Assessing the influence of combinations of blockiness, blurriness, and packet loss impairments on visual attention deployment

Alexandre F. Silva[a], Mylène C.Q. Farias[a] and Judith A. Redi[b]

[a]University of Brasília, Department of Electrical Engineering, Brasília, Brazil;
[b]Delft University of Technology, Department of Intelligent Systems, Delft, The Netherlands

ABSTRACT

In digital video systems, impairments introduced during the capture, coding/decoding processes, delivery and display might reduce the perceived quality of the visual content. Recent developments in the area of visual quality have focused on trying to incorporate aspects of gaze patterns into the design of visual quality metrics, mostly using the assumption that visual distortions appearing in less salient areas might be less visible and, therefore, less annoying. Most of these studies, however, have considered the presence of a single artifact (e.g. blockiness or blur) impairing the image. In practice, this is not the case, as multiple artifacts may overlap, and their combined appearance may be strong enough to deviate saliency from its natural pattern. In this work, our focus is on measuring the impact and the influence of combinations of artifacts on the video saliency. For this purpose, we tracked eye-movements of participants in a subjective quality assessment experiment during a free-viewing and a quality assessment tasks. Results show that the gaze locations change from pristine videos to impaired videos. These changes seem to be more related to the quality level and content of videos than to the specific combination of artifacts.

Keywords: Impairment, Perceived quality, Gaze pattern, Visual quality.

1. INTRODUCTION

In digital video systems, impairments introduced during the capture, processing (coding and decoding), delivery and display might reduce the perceived quality of the visual content. An impairment is any (visible) defect in a video signal that can be decomposed into a set of perceptual features called artifacts[1,2]. Automatically predicting the annoyance of such impairments (using quality metrics[3]) is of major importance for implementing quality control loops in video delivery systems.

Recent studies show that the assessment of video quality is closely tied to gaze deployment[4]. When observing a scene, the human eye typically scans the video neglecting areas carrying little information, while focusing on visually important regions[5]. Wang et al.[6] showed that, within the first 2000 ms of observation, gaze patterns are targeted to main objects in the image. Later, the gaze is redirected to other salient, yet not visually important, areas. This result suggests that visual coding should be focused, at first, into the main objects of the scene. The presence of artifacts may disrupt these natural gaze patterns, causing viewer's annoyance and, consequently, a lower quality judgments[7]. Therefore, saliency information should be incorporated into video quality metrics.

Several researches in the area of visual quality have focused on trying to incorporate gaze pattern information into the design of visual quality metrics[8], mostly using the assumption that visual distortions appearing in less salient areas might be less visible and, therefore, less annoying[9,10]. However, while some researchers report that the incorporation of gaze pattern information increases the performance of quality metrics, others report no or very little improvement[11]. One possible reason for such disagreement is that, still, the role played by visual attention in quality evaluation is unclear. Although it has been shown that, for images, artifacts in visually

Further author information:
A.F.S.: E-mail: alexandrefieno@gmail.com,
M.C.Q.F.: E-mail: mylene@ieee.org,
J.A.R.: E-mail: J.A.Redi@tudelft.nl

Human Vision and Electronic Imaging XX, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 9394, 93940Z · © 2015 SPIE-IS&T
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2083487

important regions are far more annoying than those in the background[12], it is still not clear if artifacts can create saliency (and therefore, attract gaze) on their own. And if so, it is unclear which type of artifacts can create saliency and at what perceptual strength. If artifacts can disrupt gaze patterns by creating saliency, this should be taken into account in the design of quality metrics that make use of saliency or gaze pattern information. Unfortunately, the existing knowledge in this direction is scattered.

Ninassi et al.[13] studied viewing behavior during both free-viewing and quality assessment of impaired images. They found two results: (1) the quality task has a significant effect on the fixation duration, which increased on unimpaired images during a quality scoring task and, (2) the type of impairment degrading the image causes modifications in gaze patterns. Redi et al.[7] also analyzed the impact of three kinds of artifacts (JPEG compression, white noise and gaussian blur) on the deviation of gaze patterns during both quality assessment and free-viewing of images. As in many studies (see Engelke et al[8] for an overview) in[14] gaze pattern deviations were measured by analyzing similarities among saliency maps. Based on this type of analysis, Redi et al.[14] report that differences between saliency maps for both tasks seem to be more related to the strength of the artifacts impairing the images than to the type of those artifacts. With respect to video, Le Meur et al.[15] examined viewing behavior during both quality assessment and free-viewing tasks. Differently from images, they found that the average fixation duration is almost the same for both tasks; whereas saliency does not change significantly when videos are impaired (coding artifacts). Redi et al.[16], investigated to what extent the presence of packet-loss artifacts influences viewing behavior. Contrary to Le Meur et al.,[15] they showed that saliency can significantly change from free-viewing to quality assessment tasks, and that these changes were related to both video content and to packet-loss annoyance. Similarly, Mantel et al.[17] found a positive correlation between coding artifacts annoyance and fixation dispersion. From these results, it seems that, for both images and videos, some artifacts (e.g. packet loss) may be able to divert gaze and viewing behavior from their natural paths. But, it is yet unclear when and how this happens. It is important to point out that most studies have focused on analyzing the impact that artifacts in isolation have on gaze patterns: e.g., only blockiness[17,18], or only packet-loss[16,19]. In real-life situation, it is very likely that different artifacts are co-present in a video. For example, packet-loss may occur in the transmission of a severely compressed video, creating perceptual degradations that are very different from the single artifacts in isolation. To the best of the authors' knowledge, there is no study that explores the impact of combinations of artifacts on gaze patterns and viewing behavior.

The aim of this paper is to examine viewing behavior during both quality assessment and free-viewing of videos impaired with multiple artifacts. More specifically, we aim at detecting differences in 1) fixation duration and 2) spatial gaze allocation for videos containing combinations of blockiness, blurriness, and packet-loss. We report the outcomes of an eye-tracking study during which observers were asked to freely view at pristine videos and score the annoyance of a set of impaired versions of those videos. The resulting eye-tracking data are converted into saliency information (i.e., saliency maps averaged across all participants for each video and under each viewing condition) and analyzed to detect any changes in gaze locations due to both task and artifact annoyance. This paper is organized as follows. Section 2 presents the experimental methodology. Section 3 presents data processing and analysis of the experiment. A discussion of our results is provided in Section 4 and we conclude the paper in Section 5.

2. EYE TRACKING: EXPERIMENTAL SETUP

To investigate the impact of the presence of multiple artifacts on users gaze patterns, we tracked eye-movements of subjects in a quality assessment experiment. The test sequences used in the experiment contain a set of combinations of the most relevant artifacts found in digital video applications: packet-loss, blockiness and blurriness. Blockiness and blurriness are among the most relevant artifacts caused by compression[20], while packet-loss is a temporal artifact introduced by the automated concealment of the loss of packets in digital transmissions[16].

2.1 Stimuli

Seven high definition videos with spatial resolution 1280×720 (50fps) were used in the experiment. Figure 1 shows sample frames of all original videos used in the experiment. The videos were all ten seconds long and were chosen with the goal of including diversity in content in our test set.

Figure 1. Frames of videos: (a) Park Joy, (b) Into Tree, (c) Park Run, (d) Romeo and Juliet, (e) Cactus, (f) Basketball, and (g) Barbecue.

To choose the originals, we followed the recommendations of the "Final Report of VQEG on the validation of objective models multimedia quality assessment (Phase I)", which states that the set of video sequences should have a good distribution of spatial and temporal activity[21]. Fig. 2 (a) shows a graph of the spatial and temporal measures of the originals used in the experiment[22]. To add blockiness and blurriness, we used the strategy described in[23]. Once blocky and blurry signals were created, impaired sequences with combinations of blockiness and blurriness were generated by linearly combining the original video with the artifact signals in different proportions. To generate "packet-loss" artifacts on those sequences, we compressed the videos with H.264 at high compression rates (in order to avoid inserting additional artifacts), then we randomly deleted packets from the coded video bitstream in different percentages (the higher the percentage of lost packets, the stronger the perceptual effect). More details on the procedure can be found in[16].

We combined blockiness, blurriness and packet-loss artifacts at different proportions and strengths:

- 2 different strengths of blockiness and blurriness, namely 0.4 and 0.6 times the strength of the full artifact signal;
- 2 different packet-loss ratios: 0.7% and 8.1%.

These settings were chosen based on previous experiments.[1] These experiments showed that there is a statistically significant difference in the annoyance produced by packet-loss ratios of 0.7% and 8.1 %. Similarly, we chose only the strengths 0.4 and 0.6 for blockiness and blurriness, which are more representative of these artifacts. An overview of all combinations used in the experiment is given in table 1. We included, besides the pristine video (combination 1), three videos impaired with a single artifact (Combinations 2 to 4), eight videos impaired with packet-loss in combination with either blockiness or blurriness (combinations 5 to 12), and eight videos with combinations of all possible artifacts. Eventually, a total of 7 originals (see Figure 1) and 19 combinations were used in this experiment, resulting in $19 \times 7 + 7 = 140$ test sequences.

2.2 Methodology and Equipment

Twenty-one unpaid subjects participated in the experiment (17 males and 4 females). They were mostly graduate students from Delft University of Technology, the Netherlands. They were considered naïve of most kinds of digital video defects and the associated terminology. No vision test was performed on the subjects, but they were asked to wear glasses or contact lenses if they needed them to watch TV. While watching the video sequences, their eye-movements were recorded using a *SensoMotoric Instruments* REDII Eye Tracker with a sampling rate of 50/60Hz. It has a pupil tracking resolution of $0.1°$ and a gaze position accuracy of 0.5 to 1. The various stimuli were displayed on a Samsung LCD monitor of 23 inches (Sync Master XL2370HD) with resolution 1920 $\times$ 1080. The dynamic contrast of the monitor was turned off, the contrast was set at 100 and the brightness at 50. The measured gamma of the monitor for luminance, red, green, and blue was 0.99, 0.97, 1.00, and 0.92, respectively. To guarantee stability of the eye-tracking equipment a constant illumination at approximately 70 lux was used. The user interface for the experiment was implemented using the *Neurobehavioral Systems* software Presentation. Subjects were kept at a fixed distance of 0.7 meters from the monitor using a chinrest. The environment was compliant to ITU-T Recommendation BT.500[24].

Table 1. Combinations used in the experiment. The parameter 'bloc' corresponds to the blockiness strength, the parameter 'blur' corresponds to the blurriness strength, and the parameter 'P' corresponds to the packet-loss ratio.

Combination	P	bloc	blur
1	0.0	0.0	0.0
2	0.0	0.6	0.0
3	0.0	0.0	0.6
4	8.1	0.0	0.0
5	0.7	0.0	0.4
6	8.1	0.0	0.4
7	0.7	0.0	0.6
8	8.1	0.0	0.6
9	0.7	0.4	0.0
10	8.1	0.4	0.0
11	0.7	0.6	0.0
12	8.1	0.6	0.0
13	0.7	0.4	0.4
14	8.1	0.4	0.4
15	0.7	0.4	0.6
16	8.1	0.4	0.6
17	0.7	0.6	0.4
18	8.1	0.6	0.4
19	0.7	0.6	0.6
20	8.1	0.6	0.6

A single stimulus setup with hidden reference[24] was chosen for this experiment. A continuous 100-point annoyance scale was displayed after every video. After a brief oral introduction, the experiment started with the calibration of the eye-tracker for each observer. This was carried out based on a 13-point grid. After this, participants were asked to freely view the 7 original (i.e., unimpaired) videos. They were instructed to watch them as if they were sitting home and watching TV. Next, participants performed a training stage get acquainted with the type of distortions they could expect in the scoring test. The video sequences provided in the training were not intended to be scored. Rather, they were meant to become visual anchors (referneces) for the annoyance scoring. These videos were thus highly impaired. Participants were instructed to score videos as annoying as those seen in the training as '100', videos half as annoying as '50', and so on. After the training, the actual scoring session started. It was divided into three sub-sessions to limit fatigue effects. Between sub-sessions, subjects were allowed to rest their eyes for a few minutes. The entire experiment took on average 60 minutes, including the calibration and two breaks.

3. ANALYSIS OF EYE-TRACKING DATA AND QUALITY SCORES

The experiment produced two types of output: eye-tracking data and quality scores. For both outputs, we had one recording per participant and video. The quality scores collected for each video sequence were averaged over subjects to obtain a mean annoyance value (MAV). This represents the level of annoyance experienced by an average observer while watching a given video.

The eye tracking data consisted of pupil movements, recorded in terms of fixation points and saccades. In this paper, we limit ourselves to the analysis of fixation data, which is considered to be one of the most informative data regarding viewing behavior. Specifically, we analyze viewing behavior by looking at two quantities: the duration of the fixations, found to be impacted by quality scoring in Ninassi et al.,[13] and the spatial deployment of gaze patterns. We quantify the latter via saliency maps;[25] which represent, per each video pixel, the probability that it will be gazed at by an average observer. This choice is in line with most of the existing literature in the area.[8]

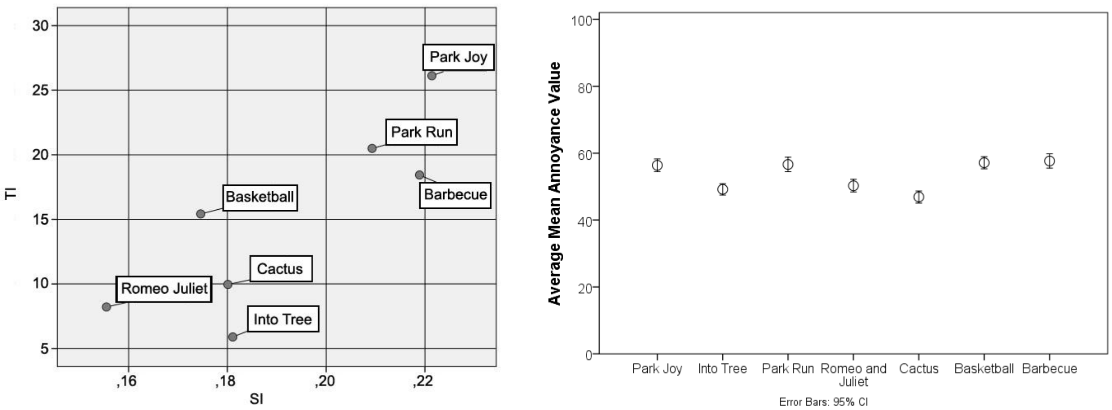


Figure 2. (a) Temporal (TI) and spatial (SI) characteristics of the videos included in the experiment (b) Average Mean Annoyance Values (MAVs) computed over all the distorted versions of each video.

For each video, we record the fixation points on which the subject's pupil rests for at least 100 ms. Then, we process the fixation data to create the saliency maps[25] that are used to determine the most visually attractive areas in a scene. For videos, the fixation data is recorded at a frame level (i.e., every 20ms). However, calculating a saliency map for each video frame gives an excessive granularity, as compared to the duration of a fixation (around 400ms). Therefore, we adopt the same strategy used in our previous work[16]: for a given video, we group the fixations from all participants in time windows of 400ms, generating *fixation maps*. To compute the saliency map corresponding to a specific time window of a specific video, we smoothed the fixation maps by applying a Gaussian patch with width equal to approximately the fovea size (2° visual angle). This procedure was applied for each video sequence, creating $\frac{10,000}{400} ms = 25$ saliency maps for each video.

We create saliency maps for each test video. For the sake of analysis, the saliency maps are clustered into the following groups:

1. FL_{PV} corresponding to pristine videos, captured during the free-viewing task;
2. SC_{PV} corresponding to pristine videos, captured during the quality assessment task;
3. SC_{G1} corresponding to test sequences with artifacts in isolation (Combinations 2 to 4 in Table 1), captured during the quality assessment task;
4. SC_{G2} corresponding to test sequences with combinations of two artifacts (Combinations 5 to 12 in Table 1), captured during the quality assessment task;
5. SC_{G3} corresponding to test sequences with combinations of three artifacts (Combinations 13 to 20 in Table 1), captured during the quality assessment task.

Our goal is to analyze if viewing behavior changes across these five groups. Thus, our analysis has two independent variables: task (free-viewing or quality assessment) and degradation (pristine or impaired). For impaired videos, we are also interested in checking whether the number and/or type of artifacts impact the saliency maps and the fixation durations.

3.1 Similarity measures for detecting saliency changes

We use similarity measures as indicators of changes in saliency distribution and therefore in gaze patterns[14,26]. The following measures are adopted to estimate the extent to which saliency maps corresponding to a certain time window of a certain video changes across the groups indicated above.

1. Linear correlation coefficient (LCC) $\in$ [1 -1], which quantifies the strength of the linear relationship between two saliency maps.
2. Structure Similarity Index (SSIM[27]) $\in$ [0 1], which indicates the extent to which the structural information of a map is preserved in relation to another map.

In both similarity measures, a value close to 1 indicates high similarity, while a value of 0 indicates dissimilarity, in turn suggesting a consistent change in the image saliency and consiquently of the spatial allocation of gaze. We also use the measure Upper Empirical Similarity Limit (UESL), to account for content and inter-observer variability[14]. UESL represents the similarity of saliency maps obtained under the same experimental conditions (e.g., while observing the same video, at the same level of impairment, and under the same task), but for two different groups of participants. As such, it expresses the extent to which two saliency maps are similar given individual differences in participants. UESL represents a useful benchmark to understand whether dissimilarity in maps, as measured after a change in experimental conditions (e.g., between a free-viewing map and a quality assessment map), is due to inter-subject variability rather than to the change in experimental conditions.

We calculate UESL based on LCC using with the following equation:

$$UESL(LCC, v_i) = \frac{1}{T}\sum_{t=1}^{T} LCC\left(SM\left(v_{i,t}^{FL_{PVC}}\right), SM\left(v_{i,t}^{FL_{PVO}}\right)\right), \tag{1}$$

where $SM(v_{i,t}^{FL})$ indicates the saliency map computed for time slot t, video i, and observer group FL. The saliency maps $SM(v_{i,t}^{FL_{PVC}})$ are recorded in our experiment from the pristine videos during the free-viewing task (FL_{PV}). The saliency maps $SM(v_{i,t}^{FL_{PVO}})$ were also recorded from the pristine videos during the free-viewing task, but from a previous experiment.[16] To compute the UESL based on SSIM, we simply substitute LCC by SSIM in eq. 1.

4. RESULTS

4.1 Annoyance of multiple artifacts

Fig. 2 (b) represents the mean MAV per video content, averaged across all 19 combinations. We can notice that the MAV are slightly different, across videos: the videos 'Into Tree' (49.19 ± 19.18), 'Romeo and Juliet' (50.27 ± 21.71) and 'Cactus' (46.90 ± 20.62) obtain a relatively smaller MAV than 'Park Joy' (56.38 ± 21.33), 'Park Run' (56.62 ± 24.73), 'Basketball' (57.14 ± 20.92) and 'Barbecue' (53.45 ± 22.31). These findings are in line with previous works available in literature, which observed that video content influences MAVs. Nevertheless, in this study the impact of video content on MAV seems to be smaller than what was found by Redi et al.[16] using the same videos and the same packet-loss artifacts. A possible cause for this difference is that in the present study we use two additional artifacts (blockiness and blurriness). The annoyance of blocky and blurry videos may depend less on the temporal characteristics of the video.

Figure 3 shows the distribution of MAV across the several combinations of artifacts used in the test (SC_{PV}, SC_{G1}, SC_{G2} and SC_{G3}). The average MAV for SC_{G1} (39.67 ± 12.26) is lower than for SC_{G2} (47.26 ± 15.58) and SC_{G3} (71.40 ± 12.92). *A one-way ANOVA shows that these differences are statistically significant (F = 1475.98, df = 2, p = .000).* There are, however, some exceptions. The combination 5 (P = 0.7%, blur = 0.4) has a lower MAV than the combination 3 (only blur = 0.6), which has the smallest MAV of SC_{G1}. Maybe, in this case, the presence of packet-loss artifacts is masking the blurriness. Another exception is combination 2 (only bloc = 0.6) that is more annoying than half of the combinations in SC_{G2}, which are mostly combinations of packet-loss and blurriness artifacts. This suggests that the blockiness artifact, when in isolation or in the presence of other artifacts, is more annoying than the other two artifacts. However, further analysis is needed to investigate the interactions among artifacts.

4.2 Fixation duration

In order to analyze the viewing behavior, we study the fixation duration recorded during free-viewing and quality assessment tasks. Figure 4 shows the average fixation duration per video content for both tasks. We run a repeated measures ANOVA using the video groups as the independent variables and the fixation duration

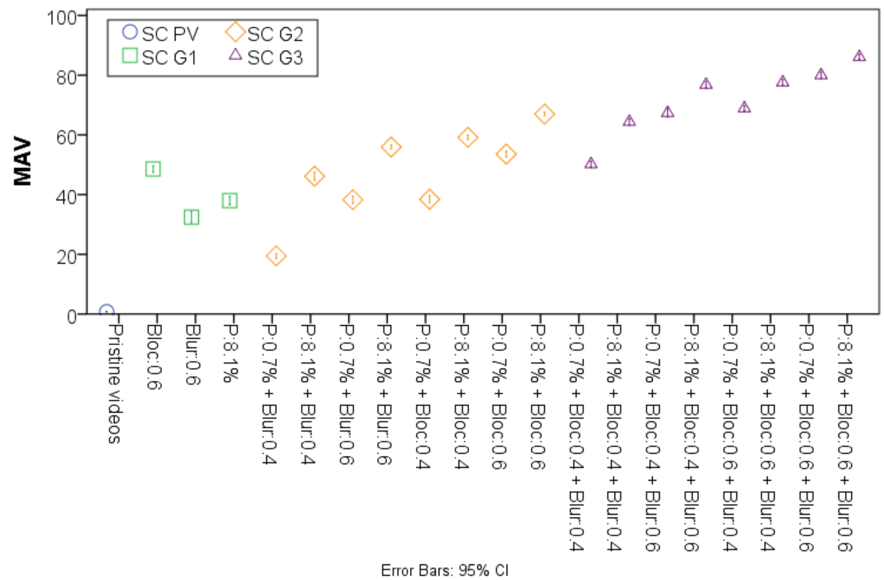


Figure 3. Average MAV over all videos for all combinations of artifacts (see Table 1).

as the dependent variable. We do not find a significant difference between the average fixation duration for free-viewing and quality assessment tasks (F(5.809, 116.175) = 2.113, p = 0.059). These results suggest that the average fixations duration is similar when free-viewing and quality assessment tasks are considered, which contraddicts the finding of Redi et al.,[16] who showed significant differences in the average duration fixations between free-viewing and quality assessment tasks. Our result is instead in line with what found by Le Meur et al.[18] As in our experiments, videos in[18] included blocky artifacts, whereas videos in[16] were impaired by packet loss. We may hypthesize then that packet loss artifacts casue an increase in the duration of fixation; but since in our experiment packet loss artifacts were overlapping with blocking artifacts (which seem not to impact on fixation duration), their effect may be reducted.

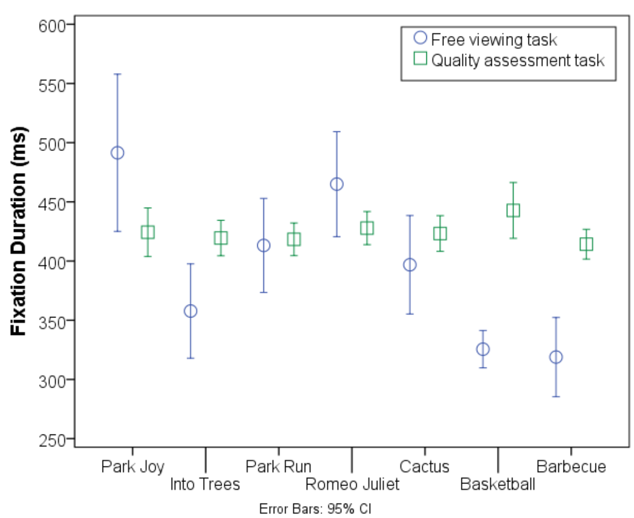
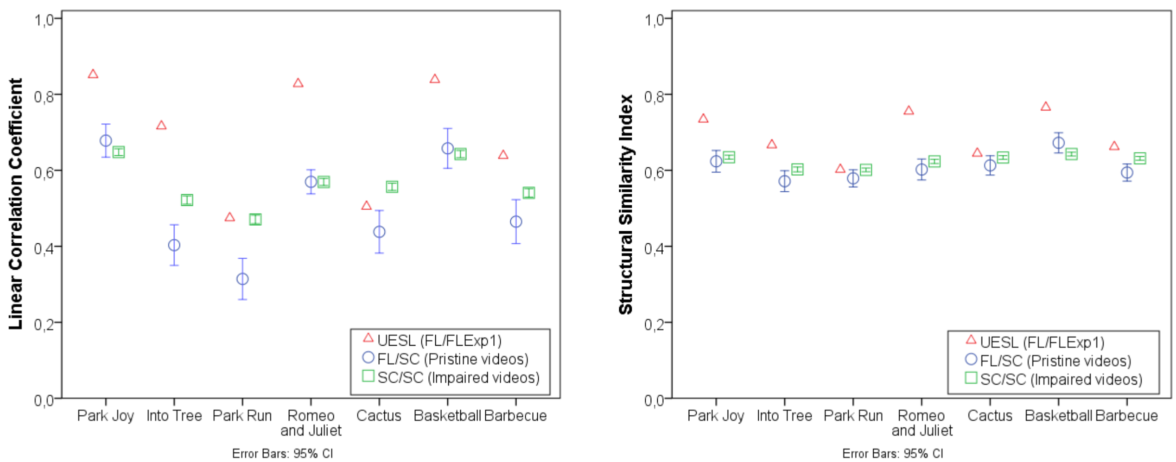


Figure 4. Average fixation duration for free-viewing (blue circles) and quality assessment (green squares) tasks.

4.3 Similarities among saliency maps

To take a closer look at the viewing behavior for sequences with combinations of blockiness, blurriness, and packet-loss artifacts, we look into the similarities between saliency maps for free-viewing and quality assessment tasks. Figure 5 shows (a) LCC and (b) SSIM between saliency maps computed for the same video under different tasks: quality assessment (SC_{PV}) and free-viewing (FL_{PV}). The UESL values are also included to represent inter-observer variability. Notice from this graph that the similarity between quality assessment and free-viewing maps is systematically lower than the UESL, showing that task does have an impact on viewing behavior.

To check whether the presence of artifacts altered the viewing behavior, we calculate the similarity between saliency maps obtained for pristine and impaired videos, during a quality assessment task. We consider all combinations used in SC_{G1}, SC_{G2}, and SC_{G3} (133 impaired videos used in quality assessment tasks) as a single video group (SC_SC). Since the saliency distributions are gathered for the same task (quality assessment), we expect the similarity measures for SC_SC to be close to the UESL values. Figure 5 shows the results of this test. It can be noticed that the saliency of pristine and impaired videos (quality assessment task) is different. Similarity measures are lower than UESL values, showing that the presence of artifacts (in this case, combined) alters saliency maps. These results are in agreement with what was found by Redi et al.[16].



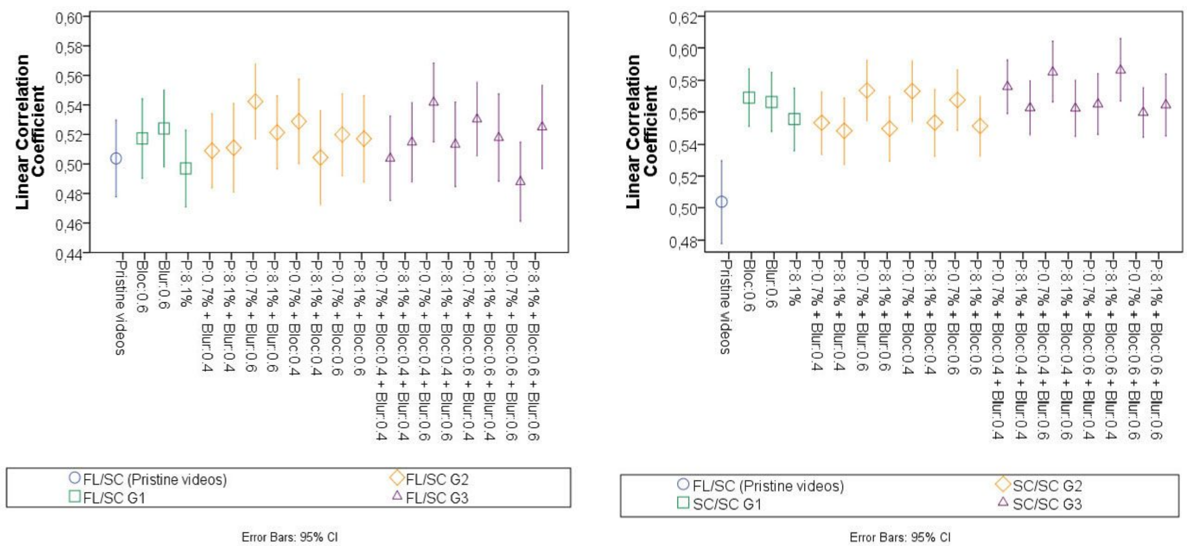
(a) (b)

Figure 5. (a) LCC and (b) SSIM Similarity measures computed between maps obtained from pristine videos during free-viewing and quality assessment tasks.

Next, we verify if this attention change is influenced by the different combinations of artifacts. Figure 6 (a) shows the similarity (in terms of LCC) between the FL_{PV} maps and the SC_{PV} maps. Additionally, it shows the similarity between the maps corresponding to free looking of videos and the maps for the scoring of the coorresponding video impaired with all different combinations of artifacts (thus, belonging to $FL/SC_{G1}, FL/SC_{G2}$ and FL/SC_{G3}). It can be seen from this figure that, at least quantitatively, the change in saliency is independent of the number or the type of artifacts. A one-way ANOVA reveals that the similarities among saliency maps for FL/SC and $FL/SC_{G1}, FL/SC_{G2}$ and FL/SC_{G3} are not statistically significant (F = 0.971, df = 19, p = 0.493).

Similarly, we compare the saliency maps for FL/SC and $SC/SC_{G1}, SC/SC_{G2}$, SC/SC_{G3}. As shown in Fig. 6 (b), the specific combination also does not seem to play a role in the saliency changes. Compared to the impact of task (see the blue mark on the left side of Fig 6 (b) corresponding to the comparison of the saliency maps of pristine videos for free-viewing and quality assessment tasks), the impact of specific artifact combinations seems negligible. A one-way ANOVA reveals indeed that the similarity for FL/SC is significantly lower than that of $SC/SC_{G1}, SC/SC_{G2}$ and SC/SC_{G3} (F = 3.155, df = 19, p = 0.000).

Although the specific artifact combinations do not seem to have an impact on gaze locations, there may still be a relationship between the perceived quality of the video and saliency distribution. To check this hypothesis, we measure the similarity of saliency maps of pristine and impaired videos in quality assessment task. Figure 7 shows how $LCC(SC_{PV}, SC\_SC)$ varies depending on the MAV of the impaired videos. We consider 3 categories of MAV: $MAV < 30$, $30 < MAV < 60$ and $MAV > 60$. A one-way ANOVA reveals that the LCC between these maps is significantly different among categories of MAV (F = 10.483, df = 2, p = 0.000). Notice that the similarity among saliency maps obtained from scoring pristine and impaired videos increases with the annoyance of the artifacts.



(a) (b)

Figure 6. Similarity among saliency maps computed LCC for (a) pristine videos during free-viewing and quality assessment tasks and (b) pristine and impaired videos during quality assessment tasks.

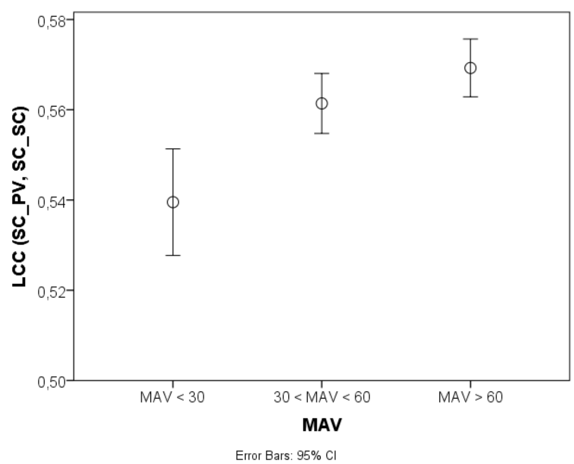


Figure 7. LCC Similarity among saliency maps obtained scoring pristine and impaired videos, for different categories of MAV.

5. DISCUSSION AND CONCLUSIONS

In this paper, we studied the effect combinations of artifacts have on viewing behavior. With this goal, we tracked eye movements of 21 participants while they were watching videos impaired with combinations of blockiness, blurriness, and packet-loss. Then, we analyzed the viewing behavior of our participants in terms of fixation durations and spatial gaze allocation. Our results indicate that the presence of impairments has no impact on the duration of fixations. Nevertheless, analyzing saliency maps we were able to detect changes in gaze deployment. In particular, we measured the similarity of saliency maps corresponding to the same video captured for different tasks (free-viewing or quality assessment), types of impairments (different combinations of packet-loss, blockiness and blurriness), or a combination of the two. Our results show that differences in viewing behavior exist due to a change in task. Also, the presence of impairments in the video impacts the saliency distribution. We did not find an effect of a specific type of artifact combination on saliency changes.

Interestingly, the similarity measure of the saliency maps increased with the increase of the artifact annoyance. This is a counter intuitive result, as one would expect more annoying artifacts to be visually stronger and thus create saliency on their own. A possible explanation for this result is that, whereas for combinations with low MAV the impact of localized packet-loss artifacts is more evident, for more annoying combinations, this localized effect may have been masked by the presence of other artifacts. So, the source of annoyance may become indistinctively diffuse across the whole video area. For this reason, further analysis is needed to link the change in saliency to physical properties of the video.

Two main points need to be addressed in the future. First, we need to establish whether the change in saliency detected by our similarity metrics relates to the divergence of fixations outside the core region of interest of the video or it comes from a convergence of fixations within it. This may indicate if artifacts create saliency on their own. Second, it is necessary to further understand the link between saliency changes and artifact type, annoyance, and location.

REFERENCES

[1] Farias, M. C. Q., Heynderickx, I., Espinoza, B. L. M., and Redi, J. A., "Visual artifacts interference understanding and modeling (varium): A project overview," in [*Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics*], VQPM (2013).

[2] Keelan, B., [*Handbook of image quality: characterization and prediction*], Marcel Dekker, Inc., New York (2002).

[3] Lin, W. and Jay Kuo, C. C., "Perceptual visual quality metrics: A survey," *Journal Vis. Comun. Image Represent.* **22**(4), 297–312 (2011).

[4] Vu, E. and Chandler, D., "Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience," in [*Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on*], 73–76, IEEE (2008).

[5] Desimone, R. and Duncan, J., "Neural mechanisms of selective visual attention," *Annual review of neuroscience* **18**(1), 193–222 (1995).

[6] Wang, J., Chandler, D. M., and Le Callet, P., "Quantifying the relationship between visual salience and visual importance," in [*IS&T/SPIE Electronic Imaging*], 75270K–75270K, International Society for Optics and Photonics (2010).

[7] Redi, J., Liu, H., Zunino, R., and Heynderickx, I., "Interactions of visual attention and quality perception," *Proc. SPIE* **7865**, 78650S–78650S–11 (2011).

[8] Engelke, U., Kaprykowsky, H., Zepernick, H.-J., and Ndjiki-Nya, P., "Visual attention in quality assessment: Theory, advances, and challenges," in [*IEEE Signal Processing Magazine*], **6**, 50–59 (2011).

[9] Farias, M. C. Q. and Akamine, W. Y. L., "On performance of image quality metrics enhanced with visual attention computational models," *Electronics Letters* **48**, 631–633(2) (2012).

[10] Redi, J., Liu, H., Gastaldo, P., Zunino, R., and Heynderickx, I., "How to apply spatial saliency into objective metrics for jpeg compressed images?," in [*Image Processing (ICIP), 2009 16th IEEE International Conference on*], 961–964, IEEE (2009).

[11] Le Meur, O. and Le Callet, P., "What we see is most likely to be what matters: Visual attention and applications," in [*Image Processing (ICIP), 2009 16th IEEE International Conference on*], 3085–3088, IEEE (2009).

[12] Alers, H., Redi, J., Liu, H., and Heynderickx, I., "Studying the effect of optimizing image quality in salient regions at the expense of background content," *Journal of Electronic Imaging* **22**(4), 043012–043012 (2013).

[13] Ninassi, A., Le Meur, O., Le Callet, P., Barba, D., Tirel, A., et al., "Task impact on the visual attention in subjective image quality assessment," in [*Proceedings of European Signal Processing Conference*], (2006).

[14] Redi, J. and Heynderickx, I., "Image quality and visual attention interactions: Towards a more reliable analysis in the saliency space," in [*Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*], 201–206 (Sept 2011).

[15] Le Meur, O., Ninassi, A., Le Callet, P., and Barba, D., "Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric," *Signal Processing: Image Communication* **25**(7), 547–558 (2010).

[16] Redi, J., Heynderickx, I., Farias, M. C. Q., and Macchiavello, B., "On the impact of packet-loss impairments on visual attention mechanisms," *IEEE International Symposium on Circuits and Systems (ISCAS)* , 1107–1110 (2013).

[17] Mantel, C., Nathalie, G., Patricia, L., Gelu, I., and Thomas, K., "Characterizing eye movements during temporal and global quality assessment of h.264 compressed video sequences," *Proc. SPIE* **8291**, 82910Y–82910Y–12 (2012).

[18] Le Meur, O., Ninassi, A., Le Callet, P., and Barba, D., "Do video coding impairments disturb the visual attention deployment?," *Signal Processing: Image Communication* **25**(8), 597–609 (2010).

[19] Engelke, U., Pepion, R., Le Callet, P., and Zepernick, H.-J., "Linking distortion perception and visual saliency in h. 264/avc coded video containing packet loss," in [*Visual Communications and Image Processing 2010*], 774406–774406, International Society for Optics and Photonics (2010).

[20] Farias, M. C. Q. and Mitra, S. K., "Perceptual contributions of blocky, blurry, noisy, and ringing synthetic artifacts to overall annoyance," *Journal of Electronic Imaging* **21**(4), 043013 (2012).

[21] (VQEG), V. Q. E. G., "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase i," tech. rep., Video Quality Experts Group (VQEG) (2008).

[22] Ostaszewska, A. and Kloda, R., "Quantifying the amount of spatial and temporal information in video test sequences," in [*Recent Advances in Mechatronics, Springer*], 11–15 (2007).

[23] Farias, M., *No-Reference and Reduced Reference Video Quality Metrics: New Contributions*, PhD thesis, University of California, Santa Barbara, California (2004).

[24] International Telecommunication Union, *ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures* (1998).

[25] Koch, C. and Ullman, S., "Shifts in selective visual attention: towards the underlying neural circuitry," in [*Matters of Intelligence*], 115–141, Springer (1987).

[26] Riche, N., Duvinage, M., Mancas, M., Gosselin, B., and Dutoit, T., "Saliency and human fixations: State-of-the-art and study of comparison metrics," in [*Computer Vision (ICCV), 2013 IEEE International Conference on*], 1153–1160 (Dec 2013).

[27] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on* **13**(4), 600–612 (2004).