



HAL
open science

Characterizing eye movements during temporal and global quality assessment of h.264 compressed video sequences

Claire Mantel, Nathalie Guyader, Patricia Ladret, Gelu Ionescu, Thomas Kunlin

► **To cite this version:**

Claire Mantel, Nathalie Guyader, Patricia Ladret, Gelu Ionescu, Thomas Kunlin. Characterizing eye movements during temporal and global quality assessment of h.264 compressed video sequences. IS&T/SPIE Electronic Imaging: Science and Technology, Jan 2012, San Francisco, United States. 11 p. hal-00684679

HAL Id: hal-00684679

<https://hal.science/hal-00684679>

Submitted on 2 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Characterizing eye movements during temporal and global quality assessment of h.264 compressed video sequences

Claire Mantel^{a,b}, Nathalie Guyader^a, Patricia Ladret^a, Gelu Ionescu^a and Thomas Kunlin^b

^aGIPSA-Lab - Department of Signal and Images, Grenoble Institute of Technology - UMR CNRS
5216, 961 rue de la Houille Blanche, Grenoble - France;

^bSTMicroelectronics S.A., 12 Rue Jules Horowitz B.P. 217, Grenoble - France

ABSTRACT

Studies have shown that the assessment of image or video quality is closely linked to the deployment of visual attention measured through the eye movements of observers. However, this link is not fully understood yet. While most studies have considered the evaluation of global quality (i.e. produced by all types of defects) this study focuses on the evaluation of what we call temporal quality (i.e. produced by defects varying from one frame to another). In this study, we measured whether assessing the temporal quality of a video influences the visual exploration of the corresponding video.

We set-up a subjective experiment in which the eye movements of observers are recorded during three different tasks: a free-viewing task (FT), a global quality assessment task (QA-G) and a temporal quality assessment task (QA-T). The FT acts as a reference representing the standard visual exploration for our video database in perfect quality. Eye movements recorded during this task are compared to those obtained during the two quality assessment tasks for each sequence at different quality levels.

As previously shown, we found that observers assessing global quality gaze at locations dissimilar to those fixated during the FT. It is however not the case for locations fixated by observers assessing temporal quality. The fixated locations from QA-T are more similar to the ones from FT. Moreover, our results suggest that video quality level does not influence the locations fixated but has an impact on the fixation durations: the lower the quality is, the longer observers maintain their fixations. Finally, fixation locations are less variable between observers during the quality assessment tasks than during the FT for either perfect or poor quality level.

We also analyzed the evolution over time of these eye movement parameters (fixation locations and durations). It appears that all indicators are similar for the three tasks during the first 1 or 2 seconds of the videos and that differences between tasks appear later on. Hence, it seems that the task to assess video quality plays a role afterwards on the deployment of the participants' visual attention.

1. INTRODUCTION

Compressing a video sequence creates various kinds of artifacts. The impairments produced by each type of artifact combine to create the global impression of quality. This paper mainly focuses on a specific type of artifacts that we call 'temporally annoying artifacts'. We define these artifacts as defects induced by compression that are annoying because of their evolution over time, regrouping in this category variation of spatial defects (¹) and specifically temporal defects (such as mosquito noise^{2,3}). Those 'temporally annoying artifacts' contain both spatial and temporal aspects but their transient nature is their main characteristic and they cannot be isolated from other artifacts. Therefore those defects play a role complicated to evaluate in the overall quality of a video. We call '*temporal quality*' the impression of quality produced by a video while considering only those '*temporally annoying artifacts*'.

This paper proposes a behavioral experiment whose aim is to better characterize how observers visually explore videos while assessing global quality and while assessing 'temporal quality'. This understanding might help developing quality metrics to measure temporal quality or improving the way video quality metrics take into account 'temporally annoying artifacts'.

Further author information: (Send correspondence to C.M.)
C.M.: E-mail: claire.mantel@gipsa-lab.grenoble-inp.fr

In the field of video quality, the first metrics that were developed were composed of image algorithms applied independently to frames one after another. The first reason to this approach is that the memory space and computing time required to work on a series of images is much larger than the one to work on one image at a time. Thanks to hardware performances progress, this limitation is less and less an issue. The second cause is the importance granted to the spatial aspects of a video sequence compared to the temporal ones concerning quality perception. Spatial defects are more simple to work on because it is possible to isolate the spatial aspects of compression defects by working on frames extracted from a video sequence.

As a result, there have been few studies on the perception of temporal noises and, consequently, few studies have worked on how to measure them. Several studies however classify temporal defects among the main artifacts for h.264 video compression.^{1,4,5} It is then interesting to define the specificity of temporal quality assessment (QA-T) as a preliminary for designing a metric dedicated to evaluating 'temporally annoying artifacts' induced by h.264 video compression.

To better understand what parts of videos the observers use while they assess video quality, it is interesting to record their eye movements⁽⁶⁻⁹⁾. Those recordings can then be compared with other recordings from different quality levels but also with the recordings of eye movements from observers freely exploring the same videos. Redi et al. showed that locations fixated during global quality assessment (QA-G) are different from those obtained during free-viewing exploration and that this deviation depends on degradations: the lower the quality is, the higher the deviation is. They found that the main difference is a shrinking of the area fixated: as quality level decreases, observers seem to concentrate their gaze on smaller regions⁽⁹⁾. LeMeur et al. found that QA-G task moderately modifies fixated locations for videos. However, this modification does not seem to depend on the quality level but rather on the fact that observers watch several times the same video under different quality level.⁸

The aim of this paper is to characterize the influence of temporal quality assessment (QA-T) over eye-movements. The results focus on two parameters extracted from eye movements, or more exactly from eye fixations: first, the fixated locations and second, the durations of the fixations. We precisely study the influence of task (global and temporal quality assessment) and of quality level on those two parameters. We also analyze the evolution of those parameters over time. To do so, we recorded eye-movements of observers while they were looking at video sequences in three different conditions: free exploration without instructions, exploration to assess global quality and exploration to assess temporal quality. The free-viewing exploration acts as a standard representing '*natural viewing*' and the global quality assessment exploration acts as a standard representing '*global quality rating viewing*'. To shorten notations, the three tasks are designated with the following abbreviations: QA-G for the Global Quality Assessment task, QA-T for the Temporal Quality Assessment task and FT for the free-viewing task.

The experiment we set up is detailed in section 2. To study the eye-movements recorded during the experiment, we computed three indicators: dispersion of locations fixated between observers, resemblance of locations fixated during QA-G and QA-T to the locations fixated during FT and, finally, fixation durations. These indicators are analyzed in section 3, before the conclusion of this paper presented in section 4.

2. THE SUBJECTIVE EXPERIMENT

The purpose of this experiment is to collect data on the observers' eye movements while they watch video sequences under three different tasks: FT, QA-G and QA-T. For the two quality assessment tasks (QA-G and QA-T) video sequences are displayed at different quality levels.

2.1 Stimuli

We used 10 different video sequences from the CDVL database⁽¹⁰⁾. A frame extracted from each sequence is shown in figure 1. Those sequences are originally 1080p videos with a 4:2:2 color sampling but for practical reasons we subsampled the color space to 4:2:0 and we downsampled the videos to 720p using Lanczos filter.

Each video sequence was then compressed at 4 bitrates in the h.264 norm⁽¹¹⁾ using two different coding configurations: constant bitrate (CBR) and variable bitrate (VBR). The chosen configurations correspond to the two main categories of



Figure 1: A frame extracted from each sequence used for the experiment

encoding: CBR for broadcasting TV and VBR for storage (Blu-Ray). We used two different coding configurations because they play an important role in the resulting quality (⁴). Each sequence was cut to 10s, taking care of avoiding its first and last 2s during which the coding quality can be unstable. The stimuli therefore consisted in 9 different versions of the 10 video sequences: the original (perfect) quality and the compressed 8 versions (2 configurations x 4 bitrates).

The 4 bitrates are different for each sequence as can be seen in Figure 2. The bitrates were chosen fitly for each video to produce 4 different quality levels that are, in our opinion, similar from one sequence to another.

Sequencename	Frame Per Second	Quality 1	Quality 2	Quality 3	Quality 4
BbCrowdRedPlain	25	700Kb/s	1,2Mb/s	2Mb/s	3Mb/s
BbScore	25	700Kb/s	1,2Mb/s	1,8Mb/s	3Mb/s
BlueSpruce	25	1,5sMb/s	2,5Mb/s	3,5Mb/s	5Mb/s
DoubleEndBagintro	29,97	500Kb/s	1Mb/s	2Mb/s	3Mb/s
Goose	29,97	1Mb/s	2Mb/s	3Mb/s	5Mb/s
Kickoff	29,97	500Kb/s	1Mb/s	2Mb/s	3Mb/s
AspenWalk	29,97	1Mb/s	1,5Mb/s	2,5Mb/s	4Mb/s
NightLights	29,97	500Kb/s	1,2Mb/s	2,5Mb/s	4Mb/s
RedGoldenCage	25	500Kb/s	1Mb/s	2Mb/s	3Mb/s
Tulip	25	1Mb/s	2Mb/s	3Mb/s	5Mb/s

Figure 2: Frame per second and bitrates chosen for each sequence

2.2 Methodology

The progress of an experiment session is schematized in figure 3. Participants began by watching each of the 10 sequences in perfect quality in a random order without any instruction: this part of the experiment is later referred to as the free-viewing task (FT).

The experimenter then explained to the participant that the rest of the test session consisted in a quality assessment task: observers had to watch videos and to rate their quality. One half of observers participated in the Global Quality rating task (QA-G) and was asked to assess the quality of videos without further explanations.

Conversely, the other half of observers participated in the Temporal Quality rating task (QA-T) and were asked to rate the temporal quality of the video sequences. The explanation was done in three phases:

- First the observers were given a semantic explanation of the concept of temporal quality and 'temporally annoying artifacts'. Temporal quality was defined as the impression of quality produced by a video while considering only 'defects that are annoying because they vary from one frame to another'. It was specified that the 'temporally annoying artifacts' can be 'thin as a pixel or thick as groups of pixels'.

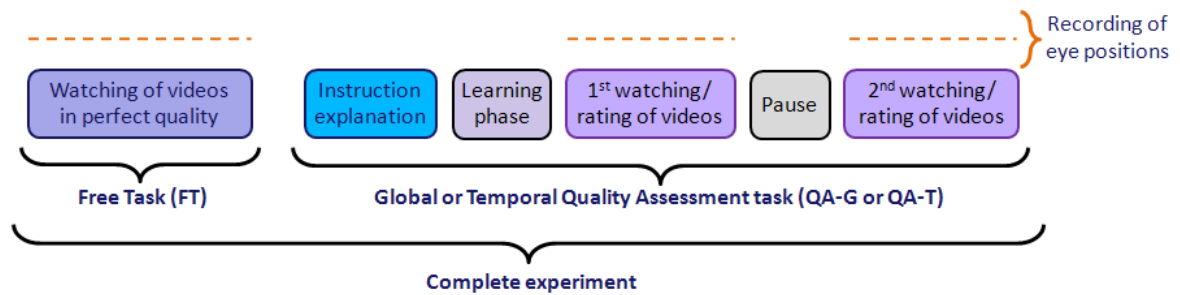


Figure 3: Organization of an experiment session

- Then observers were asked to put in practice this definition: they watched compressed videos and had to tell where they saw 'temporally annoying artifacts'
- Finally, their identifying of 'temporally annoying artifacts' was validated by the experimenter who pointed out the locations containing 'temporally annoying artifacts' in the above-mentioned videos

We used 5 different video sequences for observers to grasp the concept of temporal quality. Those videos are different from those used during the experiment. Examples of the videos used and are visible in Figure 4. During this explanation, we also took care of displaying a video presenting static blocking defects to observers. It allowed to emphasize that even though it was a very noticeable defect, it should not be taken into account for rating temporal quality as it did not belong in the 'temporally annoying artifacts'.

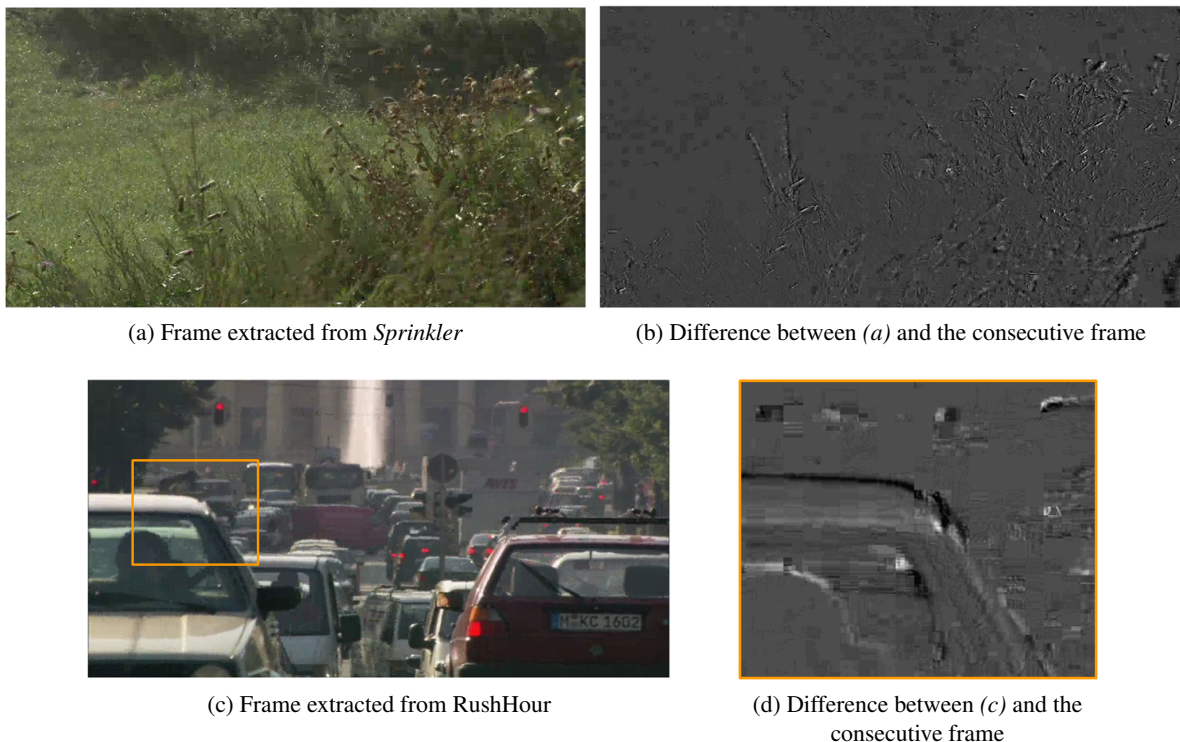


Figure 4: (a) and (c) are frames from two of the videos used to explain what are 'temporally annoying artifacts'. Image (b) and (d) illustrate the variation from one frame to another. Those variations are due to the motion in the video and to the 'temporally annoying artifacts' present in the videos (the difference frame was not shown to observers, it only serves here as a way to show 'temporally annoying artifacts' on paper).

During the learning phase, observers performed a few tryouts to make sure they took correctly in hand the rating pro-

cess. Then, they performed the QA-G or QA-T two times: they watched and rated all of the 90 stimuli displayed in random order two times (following recommendation from ITU¹² to check the rating of a video).

To avoid fatigue and weariness among participants, we split each of the two quality ratings performed by one observer in two 15min parts separated by a short break. The average overall duration of a test session was 1h15 for observers participating in the QA-G task and 1h20 for those participating in the QA-T task (the difference is due to the explaining of what we call the temporal quality).

2.3 Observers

36 observers participated in the experiment: 18 of them performed the QA-G task and the other 18 the QA-T task. The participants were 18 men and 18 women and their age ranged from 23 to 35. They all had a normal or corrected-to-normal visual acuity and correct color vision (tested with Snellen chart and Ishihara plates). All observers were naive regarding the purpose of the experiment and were no expert in video compression.

2.4 Set-up

Video sequences were displayed without upscaling on a Sony Bravia KDL40EX500-AEP TV placed at 1,5m from the observers which corresponds to 3 times the height of the videos, as advised by VQEG in.¹³ The experimental room has dark blue walls, it is isolated from exterior lighting and was lit by two fluorescent daylight-colored (6500K) light bulbs. We used an EyeLink II eye-tracker from SR Research. We recorded eye movements at a frequency of 500Hz in a monocular pupil tracking mode.

Observers were asked to rate the quality of videos on a 9-point scale, shown in Figure 5. Only the upper and lower grades of this scale were defined, respectively with *excellent* and *really bad*, to avoid influencing the observers answers (as advised in^{14,15}). We used the same scale but different questions for the two quality assessment tasks (QA-G and QA-T).

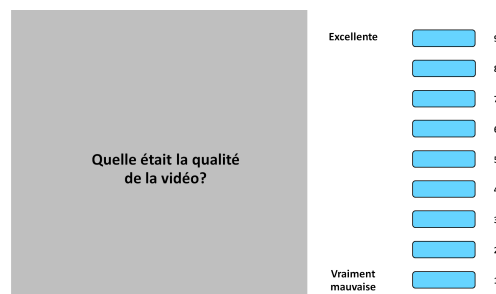


Figure 5: The rating scale for the global quality evaluation task (QA-G) in French. The question asked to participants translates as 'What was the quality of the video?'

3. RESULT ANALYSIS

We collected several type of data during this experiment: eye-movements recording and quality grades from observers. The ratings obtained during the experiment seem coherent and consistant with a quality assessment experiment, as can be seen in Figure 6: the grades decrease with the decrease of quality.

However, the purpose of this paper is to specify the differences between the way an observer watches a video in a free-viewing task (FT), while assessing its global quality (QA-G) or while assessing its temporal quality (QA-T). Hence in this result analysis, we do not focus on the quality rating but rather on the eye movements recorded during the different tasks and for the different quality levels. From our data, we extracted three indicators: dispersion, that represents the variability between observer 'gazes, ROC, that compares the spatial positions of the locations fixated between the FT and the two QA tasks, and finally durations of the recorded fixations*.

*We also analyzed saccade speed and duration. As they seemd redundant with fixation analisys, we do not present them here.

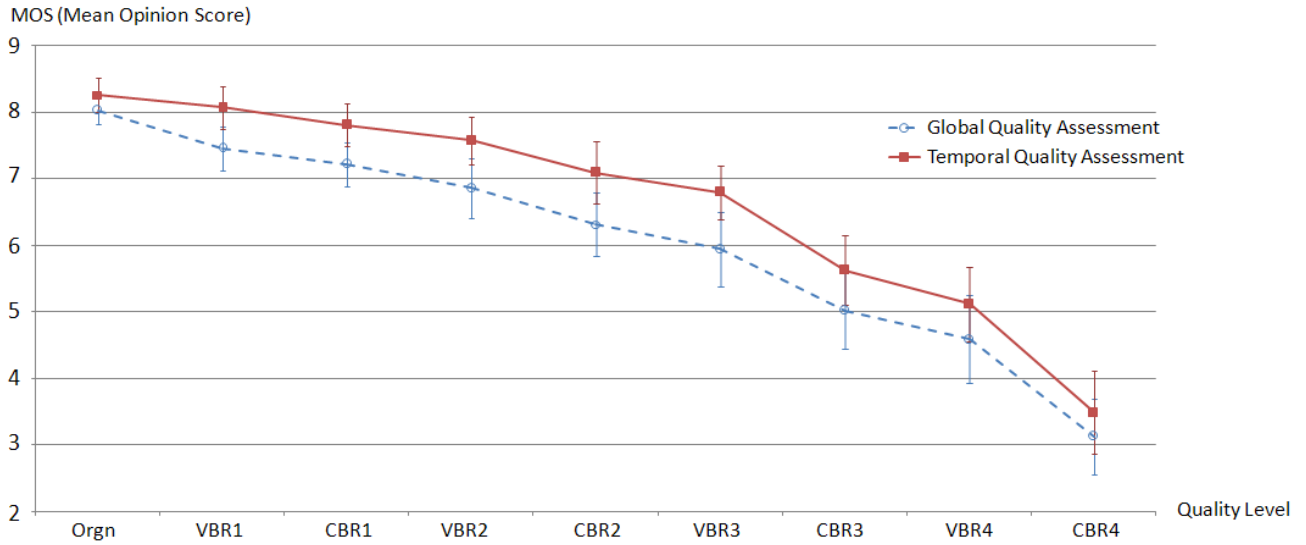


Figure 6: MOS (Mean Opinion Scores) averaged over all observers and all sequences and their corresponding Confidence Interval for the QA-G and QA-T tasks

For the two first indicators, we used the raw data recorded by the eye tracker each 2 ms. The eye movements recordings contained 20 eye positions per frame for 25Hz videos (or 16 eye positions per frame for videos at 30Hz) for each observer. We computed the median position of these 20 positions, hence we got one position per observer for each frame of a video sequence. We used these positions for all observers to compute the dispersion, which is the variability between observers. We also used these data to compare eye positions recorded during the FT and the two QA tasks. For the last indicator we used the fixation-events detected by the eyelink rather than using the raw eye positions. We extracted the duration of these detected fixations.

3.1 Protocol Validation

Before analyzing the results of this experiment, we tried to evaluate whether observers performing the QA-T task actually rated 'temporally annoying artifacts'. To do so, we studied the feedback of observers on the experiment.

At the end of each test session, we asked the participant to rate the difficulty he had to perform the task he was assigned for each video. Observers that participated in QA-T task had in addition to rate the difficulty they had for separating the temporal quality from the global quality. The difficulty to perform the QA-G task, QA-T task and separating temporal from global in the QA-T task are displayed in Figure 7 averaged over all participants and with the corresponding confidence intervals[†]. We used two statistical inference tests to assess the equality of the participants' answers: Mann-Whitney's U test for the means equality and Brown-Forsythe test for the variances. These tests showed that, for each video, the assumptions of equal mean and variance for the QA-G and the QA-T tasks were not rejected (with p - values above 0,25 in both cases).

This means that we did not find any significant difference on the perceived difficulty to rate global or temporal quality of videos. We also compared the difficulty observers had to separate temporal and global quality during QA-T to the difficulty observers to assess global quality during QA-G. Tests showed that we did not statistically rejected the equality of means for each video except for the video 'GoosePark' (for 'GoosePark' $p = 0,03$, for all other videos p are above 0,19). Moreover, we did not statistically rejected the equality of variances for each video except for the video 'Tulip' (for 'Tulip' $p = 0,01$, for all other videos p are above 0,3). It indicates that the perceived difficulty and the variations between observers of both questions do not present a statistically significant difference.

[†]The ratings attributed to the BlueSpruce video are quite lower than those for other sequences. We suppose this originates in the content of the sequence: a traveling on a spruce with snow falling on it that occupies the screen really homogeneously. We also believe we chose too high bitrates because we overestimated the impact produced by the present defects on the perceived quality

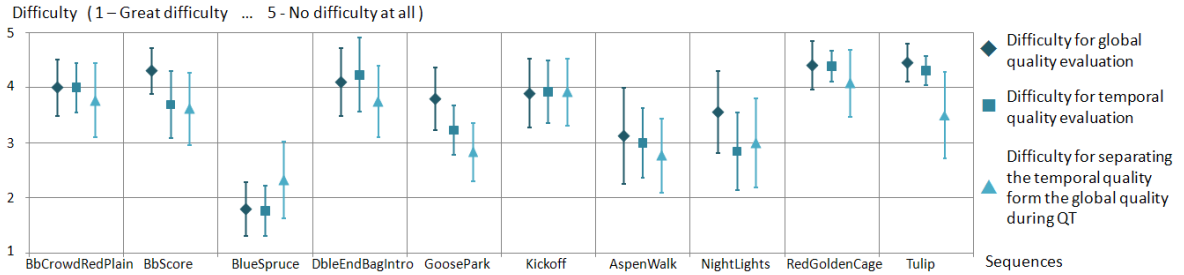


Figure 7: Evaluations of the difficulty observers had executing the task they were assigned for each video

Those elements seem to indicate that our experiment protocol is valid and that observers performing QA-T managed to rate the temporal quality separately from the global quality. We therefore consider in the rest of this analysis that instructions have been correctly executed. However, like all data in subjective quality assessment, the measurements are based on the perception of the participants.

3.2 Measure of the resemblance between locations fixated

During the free-viewing task (FT), no instruction was given to observers and they ignored the focus of the experiment so FT represents the 'natural' way to watch a video and it therefore acts as a reference.

To compare the locations fixated during each quality assessment (QA) task to those fixated during FT, we computed two indicators: the dispersion of locations fixated by all participants and the Area Under Curve (AUC) using a receiver operating characteristic (ROC) metric. AUC measures the differences between locations fixated during FT and locations fixated during the two quality assessment tasks.

Dispersion represents the variability between the different observers: the lower it is, the closer the locations gazed at by participants were. It is computed as the average distance between the positions gazed at by all participants.

AUC measures the resemblance of a set of fixated locations to a reference. As the free-viewing task represents our reference, we applied AUC to evaluate how much fixations from both QA tasks are similar to fixations from FT. Figure 8 illustrates the computation steps for the Area Under Curve (AUC) indicator. We first built a reference Visual Attention Density Map ($VA_{FTCurrent}$) from the locations fixated during FT. We applied isotropic Gaussian filters with a standard deviation of 1° centered on each fixation to create the Visual Attention Density Map ($VA_{FTCurrent}$). The degree of similarity between location fixated from QA-G or QA-T and this reference $VA_{FTCurrent}$ is then assessed by computing a receiver operating characteristic (ROC) metric (as explained by Le Meur et al.¹⁶). The ROC curve represents the True Positive rate as a function of the False Positive rate and the AUC is the area under this curve. The true positives are the locations fixated during QA-G that are also fixated during FT. The false positives are locations fixated during QA-G for a frame of another video at another quality level (the frame number, video and quality level are randomly picked) that are also fixated during FT. The AUC values are between 0,5 and 1. The more similar the fixated locations are to the reference, the closer the AUC will be to 1; random locations on the contrary produce an AUC of 0,5.

3.3 Influence of quality on the eye movements: fixation locations and durations

To study the influence of quality over the chosen indicators, we present their variations as a function of the quality level. The *Orgn* quality level designates the original quality of the videos (the perfect quality). The terms *VBR* and *CBR* designates the two compression configurations, respectively Constant Bitrate and Variable Bitrate. The numbers associated to those configurations indicates the compression level, 1 being the lowest compression (highest quality) and 4 the highest compression (lowest quality).

Dispersion

To study how widely the fixated locations by the observers are spread during each task, we computed the dispersion between locations fixated. As previously said, for each video frame and for each observer, we computed the median position of the locations gazed at. We then computed the dispersion (the Euclidean distances between one point and all the others points) between those median positions of all observers for each frame. Then, the dispersion values are averaged over all

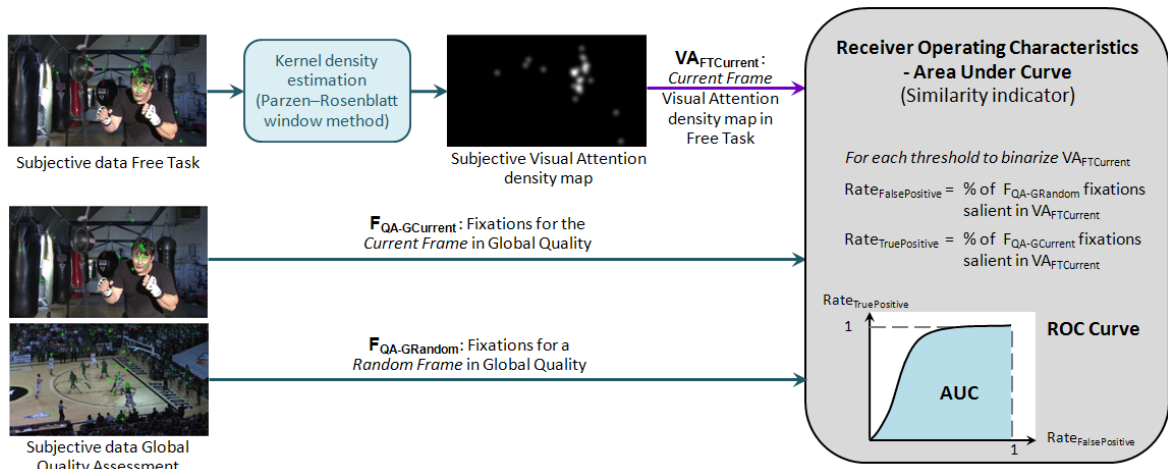


Figure 8: Computation process of a similarity indicator: the Receiver Operating Characteristics (ROC) - Area Under Curve (AUC). In this example it evaluates the similarity of locations fixated during QA-G with locations fixated during FT (serving as a reference)

the frames of the video. The dispersions for the three tasks are shown in Figure 9 as a function of quality level. The lower dispersion is, the closer the locations that observers fixated are.

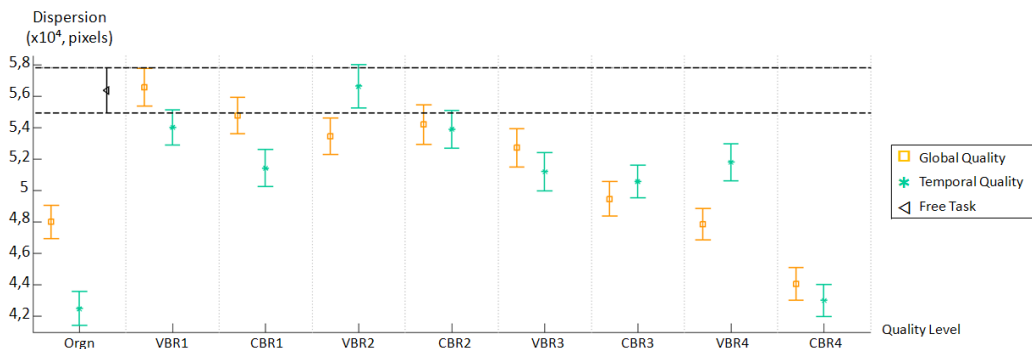


Figure 9: Dispersion between the locations that participants fixated as a function of quality level for the three tasks (Global Quality, Temporal Quality and Free Task). Note that FT value exists only for the original quality so we drew the corresponding confidence interval for all quality levels

Compared to the dispersion values for FT, the dispersion values for both quality tasks (QA-G and QA-T) are much lower when the quality of the displayed video are either perfect (*Orgn*) or low (*CBR3*, *VBR3*, *CBR4*, *VBR4*). It means that observers looked at closer locations when assessing the quality of perfect or low quality videos than when watching without instruction videos of perfect quality. The difference is smaller for the other quality levels.

AUC

To study the resemblance of the participants' fixations during each QA task and the free task, we computed the AUC indicator, shown in Figure 10 as a function of quality level. The AUC for Global Quality/Free Task measures the resemblance of the locations fixated during QA-G compared to the locations fixated during FT used as reference. The higher an AUC value is, the higher the resemblance is.

The AUC values for all quality levels are considerably higher for the FT/FT AUC than for QA-G/FT AUC or QA-T/FT AUC. Those results concord with the conclusions from Ninassi et al. or LeMeur et al. that observers performing a QA task do not gaze at the same locations as during a free task.^{1,8}

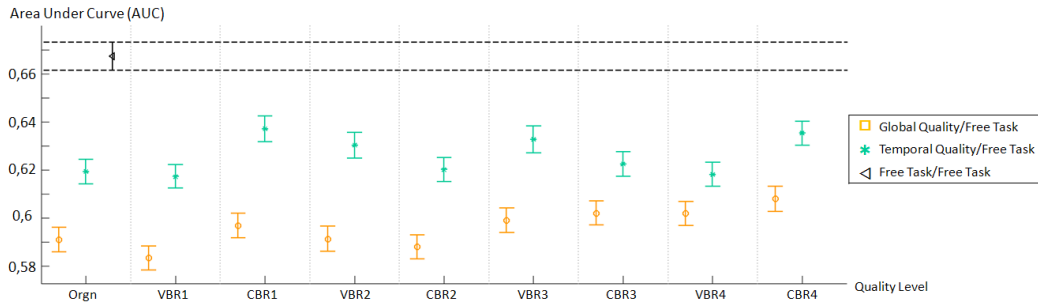


Figure 10: AUC indicator of resemblance of the locations fixated by the participants between QA-G and FT and between QA-T and FT as a function of quality level. Note that FT value exists only for the original quality so we drew the corresponding confidence interval for all quality levels. The resemblance of the positions fixated by one half of the participants in FT and the other half in FT acts as a reference for the natural eye movements

This difference is smaller for QA-T than for QA-G, suggesting that participants look at locations closer to the FT ones when they rate temporal quality than when they rate global quality.

The quality level of the video displayed has no visible influence on the resemblance of fixations from a quality rating task to the FT reference.

Fixation duration

To further characterize the way participants look at videos, we computed the average fixation durations for the three tasks. The fixation duration, averaged over all participants and all sequences, is shown for each task in Figure 11 as a function of quality level.

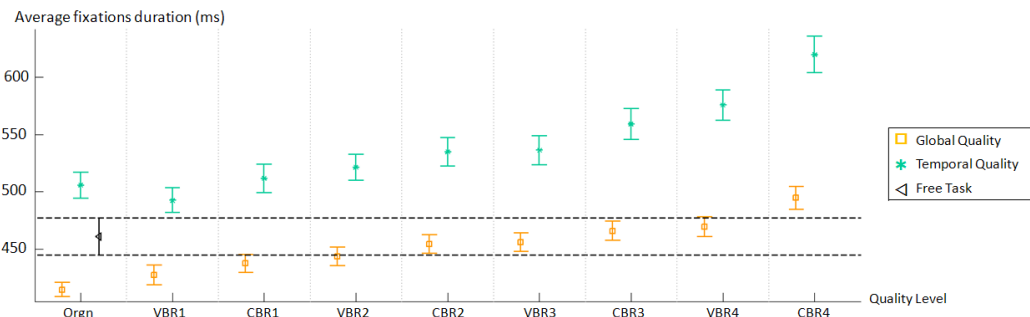


Figure 11: Duration of visual fixations for each of the 3 tasks as a function of quality level, averaged over all participants and all sequences. Note that FT value exists only for the original quality so we drew the corresponding confidence interval for all quality levels.

The average fixation duration of observers performing a quality assessment strongly depend on the quality level of the displayed video. The lower the quality level of the videos displayed is, the longer observers stare at objects.

Observers rating temporal quality clearly stare at objects longer than during free-viewing or global quality rating.

3.4 Evolution of the parameters over time

To study the evolution of eye movements over time, we computed the evolution as a function of time of the three indicators presented above: dispersion of the participants' gazes, resemblance of locations fixated during QA tasks to the FT reference (through AUC) and fixation duration. The results for each task averaged over all sequences and all quality levels are visible in Figure 12 for the dispersion. The fixations duration for each task averaged over all sequences for the VB4 quality is visible in Figure 13. We are analyzing fixations duration for the VB4 quality level because it is the quality level for which the the average values for each tasks differ the most (cf Figure 11).

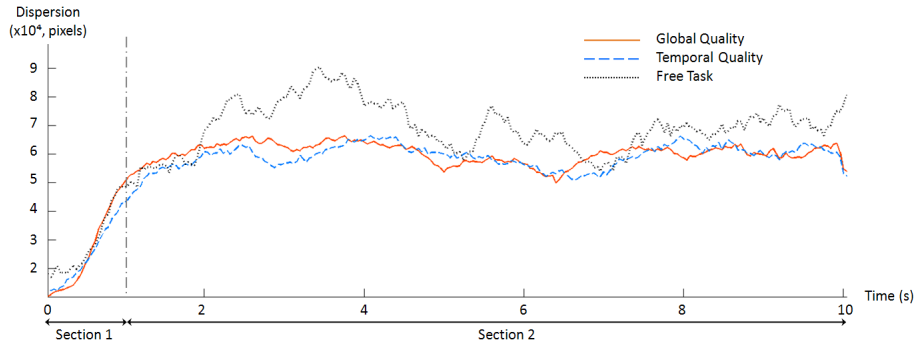


Figure 12: Dispersion of the locations fixated by the participants as a function of time for each task (Global Quality, Temporal Quality and Free Task)

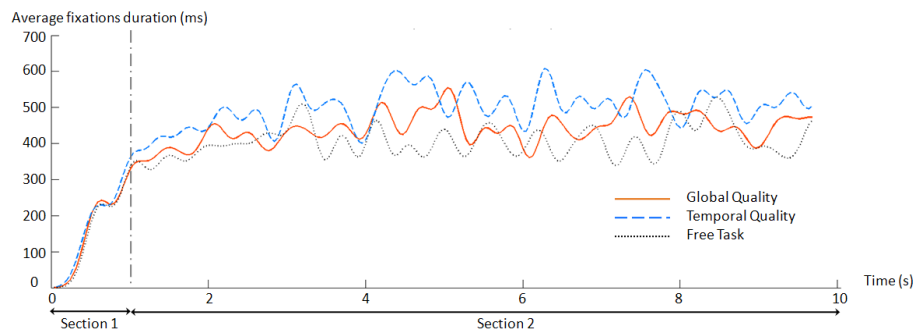


Figure 13: Duration of visual fixations for the VB4 quality for each of the 3 tasks as a function of time, averaged over all participants and all sequences

Taking only into account the evolution of the curves over time and not their respective level, it is clearly visible that the three curves contain two parts: during approximately 1 to 2 seconds, the indicators are evolving a lot whereas they stabilize during the rest of the video sequences. As the results are similar for the evolution of AUC over time, we did not include the corresponding figure in this paper. This evolution is common to the free-viewing task as well as the quality rating ones, it suggests that the participants' attention is mainly attracted by the low-level features from the videos during the first 1 or 2 seconds and that the instructions only seem to influence the participants' eye movements afterwards.

4. CONCLUSION

The aim of this paper was to characterize the influence of temporal quality assessment task (QA-T) on eye-movements of observers compared to the influence of Global Quality assessment task (QA-G) and compared to a Free Task (FT). To collect ground truth data, we set up an experiment combining quality assessment and recording of eye-movements with an eye-tracker. We analyzed the recorded eye-movements through three indicators: dispersion of fixated locations for each task, ROC/AUC between one task and the FT reference and fixation duration for each task.

Results showed that the quality level of the displayed video had no visible effect on the locations that observers fixated during both quality assessment tasks (QA-G and QA-T). However the quality highly influenced the fixations duration, i.e. the time during which observers kept their gaze fixated on one object. For both quality assessment tasks, the lower the quality of the displayed video were, the longer the fixations lasted.

Another interesting finding was that the instruction given to participants had no noticeable influence over the studied indicators (AUC, dispersion or fixation duration) during the first 1 to 2 seconds. It suggests that during those first diffusion instants, low-level (signal) properties of the videos might be the main factor driving the eye-movements. The assigned task only seemed to affect those indicators during the last 8 to 9 seconds of the diffusion, whichever the considered task (FT, QA-G, or QA-T).

We finally determined two elements that differed according to the type of quality assessed (global or temporal). The average fixation duration during QA-G was similar to that of FT whereas the average fixation duration during QA-T was

much longer to that of FT. Conversely, the locations fixated during QA-T were resembling more closely to those fixated during FT than those fixated during QA-G.

REFERENCES

- [1] Ninassi, A., Le Meur, O., Le Callet, P., and Barba, D., “Considering temporal variations of spatial visual distortions in video quality assessment,” *IEEE Journal Of Selected Topics In Signal Processing : Special Issue On Visual Media Quality Assessment* **3**(2), 253–265 (2009).
- [2] Segall, C. and Katsaggelos, A., “Pre- and post-processing algorithms for compressed video enhancement,” *Signals, Systems and Computers, 2000. Conference Record of the Thirty-Fourth Asilomar Conference on* **2**, 1369–1373 (2000).
- [3] Coudoux, F., Gazalet, M., and Corlay, P., “A post-processor for reducing temporal busyness in low-bit-rate video applications,” *Signal Processing: Image Communication* **18**, 455–463 (July 2003).
- [4] Keimel, C., Oelbaum, T., and Diepold, K., “No-reference video quality evaluation for high-definition video,” *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* , 1145 –1148 (April 2009).
- [5] Xia, J., Shi, Y., Teunissen, K., and Heynderickx, I., “Perceivable artifacts in compressed video and their relation to video quality,” *Signal Processing: Image Communication* **24**(7), 548 – 556 (2009).
- [6] Larson, E. C., Vu, C. T., and Chandler, D. M., “Can visual fixation patterns improve image fidelity assessment?,” in [*IEEE International Conference on Image Processing (ICIP)*], 2572–2575 (2008).
- [7] Ninassi, A., *De la perception locale des distorsions de codage à l’appréciation globale de la qualité visuelle des images et vidéos. Apport de l’attention visuelle dans le jugement de qualité.*, PhD thesis, Université de Nantes (2009).
- [8] Le Meur, O., Ninassi, A., Le Callet, P., and Barba, D., “Do video coding impairments disturb the visual attention deployment?,” *Signal Processing Image Communication* , – (2010).
- [9] Redi, J., Liu, H., Zunino, R., and Heynderickx, I., “Interactions of visual attention and quality perception,” in [*IS&T/SPIE Electronic Imaging*], *Human Vision and Electronic Imaging XVI* (2011).
- [10] [http : //www.cdvl.org](http://www.cdvl.org), *Consumer Digital Video Library* (2010).
- [11] [http : //www.videolan.org/developers/x264.html](http://www.videolan.org/developers/x264.html), *Encodeur x264* (2003).
- [12] ITU-T, *Recommandation P910. Subjective video quality assessment methods for multimedia applications* (2008).
- [13] VQEG, *Tutorial on the objective perceptual assessment of video quality: Full reference television* (2005).
- [14] Jones, B. L. and McManus, P. R., “Graphic scaling of qualitative terms,” *SMPTE Journal* **95**(11), 1166–1171 (1986).
- [15] Reiter, U. and Korhonen, J., “Comparing apples and oranges: Subjective quality assessment of streamed video with different types of distortion,” in [*Quality of Multimedia Experience, QoMEx 2009. International Workshop on*], 127 – 132 (2009).
- [16] 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* , – (2010).