Measurement of compression-induced temporal artifacts in subjective and objective video quality assessment

Claire Mantel, Patricia Ladret, Thomas Kunlin

Measurement of compression-induced temporal artifacts in subjective and objective video quality assessment

Claire Mantel[a, b], Patricia Ladret[a] and Thomas Kunlin[b]

[a]GIPSA-Lab - Department of Signal and Images, Grenoble Institute of Technology - UMR CNRS 5216, 961 rue de la Houille Blanche, Grenoble - France;
[b]STMicroelectronics S.A., 12 Rue Jules Horowitz B.P. 217, Grenoble - France

ABSTRACT

Temporal pooling and temporal defects are the two differences between image and video quality assessment. Whereas temporal pooling has been the object of two recent studies, this paper focuses on the rarely addressed topic of compression-induced temporal artifacts, such as mosquito noise. To study temporal aspects in subjective quality assessment, we compared the perceived quality of two versions of a mosquito noise corrector: one purely spatial and the other spatio-temporal. We set up a paired-comparison experiment and choose videos whose compression mainly creates temporal artifacts. Results proved the existence of a purely temporal aspect in video quality perception. We investigate the correlation between subjective results from the experiment and three video metrics (VQM, MOVIE, VQEM), as well as two temporally-pooled image metrics (SSIM and PSNR). SSIM and PSNR metrics find the corrected sequences of better quality than the compressed ones but do not distinguish spatial and spatio-temporal processings. The confrontation of those results with the VQM and Movie objective metrics show that they do not account for this type of defects. A detailed study highlights that either they do not detect them or the response of their temporal component is masked by the one of their spatial components.

Keywords: Video quality, Temporal artifacts, Subjective experiment, Quality assessment, Temporal filtering

1. INTRODUCTION

The disparity between image and video quality metrics is located on two different levels. Firstly a new type of impairments, the temporal ones, arises only in videos. There are various kinds of temporal compression artifacts: motion jerkiness, motion trail, temporal fluctuations (on block-level or on pixel-level) and temporal scaling for scalable coding. Secondly, the metric needs a way to combine all the frame-level ratings into a unique grade for the sequence: the temporal pooling. Temporal pooling methods have not been thoroughly investigated because they seem to matter less than the measurement itself and because the impairments level is often approximated as constant over time. The global grade of many video quality metrics have been designed as the Minkowski summation of a frame-by-frame metric ([1–4]). It has however recently been the point of two different papers: one from Rimac-Drjle et al. in[5] and the other from Keimel and Diepold in.[6] In both papers the authors apply different pooling strategies to existing quality metrics and study their influence on the correlation of the metrics results with subjective ratings. They show that the temporal pooling method can enhance the efficiency of a metric up to 10%. There have been few studies on the perception of temporal noises and consequently few studies on how to measure them. In[7] Yang uses the SSD (sum of squared differences between two consecutive images) as a fluctuation metric to assess the temporal efficiency of their fluctuation reduction filter. They base themselves on the recommendation from JVT in[8] that does not contain any token justifying it. The SSD is to temporal quality what the MSE is to spatial quality: a completely signal-based, rough evaluation of the degradation.

The evolution of compression norms and displays sizes induces a parallel evolution in defects characteristics. For example the blocking distortion was notoriously the major MPEG2 impairment. Both its importance and nature have changed in h.264 due to the use of 4x4 blocks for transform computation and quantization and to the integration of deblocking filters in encoders. Blocks are now annoying mainly because they are not stable over

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

time. Two video quality metrics consider the temporal fluctuations of defects, each in its own way. In the VQA metric detailed in,[9] Ninassi et al. account for the temporal variations of spatial artifacts through the gradient of distortion level over time (computed after motion-compensation). The second metric is a No-Reference one: in,[10] Keimel at al. describe the quality of a video sequence with 6 features: *blur*, *blocking*, *activity*, *predictability*, *dblur* and *dblocking* (representing respectively blur and blocking variations between consecutive frames). During the calibration phase of the metric, a weight is attributed to each parameter via PCA to fit best subjective data. The *dblur* and *dblocking* parameters are given important weights, proving them to be among the three most important parameters to account for quality grades.

In this article we study temporal compression-induced noises through the Mosquito Noise (MN) compression defect. MN occupies a peculiar place in the field of compression noises as it is annoying mainly because of its temporal variation: it is not a major defect for still images ([11]). We investigated in[12] the subjective perception of Mosquito Noise (MN) by setting-up an experiment with videos presenting mainly MN and our MN-dedicated corrector, the TVIF. As the perception of temporal continuity and its role as to video quality are established, this paper focuses on the response of existing objective quality metrics to such temporal impairments.

To this aim we confront the subjective ratings obtained through our previous experiment with objective metrics. We used the Full-Reference metrics VQM ([13]) and Movie ([14]), the No-reference metric VQEM from Maalouf et al. described in,[15] as well as the combination of SSIM and PSNR image metrics with different temporal pooling methods (the average, the 10% worst grade and the average over the last 3s).

The experimental set-up is described in 2. The comments from the observers that affirms their perceiving mainly temporal artifacts is detailed in section 3. The relationship between subjective ratings from the experiment and objective metrics scores are analyzed in section 4. Finally a conclusion is presented in section 5.

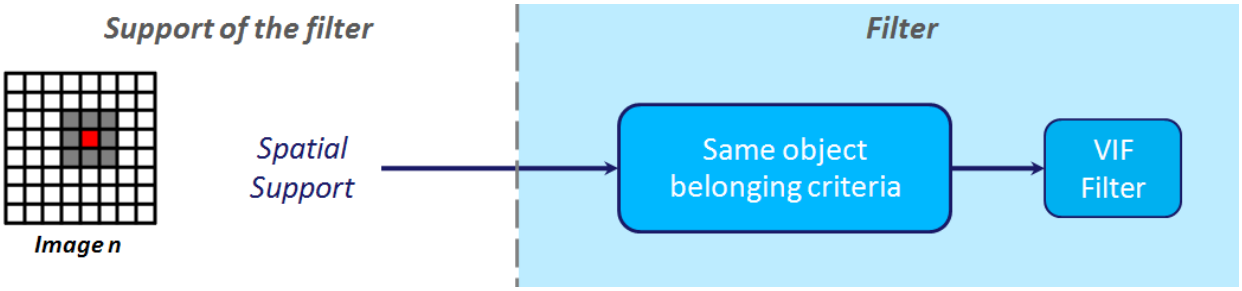
2. EXPERIMENTAL SET-UP

The purpose of this experiment was to evaluate two different versions of our MN-corrector: a spatial and a spatio-temporal. What interests us here is the difference between the two versions of the corrector, a more complete description of our MN-correction algorithm is available in.[11] The spatial version of our MN-corrector filters a pixel using pixels from its neighborhood in the current frame as the support. The spatio-temporal version differs from it by adding pixels from the previous frame to the support of the filter. They constitute the temporal part of the support. As shown in Figure 1, the filtering part is exactly the same in both versions, only the filter support changes.

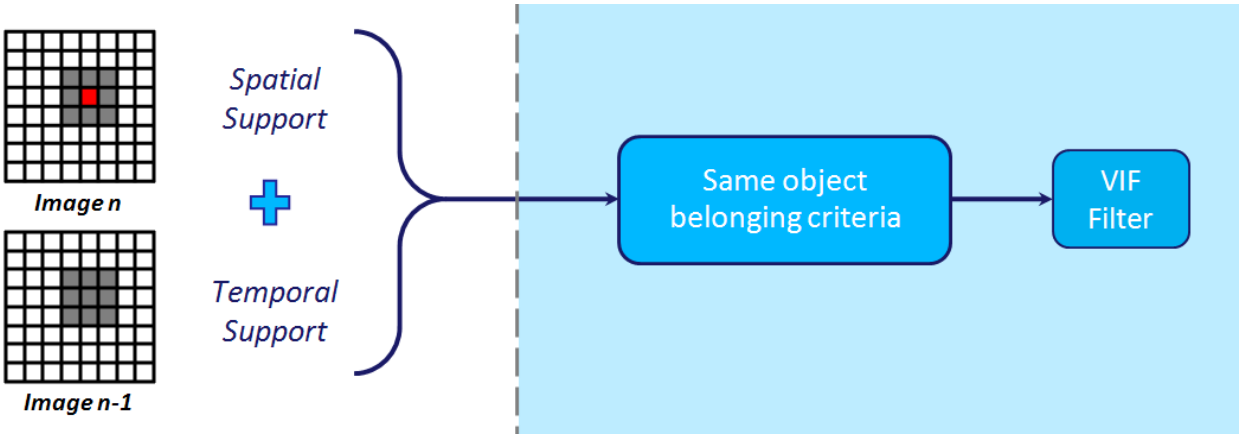
While designing the experiment, we took great care of enabling observers to assess temporal quality and chose videos that exhibit mostly temporal artifacts.

This experiment was realized using 6 different sequences that were available unprocessed. We encoded each sequence at the bitrate that, in our opinion, created the most temporal artifacts and for which they were the main defects. Each compressed sequence was then corrected with both versions of our corrector: spatial and spatio-temporal. During this experiment, we displayed those four different versions of each sequence: original, compressed, spatially corrected and spatio-temporally corrected (2).

We set-up a subjective paired-comparison experiment to obtain ground truth on the relative quality of those four versions of videos. Observers watched two different versions displayed side-by-side and were asked to rate which quality was better. They had to use a 7-points scale ranging from three degrees of preference towards the left video '*1 - Left is much better*', '*2 - Left is better*' and '*3 - Left is slightly better*' to a neutral answer '*4 - Left and Right are equivalent*' and three degrees of preference towards the right video (respectively graded 5, 6 and 7). They graded every possible combination of the four versions.



(a) Spatial version of our Mosquito Noise filter



(b) Spatio-Temporal version of our Mosquito Noise filter

Figure 1. Difference between the Spatial and the Spatio-Temporal versions of our MN corrector.

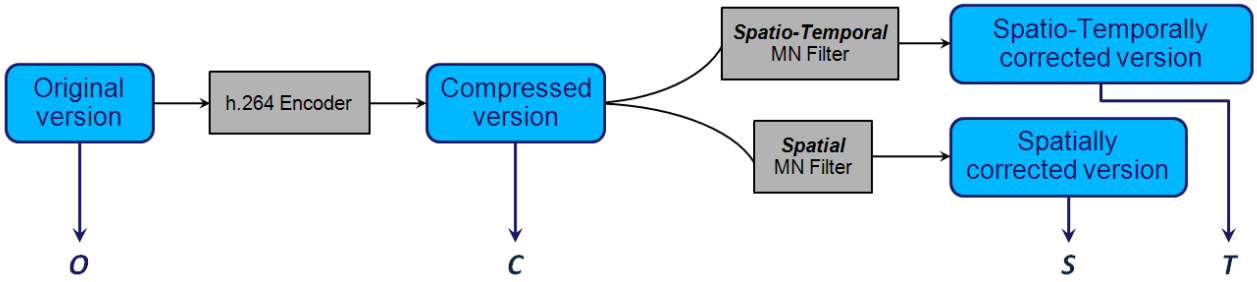


Figure 2. The 4 versions of sequences to be compared.

Figure 3 displays the subjective testing results averaged over all sequences and all the subjects with the associated confidence intervals. To analyze the results, we used those averages and t-tests to state if the difference to the central value is significant. The analysis of subjective quality ratings demonstrates the preference of observers for the spatio-temporal version of the algorithm over the purely spatial one. This ascertains importance of purely temporal aspects for quality perception.

3. THE OBSERVERS' PERCEPTION OF DEFECTS

At the end of each experiment session, a debriefing was organized by the organizer to get information about the observers' opinion on the experiment and their rating strategy. During this debriefing, observers were asked a series of questions about their opinion on the experiment and their rating strategy. They had to say for each video if they found it easy to rate, what type of defects they saw and where they saw them.

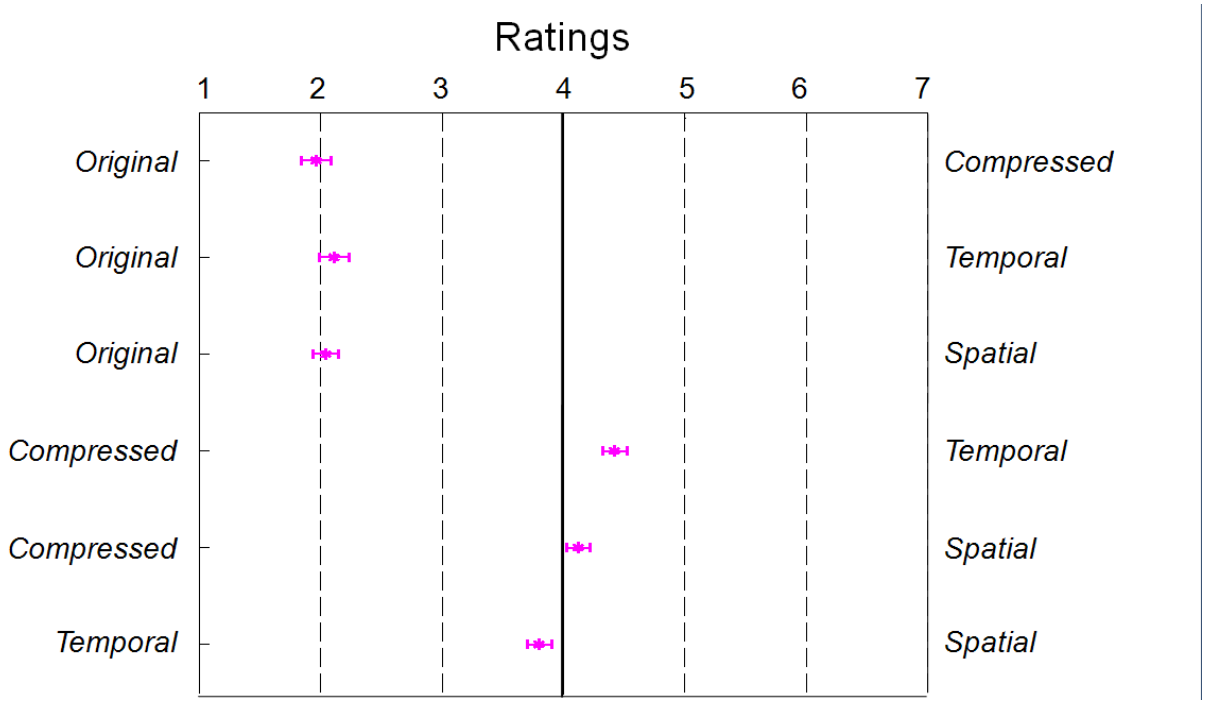


Figure 3. The results of the experiment averaged over all sequences and all the subjects with the associated confidence intervals. Each line corresponds to a comparison. When one of those averages per comparison is below the central value (4), it means that the first element of the comparison was graded better than the second. For example the O-C comparison star is close to the rating 2, which indicates that the original version (O) was clearly preferred to the compressed one (C).

Some subjects also said that they did not identify a precise defect but rated the comparison from a global impression. This tendency is stronger for the video with the most motion: *Aspen*: about a third of observers used their overall opinion to assess it.

Almost all observers said they saw impairments next to the contours. However, the type of defect differs from one subject to another and from a video to the other. Some defects are easily described by non-experts: everyone can say that an image is blurred, that motion is jerky or that colors are dribbling, which corresponds to the color bleeding artifact.On the contrary, there are several ways to describe MN: a 'swarming', a 'wobbling' or a 'crackling' next to the edges fit the definition.

We regrouped comments in 8 categories describing artifacts: *Pixelation*, *Blurring*, *Wobblingedges*, *Jerkiness*, *Motion trail*, *Color bleeding*, *Edge dripping* and *Flattened texture*.

We kept pixelation as a category because several observers used this term precisely but we believe that it corresponds to edge busyness impairments. Indeed, observers located it next to edges and pixelation is hardly realistic while talking about H264 compressed videos at bitrates between 1,5M and 5,6M in 720p seen from 1,5m.

The list of defects cited by at least 4 observers and the proportion of observers that perceived them is shown in Figure 4.

It clearly appears that the three main defects are *Pixelation*, *Wobbling edges* and *Blurring* for every sequence. The *Pixelation* and *Wobbling edges* impairments are what we were hoping for while choosing and processing the videos. Those comments therefore confirm that we managed to obtain videos whose major artifacts are temporal.

Sequence Name	Defects perceived by observers (in % of observers)
CrowdRun	**Pixelation** (50%), **Wobbling edges** (40%), Blurring (40%)
DucksTakeOff	**Wobbling edges** (44%), **Pixelation** (40%), Blurring (40%)
Shields	**Wobbling edges** (55%), Blurring (45%), Flattened texture (25%), Color bleeding (20%)
RushFieldsCrowd	**Wobbling edges** (55%), Blurring (45%), Color bleeding (20%), **Pixelation** (15%)
Tractor	**Wobbling edges** (25%), Blurring (25%), Color bleeding (25%), **Pixelation** (15%)
Aspen	**Wobbling edges** (38%), Blurring (35%), **Pixelation** (15%), Flattened texture (15%)

Figure 4. Types of defects perceived by the observers for each video

4. COMPARISON OF THE SUBJECTIVE DATA WITH OBJECTIVE METRICS

This section studies the relations of the objective metrics with the perception of temporal artifacts. Even though the previous section showed the importance of such quality features for global quality perception, they are clearly not the only or most prominent ones. Therefore this section is absolutely not an evaluation of the global efficiency of those metrics but only of their response towards temporal defects. Section 4.1 present the analysis process, while section 4.2 interprets the correlation results for all sequences at once and entering the specificity of each video sequence.

4.1 Correlation Computation

The aim here is not to study the global efficiency of the metrics so the correlation between the totality of subjective and objective ratings is useless. The data are to assess whether the metrics render the quality variation perceived by observers between spatial and spatio-temporal filtering.

To analyze the response of metrics to temporal processing, the differences between the sequences must be surpassed. In order for them not to interfere with the results, the grades are analyzed for each video separately.

Given the small number of data (6 samples per sequence only) to study the correlation, we chose the Spearman rank-order correlation coefficient (or Spearman's ρ) to quantify the association between subjective and objective ratings. This robust test informs us on any monotonous relation between the two sets of ratings. This way, we avoid making any hypothesis on the distribution or the regression model to use and there is a threshold specified in the Spearman table for tests with 6 samples only. It also validates associating the subjective rating of the CS comparison to $M(S) - M(C)$, i.e. the difference between the objective grades of the compressed and the spatial video. Indeed the Spearman rank correlation does not take into account the amplitude of this difference but only its sign.
The Spearman rank-order correlation coefficients measuring the dependence between the subjective and objective ratings are shown for video quality metrics in Table 1. The order in which each objective metric or parameter sorts the different versions out is also presented. As VQM and Movie also outputs the spatial and temporal quality indices, we computed the corresponding ROCC as well.

The threshold for the Spearman test for 6 samples is 0,886, meaning that every correlation coefficient above this value indicates a 95% probability of correlation. Even though the opposite decision (i.e. stating that the data are uncorrelated) can unfortunately not be evaluated statistically, the coefficient levels can still be interpreted one relatively to another.

Sequence Name	VQM			Movie			VQEM
	Global	Spatial	Temporal	Global	Spatial	Temporal	Global
CrowdRun	0,543	0,543	**1**	0,543	0,543	**1**	**1**
	C>S>T	*C>S>T*	**T>S>C**	*C>S>T*	*C>S>T*	**T>S>C**	**T>S>C**
DucksTakeOff	0,714	0,714	**0,943**	0,6	0,6	0,714	**0,943**
	S>C>T	*S>C>T*	**T>S>C**	*C>S>T*	*C>S>T*	*S>C>T*	**T>S>C**
Shields	0,543	0,543	**1**	0,543	0,543	0,543	**1**
	C>S>T	*C>S>T*	**T>S>C**	*C>S>T*	*C>S>T*	*C>S>T*	**T>S>C**
RushFieldCuts	0,6	0,6	**0,943**	0,6	0,714	0,6	0,829
	C>S>T	*C>S>T*	**T>S>C**	*C>S>T*	*S>C>T*	*C>S>T*	**T>S>C**
Tractor	0,686	0,686	**0,914**	0,686	0,686	0,686	**0,914**
	C>S>T	*C>S>T*	**T>S>C**	*S>T>C*	*S>T>C*	*S>T>C*	**T>S>C**
Aspen	0,6	0,6	**0,943**	0,6	0,6	0,6	**0,943**
	C>S>T	*C>S>T*	**T>S>C**	*S>T>C*	*C>S>T*	*C>S>T*	**T>S>C**

Table 1. The Spearman rank-order correlation coefficient between subjective and objective quality measurements for each video and the ranking of the compressed, spatially corrected and spatio-temporally corrected versions (respectively C, S and T). The digits written with bold lettering corresponds to coefficients above the significance threshold for six samples: 0,886 (meaning that every correlation coefficient above this value indicates a 95% probability of correlation).

4.2 The metric perception of spatial and spatio-temporal filtering

The coarser point of view is the analysis of global grades for each metrics (shaded in table 1). Without a single correlation coefficient above 0.71, the two Full-Reference metrics are clearly indicating that the metrics do not tally with the subjective results. Indeed, the two metrics ranked the processed videos the other way round than observers: they rate the TVIF corrector as a degradation compared to the compressed one, the spatio-temporal one being even worse than the spatial. The global grades of VQM and Movie do not seem to account for temporal defects such as MN.

On the contrary, the VQEM ranks the different versions in the right order every time. The original method to mix spatial and temporal properties seems to measure temporal artifacts rather well. However this metric presents surprising results for the *Shields* and *RushFieldCuts* sequences where the quality of the original version is on a level with the other versions. This phenomenon is at its peak for *RushFieldCuts* where the metrics rates that the quality of the spatio-temporally corrected version is closer to the original than to the spatially corrected one. Since this sequence contains scene change, an explanation would be the sensitivity of their coherence parameter to scene change.

To complete this bold result, the availability of the temporal, spatial and global ratings for each metric allows assessing two kinds of efficiency: the ability of the temporal index to detect temporal defects and the proportion in which the global grade accounts for it. For VQM we computed the weighted sum of its 6 spatial parameters, i.e. every parameters but *'ct_ati_gain'*, to act as its spatial grade.

The temporal parameter of VQM obtains values above the Spearman threshold for every sequences, meaning that it is statistically dependent from the subjective results. The *'ct_ati_gain'* parameter actually attributes the right ranks to every versions of the sequences. As to the spatial grade, its performance is similar to that of the complete metric: rather low.
The same phenomenon is visible in the detail of the Movie metric for the CrowdRun video: the temporal index indicates perfectly the quality of each version relatively to one another, but this right information is masked in

the global grade.

In short, for every video for VQM and one for Movie, even though the temporal defects are well detected and measured by the temporal index, this does not reflect on the final rating. Even though the correct quality information is at their disposal it is not taken into account. The contradiction of global ratings with temporal ones must be caused by the manner in which the two pieces of information, spatial and temporal, are gathered to build the global grade. It seems that the temporal index is somehow masked by the spatial one.

Finally, the temporal index of Movie is significantly correlated to subjective results for one video out of six. For the others, the correlations are below 0.71. For all videos but *CrowdRun*, the temporal part of Movie evaluates the spatial correction results better than the spatio-temporal ones, which is rather surprising. Actually, the only difference between the two versions of the corrector can roughly be approximated by a temporal smoothing. So how can it possibly degrade the temporal quality of a sequence? We believe that the granularity of a motion-computation method cannot be sharp enough to assess artifacts such as MN, whose both spatial and temporal frequencies are pretty high. The problem thus roots in the current limitations of motion compensation methods: their accuracy is not perfect, and they are misfit to the characteristics of some temporal noises. Those which arise from moving areas and are located next to them, such as MN, present a detected orientation really close from the motion one, even though the speed is quite distinct. A second cause comes from the assumption made by Seshadrinathan et al. in[14] that temporal noises as *'mosquito noise and stationary area fluctuations'* recreate *'the visual appearance of motion'*. As shown in section 3, temporal noises are not described by observers as some kind of erroneous added motion but as objects the visual perception cannot stabilizes on. Hence, is it pertinent to really synthesize it as such or would it require another dedicated component?

To summarize, two tendencies emanate clearly from this table: firstly the temporal parameter of VQM properly detects the type of temporal artifacts studied here (temporal fluctuations and MN) whereas the one from Movie seem to confuse them with motion. Secondly even when the temporal parameter measures temporal defects (for every sequence for VQM and once for MOVIE), it does not reflect on the global note of the metric.

4.3 Image Metrics

To try and get more information about the objective assessment of the defects present in our compressed video and the way the correctors modify it, we used two image quality metrics: the PSNR and the SSIM presented by Wang et al. in.[16]

As shown in Sections 1 and previous work, the choice of a temporal pooling method has wrongly been regarded as secondary compared to the other elements of a video metric. We computed three different methods of temporal pooling: the average, the 10% worst grade and the average over the last 3s. The last two pooling methods can be justified through characteristics of the human visual perception: people tend to remember the worst part of what they watched, which explains the 10% worst grade and they remember best what they last saw, which grounds averaging over the last 3s.
The Spearman rank correlation coefficients and the rank attributed by each metric to all the versions of a sequence are displayed in Table 2. For both metrics the temporal pooling method that performs best is averaging over the 10% worst rates. Considering the specificity of the videos, this information cannot be extended to quality assessment in general.

The correlation coefficients are globally better than those obtained for VQM or Movie complete ratings. For all videos for SSIM and all but *Shields* for PSNR applying our corrector (spatial or temporal) is measured as an enhancement compared to the compressed sequence. It means that, contrary to the VQM and Movie assessments, the filtered versions are more similar to the original sequence than the compressed one, be it on a basic pixel by pixel gray-level difference or through the signal processing approach of structural similarity.

Sequence Name	PSNR			SSIM		
	Average	10% worst	Last 3s	Average	10% worst	Last 3s
CrowdRun	**1**	**1**	0,886	**1**	**1**	**1**
	T > S > C	**T > S > C**	$S > T > C$	**T > S > C**	**T > S > C**	**T > S > C**
Ducks Take Off	**0,943**	**0,943**	**0,943**	**0,943**	**0,943**	**0,943**
	T > S > C	**T > S > C**	**T > S > C**	**T > S > C**	**T > S > C**	**T > S > C**
Shields	0,6	0,543	0,6	0,6	0,886	0,886
	$S > C > T$	$C > S > T$	$S > C > T$	$S > C > T$	$S > T > C$	$S > T > C$
Rush Field Cuts	**0,943**	**0,943**	**0,943**	**0,943**	**0,943**	**0,943**
	$S > T > C$	$S > T > C$	$S > T > C$	**T > S > C**	$S > T > C$	**T > S > C**
Tractor	0,707	0,815	0,707	0,815	**0,924**	0,815
	$S > T > C$	$S > T > C$	$S > T > C$	$S > T > C$	**T > S > C**	$S > T > C$
Aspen	0,771	**0,943**	0,771	**0,943**	**0,943**	0,771
	$S > T > C$	**T > S > C**	$S > T > C$	**T > S > C**	**T > S > C**	$S > T > C$

Table 2. The Spearman rank-order correlation coefficient between subjective and objective quality measurements for each video and the ranking of the compressed, spatially corrected and spatio-temporally corrected versions (respectively C, S and T) for various pooling of image metrics PSNR and SSIM. Its significance threshold for six samples is 0,886

On the other hand, the spatial correction is evaluated as better than the spatio-temporal one in about half cases by SSIM and in more than half by PSNR. Those image quality metrics are not able to distinguish between spatial and spatio-temporal corrections, proving that their differences stands within their temporal properties.

We showed in[12] that observers perceive the spatio-temporally corrected sequence as better than the spatially corrected one. The combination of those two results clearly means that the enhancement between the two versions of our filter cannot be captured through spatial processing but that video quality metrics need a component dedicated to temporal artifacts.

5. CONCLUSION

The issue at stake in this paper was to determine to which extent the temporal properties of video are accounted for in objective quality assessment metrics.

To this aim we confronted the subjective results to three video metrics: the Full-Reference metrics VQM and Movie, the No-reference metric VQEM. We also compared the efficiency of combinations of SSIM and PSNR image metrics with different temporal pooling methods.
The evaluation of the sequences by the temporally pooled versions of PSNR and SSIM proves that the spatial and spatio-temporal corrections are closer to the original sequence than compressed one. The inability of the image metrics to distinguish between the qualities of the spatial and the spatio-temporal corrections proves that the only difference between those versions stands in their temporal properties.
The fact that the spatial correction is half the time assessed as better than the spatio-temporal one also indicates the necessity for a component dedicated to temporal impairments in video quality metrics.

Two different trends stands out from the correlation between the subjective experiment ratings and the Full-Reference video metrics results. The temporal index of the Movie metric does not seem to detect temporal fluctuation noises, such as Mosquito Noise. We impute this to the high spatial and temporal frequencies of the noise, higher than the granularity allowed by their use of motion computation.
The second trend is that even when temporal impairments are detected by the temporal components of the

metrics, as in all cases for VQM, it does not reflect on the final grade. This stresses the question of how the various parts of a metric are put together and the importance of accounting for temporal defects.
On the other hand, thanks to its coherence parameter, VQEM ranks perfectly the four versions while sometimes presenting difficulty to assess the original version quality, probably due to sensitivity to scene change.

We demonstrated that observers perceive purely temporal artifacts but that their detection and measurement are not yet mastered by video quality metrics. As there currently exist many more metrics than the two examined here, our aim is to broaden our study of temporal quality assessment by confronting a much wider range of metrics to the same subjective results.

REFERENCES

[1] den Branden Lambrecht, C. J. V. and Verscheure, O., "Perceptual quality measure using a spatiotemporal model of the human visual system," *Digital Video Compression: Algorithms and Technologies 1996* **2668**(1), 450–461, SPIE (1996).

[2] Watson, A. B., Hu, Q. J., III, J. F. M., and Mulligan, J. B., "Design and performance of a digital video quality metric," *Human Vision and Electronic Imaging IV* **3644**(1), 168–174, SPIE (1999).

[3] Winkler, S., "Perceptual distortion metric for digital color video," *Human Vision and Electronic Imaging IV* **3644**(1), 175–184, SPIE (1999).

[4] Li, Q. and Wang, Z., "Video quality assessment by incorporating a motion perception model," in [*International Conference on Image Processing, ICIP*], 173–176 (2007).

[5] Rimac-Drlje, S., Vranjes, M., and Zagar, D., "Influence of temporal pooling method on the objective video quality evaluation," *Broadband Multimedia Systems and Broadcasting, 2009. BMSB '09. IEEE International Symposium on* , 1 –5 (may 2009).

[6] Keimel, C. and Diepold, K., "Improving the prediction accuracy of psnr by simple temporal pooling," in [*Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010*], (Jan. 2010).

[7] Yang, J. X. and Wu, H. R., "Robust filtering technique for reduction of temporal fluctuation in h.264 video sequences," *Circuits and Systems for Video Technology, IEEE Transactions on* **20**, 458 –462 (march 2010).

[8] Fan, X., Gao, W., Lu, Y., and Zhao, D., "Jvt-e070 flicking reduction in all intra frame coding," tech. rep., JVT (2002).

[9] 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).

[10] 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).

[11] Mantel, C., Ladret, P., and Kunlin, T., "A temporal mosquito noise corrector," in [*Quality of Multimedia Experience, QoMEx 2009. International Workshop on*], 244–249 (2009).

[12] Mantel, C., Kunlin, T., and Ladret, P., "The role of temporal aspects for quality assessment," in [*Quality of Multimedia Experience, QoMEx 2010. International Workshop on*], (2010).

[13] Pinson, M. H. and Wolf, S., "A new standardized method for objectively measuring video quality," *IEEE Transactions on broadcasting* **50**(3), 312–322 (2004).

[14] Seshadrinathan, K. and Bovik, A., "Motion tuned spatio-temporal quality assessment of natural videos," *Image Processing, IEEE Transactions on* **19**, 335 –350 (feb. 2010).

[15] [*A no-reference color video quality metric based on a 3D multispectral wavelet transform*] (jun. 2010).

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