher qualities for 360-video would be required. However, if network resources are limited, more compression would be desired to avoid freezing. Furthermore, the results of this article show that PCL and cybersickness do not cause any serious issues for the QoE and are not much affected by the manipulations.

The effects of the manipulation on visual attention were minimal. It was found that attention was mainly directed by content, but also by surprising elements such as for example the spinner. It can be concluded that freezing does alter the visual attention, but that switching between different levels of degradation does not cause any changes and has thus no consequences for the experience in that way. If network resources are limited more compression would be preferred over freezing, which should be avoided.

Including eye-tracking metrics in the model did not further explain individual differences in subjective ratings as found in part 1. The proportion of the total viewing area looked at did not have any relation to the QoE. The only effect found was a small moderating effect. When looking more at moving objects, participants get less sensitive for quality distortions and the effect of freezing becomes slightly more negative. Effects found are small and the variance on participant-level remains high. Implementation of visual attention based on these results alone is most likely not enough to successfully improve objective QoE metrics. To do so, more research on other relations between visual attention and the QoE would be required.

Acknowledgements The economic support from the Internal funding at RISE and VINNOVA (Sweden's innovation agency, grant nr 2018-00735) is hereby gratefully acknowledged.

Authors' contributions (optional: please review the submission guidelines from the journal whether statements are mandatory)

Not applicable.

Funding Open access funding provided by RISE Research Institutes of Sweden. (information that explains whether and by whom the research was supported)

This project was funded by an internal funding at RISE and VINNOVA (Sweden's innovation agency, grant nr 2018–00735).

Data availability (data transparency)

Not applicable.

Declarations

Conflicts of interest/competing interests (include appropriate disclosures)
Not applicable.

Code availability (software application or custom code)
Not applicable.

Ethics approval (include appropriate approvals or waivers)
Not applicable.

Consent to participate (include appropriate statements)
All participating users signed a consent form before the participation in the study.

Consent for publication (include appropriate statements)
All authors have given their consent for publication.

References

1. Azevedo RGDA, Birkbeck N, Simone FD, Janatra I, Adsumilli B, Frossard P (2019) Visual Distortions in 360-degree Videos. IEEE Trans Circ Syst Video Technol 30:1–14. https://doi.org/10.1109/TCSVT.2019.2927344
2. Bindemann M, Burton AM, Hooge IT, Jenkins R, de Haan EH (2005) Faces retain attention. Psychon Bull Rev 12(6):1048–1053. https://doi.org/10.3758/bf03206442
3. Brunnström K, Barkowsky M (2018) Statistical quality of experience analysis: on planning the sample size and statistical significance testing. J Electron Imaging **27**(5):11. https://doi.org/10.1117/1.JEI.27.5.053013
4. Cummings JJ, Bailenson JN (2016) How immersive is enough? A Meta-analysis of the effect of immersive technology on user presence. Media Psychol 19(2):272–309. https://doi.org/10.1080/15213269.2015.1015740
5. Engelke, U., R. Pepion, P.L. Callet, and H.-J. Zepernick, (2010). *Linking distortion perception and visual saliency in H.264/AVC coded video containing packet loss*. Visual communications and image processing 2010. Vol. 7744. SPIE,
6. Engelke, U., M. Barkowsky, P.L. Callet, and H. Zepernick. (2010). *Modelling saliency awareness for objective video quality assessment*. In *2010 second international workshop on quality of multimedia experience (QoMEX)*. 2010.
7. Engelke U, Darcy DP, Mulliken GH, Bosse S, Martini MG, Arndt S, Antons JN, Chan KY, Ramzan N, Brunnström K (2017) Psychophysiology-based QoE assessment: a survey. IEEE J SelectTopics Signal Process 11(1):6–21. https://doi.org/10.1109/JSTSP.2016.2609843

8. FFMPEG. (2020). *FFMPEG: complete, cross-platform solution to record, convert and stream audio and video*. Ffmpeg. Org,
9. Gutierrez J, Perez P, Orduna M, Singla A, Cortes C, Mazumdar P, Viola I, Brunnstrom K, Battisti F, Cieplinska N, Juszka D, Janowski L, Leszczuk MI, Adeyemi-Ejeye A, Hu Y, Chen Z, Wallendael GV, Lambert P, Diaz C, … Garcia N (2021) *Subjective evaluation of visual quality and simulator sickness of short 360 videos: ITU-T rec. P.919*. IEEE Trans Multimedia:1–1. https://doi.org/10.1109/TMM.2021.3093717
10. Hart, S.G. and L.E. Staveland (1988). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Advances in Psychology Vol. 52*. P.A. Hancock and N. Meshkati, Editors. North-Holland. p. 139–183, https://doi.org/10.1016/S0166-4115(08)62386-9.
11. Holmqvist, K. and R. Andersson, (2017). Eye-tracking: a comprehensive guide to methods, paradigms and measures.
12. Ikehara, C.S. and M.E. Crosby. (2005). *Assessing cognitive load with physiological sensors*. In *proceedings of the 38th annual Hawaii international conference on system sciences*. 2005.
13. ITU-R (2019). *Methodology for the subjective assessment of the quality of television pictures* (ITU-R Rec. BT.500–14). International Telecommunication Union (ITU).
14. ITU-T. (2008). *Subjective video quality assessment methods for multimedia applications* (ITU-T Rec. P.910). International Telecommunication Union, Telecommunication standardization sector.
15. ITU-T. (2014). *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment* (ITU-T Rec. P.913). International Telecommunication Union, Telecommunication standardization sector.
16. ITU-T (2017). *Vocabulary for performance, quality of service and quality of experience* (ITU-T Rec. P.10/G.100). International Telecommunication Union (ITU), Place des Nations, CH-1211 Geneva 20.
17. ITU-T (2020). *Subjective test methodologies for 360° video on head-mounted displays* (ITU-T Rec. P.919). International Telecommunication Union, Telecommunication standardization sector.
18. Kuipers, F., R. Kooij, D. De Vleeschauwer, and K. Brunnström (2010). *Techniques for measuring quality of experience*, in *wired/wireless internet communications, lecture notes on computer science, volume 6074*. Springer-Verlag p 216-227.
19. Lavie N (1995) Perceptual load as a necessary condition for selective attention. J Exp Psychol Hum Percept Perform 21(3):451–468. https://doi.org/10.1037//0096-1523.21.3.451
20. Le Callet, P., S. Möller, and A. Perkis, eds. (2012). *Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*. 2012: Lausanne, Switzerland.
21. Möller S, Raake A (2014) Quality of experience - advanced concepts, applications and methods. Springer International Publishing, T-Labs Series in Telecommunication Services. Switzerland
22. Nidhi and N. Aggarwal (2014) *A review on Video Quality Assessment*. in *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*. 1–6. https://doi.org/10.1109/RAECS.2014.6799645.
23. Ninassi, A., O.L. Meur, P.L. Callet, and D. Barba. (2007). *Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric*. In *2007 IEEE international conference on image processing*. 2007.
24. Palmer, S.E., (1999). Vision science: photons to phenomenology. Bradford Bokk,
25. Pastrana-Vidal, R.R. and J. Gicquel. (2006). Automatic quality assessment of video fluidity impairments using a no-reference metric. Scottsdale, AZ, USA.
26. Porcino, T.M., E. Clua, D. Trevisan, C.N. Vasconcelos, and L. Valente. (2017). *Minimizing cyber sickness in head mounted display systems: design guidelines and applications*. In *2017 IEEE 5th international conference on serious games and applications for health (SeGAH)*. 2017.
27. Raudenbush SW, Bryk AS (2001) Hierarchical linear models applications and data analysis methods, vol 1. SAGE Publishing, Thousand Oaks, CA, USA. www.sagepub.com
28. Reiter, U., K. Brunnström, K. De Moor, L. Mohamed-Chaker, M. Pereira, A. Pinheiro, J. You, and A. Zgank (2014). *Factors Influencing Quality of Experience*, in *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Möller and A. Raake, Editors. Springer. p. 45–60, https://doi.org/10.1007/978-3-319-02681-7_4.
29. Robitza, W., (2017). *SITI: Spatial Information / Temporal Information*: https://github.com/slhck/siti.
30. Robitza, W., (2017). *Bufferer*: https://github.com/slhck/bufferer.
31. Robitza, W. and K. Brunnström. (2019). VQEGNumSubjTool - calculating number of subjects. Available from: https://slhckshinyappsio/number-of-subjects/, Access Date: 17 Nov 2019.
32. Salomoni P, Prandi C, Roccetti M, Casanova L, Marchetti L, Marfia G (2017) Diegetic user interfaces for virtual environments with HMDs: a user experience study with oculus rift. J Multimodal User Interf 11(2):173–184. https://doi.org/10.1007/s12193-016-0236-5

33. Schatz, R., A. Sackl, C. Timmerer, and B. Gardlo. (2017). *Towards subjective quality of experience assessment for omnidirectional video streaming*. In *2017 ninth international conference on quality of multimedia experience (QoMEX)*. 2017.
34. Søgaard J, Shahid M, Pokhrel J, Brunnström K (2017) On subjective quality assessment of adaptive video streaming via crowdsourcing and laboratory based experiments. Multimed Tools Appl 76(15):16727–16748. https://doi.org/10.1007/s11042-016-3948-3
35. Sweller, J., P. Ayres, and S. Kalyuga, (2011). *Cognitive load theory*. Explorations in the learning sciences, instructional systems and performance technologies. Vol. 1. Springer-Verlag New York. https://doi.org/10.1007/978-1-4419-8126-4.
36. Tobii (2019). *Tobii Pro Lab 1.11*. Available from: https://s3.amazonaws.com/lynx.tobii/TobiiProLab_1.138.26138_x64.exe, Access Date: 8 May 2020.
37. Tobii. (2019). *Tobii Pro Lab: User Manual*, Tobii Pro AB (https://www.tobiipro.com/).
38. Tran, H.T.T., N.P. Ngoc, C.T. Pham, Y.J. Jung, and T.C. Thang. (2017). *A subjective study on QoE of 360 video for VR communication*. In *2017 IEEE 19th international workshop on multimedia signal processing (MMSP)*. 2017.
39. van Kasteren, A. *The Contribution of Eye Tracking to Quality of Experience Assessment of 360-degree video*. (Doc nr: 0845668 and acr062449), Human Technology Interaction and Visual Media Quality, Eindhoven University of Technology and RISE Research Institutes of Sweden AB, Eindhoven, The Netherland and Kista, Sweden, M. Sc. thesis. 2019
40. van Kester, S., T. Xiao, R. Kooij, and K. Brunnström. (2011). *Estimating the impact of single and multiple freeze occurrences on video quality*. In *proc. of SPIE-IS&T Human Vision and electronic imaging XVI*. Burlingame, CA, USA: SPIE and IS&T. p. Paper 25.
41. Wang Q, Yang S, Liu M, Cao Z, Ma Q (2014) An eye-tracking study of website complexity from cognitive load perspective. Decis Support Syst 62:1–10. https://doi.org/10.1016/j.dss.2014.02.007
42. Winkler S (2005) Digital video quality: vision models and metrics. J. Wiley & Sons, Chichester, West Sussex
43. Zu, T., J. Hutson, L.C. Loschky, and N.S. Rebello (2018). Use of eye-tracking technology to investigate cognitive load theory. arXiv e-prints arXiv:180302499, Available from: https://uiadsabsharvardedu/abs/2018arXiv180302499Z, Access Date: March 01, 2018.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Anouk van Kasteren received a BSc in Industrial Design with a minor in cognition and social psychology from the University of Technology Eindhoven (TU/e) (2017) and a MSc in Human Technology Interaction from the TU/e (2019) with an exchange in Data Science at the Kungliga Tekniska Högskolan (KTH). She completed her master thesis project at RISE, Research Institutes of Sweden in Stockholm.

Kjell Brunnström Ph.D., is a Senior Scientist at Acreo Swedish ICT AB and Adjunct Professor at Mid Sweden University. He is an expert in image processing, computer vision, image and video quality assessment having worked in the area for more than 25 years. Currently, he is leading standardization activities for video quality measurements as Co-chair of the Video Quality Experts Group (VQEG). His current research interests are in Quality of Experience for visual media in particular video quality assessment both for 2D and 3D, as well as display quality related to the TCO requirements.

Chris Snijders is professor of the Sociology of Technology and Innovation at Eindhoven University of Technology. His research interests include Human-Technology Interaction, human processing of artefacts and artificial intelligence, and the behaviour of humans in the online world.

Affiliations

Anouk van Kasteren[1,2] • **Kjell Brunnström**[1,3] • **John Hedlund**[1] • **Chris Snijders**[2]

[1] RISE Research Institutes of Sweden AB, Kista, Sweden

[2] Eindhoven University of Technology, Eindhoven, The Netherlands

[3] Mid Sweden University, Sundsvall, Sweden