Annoyance Models for Videos with Spatio-Temporal Artifacts

Alexandre F. Silva*, Mylene C. Q. Farias† and Judith A. Redi‡
*Department of Electrical Engineering
†Department of Computer Science
University of Brasilia, Brasilia, Brazil
Email: *alexandrefieno@gmail.com, †mylene@ieee.org
‡Multimedia Computing Dept. Intelligent Systems
Delft University of Technology, Delft, Netherlands
Email: J.A.Redi@tudelft.nl

***Abstract*—Although compression and transmission artifacts are likely to appear simultaneously in digital videos, their annoyance has been traditionally studied and modeled in isolation. So, while blockiness, blurriness, and packet-loss metrics exist, hardly any attempt has been made at modeling their joint impact on visual perception. In this paper, we evaluate the perceptual impact of those three artifacts on video quality. Based on data from three different experiments, in which a pool of participants evaluated videos impaired with packet-loss, blurriness and blockiness (in isolation and in combination), we analyze how the different artifacts combine to produce annoyance and propose several models for predicting the annoyance of videos impaired with combinations of packet-loss, blurriness and blockiness.**

Keywords*—Visual quality perception, spatio-temporal artifacts, video quality, objective quality metrics.

I. Introduction

Objective quality metrics [1] are a core component of quality control loops in video delivery systems. They automatically detect and estimate visual impairment annoyance. But, their accuracy is related to the extent to which they properly model human visual perception processes. Because of this, their design is far from trivial. Most successful video quality metrics estimate impairment annoyance by comparing original and impaired videos [2]. Alternatives include artifact metrics [3], [4], which estimate the strength of individual artifacts and, then combine the artifact strengths to obtain an overall annoyance or quality model. The assumption here is that it is easier to detect individual artifact signals and estimate their strength because we 'know' their appearance and the type of process that generates them.

Interestingly, whereas much has been done on understanding and modeling the perception of single visual artifacts, little work has been devoted to studying and characterizing their joint appearance and the perception of their combinations [5]. Farias et al. [6], [7] studied the appearance, annoyance, and detectability of common digital video compression artifacts (in isolation or in combination) by measuring their strength and overall annoyance. When presented in combination and at a low strength, artifacts that would otherwise be clearly recognized were mistaken by others. Also, the presence of noise in videos seemed to decrease the perceived strength of other artifacts, while the addition of blurriness had the opposite effect. Moore et al. [8] investigated the relationships among visibility, content importance, annoyance, and strength of artifacts in digital videos, concluding that the artifacts' annoyance was tightly related to its visibility, but only weakly related to content. Huynh-Thu and Ghanbari [9] examined the impact of spatio-temporal artifacts in video and their mutual interactions. They verified that spatial degradations affect the perceived quality of temporal degradations (and vice-versa), but the contribution of spatial degradations to overall quality is greater than that of temporal degradations. These studies gave important contributions to the better understanding the visibility and annoyance of combinations of artifacts. However, their results are still rather scattered and no clear knowledge is available on how different spatial and temporal artifacts combine perceptually and whether their joint impact depends on the physical properties of the video.

In this work, our goal is to study the perceptual impact that combinations of spatio-temporal artifacts commonly found in digital video transmission (i.e., blockiness, blurriness, and packet-loss) have on annoyance. More specifically, we are interested in understanding the relationship between the artifacts' perceptual strength and their overall annoyance. We present the results of three psychophysical experiments in which we investigate the characteristics of these spatio-temporal artifacts (when presented in isolation or in combinations). Then, we test linear and non-linear annoyance models, with and without interaction terms. These models allow for a better analysis of the contribution of each artifact to the overall annoyance and of the interactions among the different artifacts. In this paper, we propose a set of accurate and psychophysically meaningful annoyance models that consist of combination functions of the artifact strengths. In addition, the 308 videos involved in the three experiments and their corresponding annoyance scores are available for download to the community as a contribution of this paper [10].

The paper is divided as follows. In Section II, we present the setup and methodology used in all three psychophysical experiments. Section III details the re-alignment procedure performed on the subjective scores gathered from these experiments. In Section IV, we describe the annoyance perceptual models and discuss their results. Conclusions are addressed in Section V.

II. Annoyance of artifact combinations

We performed three psychophysical experiments with the goal of understanding how spatial (blockiness and blurriness) and temporal (packet-loss) artifacts commonly encountered in digital video transmission interact with each other and contribute to the overall annoyance. In Experiment 1, subjects analyzed the annoyance of videos impaired by packet-loss in isolation. In Experiment 2, subjects rated the annoyance of videos impaired with blockiness and blurriness, in isolation or in combination. Finally, in Experiment 3, subjects scored videos impaired with all three artifacts in combinations. The three experiments shared identical experimental methodology, interface, protocol, and viewing conditions. The stimuli were different per experiment but derived from a common set of 7 original contents, as detailed below.

A. *Stimuli*

As original contents we used seven high definition videos, shown in Figure 1, with spatial resolution of 1280×720 and a temporal resolution of 50 frames per second (fps). The videos were all 10 seconds long and were chosen to span a good range of spatial and temporal activity distribution.

Fig. 1: Frames of videos top to bottom, left to right: Barbecue, Park Joy, Into Tree, Park Run, Romeo and Juliet, Cactus, and Basketball.

To be able to add artifacts individually and combine them arbitrarily, we used a system for generating artifacts [6]. This allowed higher control in artifact combination, visibility and strength, which would be impossible when using, for example, a H.264 codec. To generate blockiness for each video frame, we calculated the average value of each 8×8 block of the frame and of the 24×24 surrounding block, then added the difference between these two averages to the block. To generate blurriness, we used a simple low-pass filter according suggested by Recommendation P.930 [11]. To control the amount of blurriness, we can vary the filter sizes and the cut-off frequencies. In this work, we used a 5×5 moving average filter to generate blurriness. We generated test sequences with combinations of blockiness and blurriness by linearly combining the original video with blockiness and blurriness artifact signals in different proportions (i.e., 0.4, 0.6, and 0.8). To generate packet-loss artifacts, we first compressed the videos (possibly already impaired with blockiness and blurriness) at high compression rates, to avoid inserting additional artifacts. Then, Then, packets from the coded video bitstream were randomly deleted in different loss percentages (the higher the percentage, the lower the quality) and changed the interval between I-frames (time interval among artifacts).

B. *Methodology and Equipment*

The various test sequences were displayed on a Samsung LCD monitor of 23 inches (Sync Master XL2370HD) with resolution 1920 × 1080@60hz (FullHD 1080p). 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 1.937, 1.566, 1.908, and 1.172, respectively. We set a constant illumination of approximately 70 lux. Participants were kept at a fixed distance of 0.7 meters from the monitor using a chinrest. The experimental methodology used was the single-stimulus with hidden reference and a 100-point continuous-scale [12].

The participants were mostly graduate students from the authors' institutions. They were considered naive of most kinds of digital video defects and the associated terminology. No vision test was performed, but participants were asked to wear glasses or contact lenses if they needed them to watch TV. The experiment started after a brief oral introduction. First, participants performed a training stage that consisted of watching highly impaired and pristine sequences to get acquainted with the typical artifact combinations and strengths. The sequences presented during the training were not scored and were meant to be visual anchors (references) for the annoyance scoring. After the training, the actual scoring session started. Participants were asked to give an annoyance score to each test sequence. Sequences as annoying as those seen in the training session should be given a '100' annoyance score, sequences half as annoying a '50' annoyance score, and so on. To avoid fatigue, the session was divided into sub-sessions, between which participants could take a break for as long as they wanted to. All experimental sessions lasted between 45 and 60 minutes.

C. *Experiments*

Experiment 1 investigated the annoyance of packet-loss artifacts in isolation. 16 participants scored the annoyance of test sequences for which different percentages of deleted packets (PDP) were used (PDP = 0.7%, 2.6%, 4.3%, and 8.1%). We also varied the number of M frames between the I-frames to change the time interval among artifacts (M = 4, 8, and 12) [13], [14]. A total of 7 originals and 12 combinations were used, resulting in $12 \times 7 + 7 = 91$ test sequences.

Experiment 2 focused on blockiness and blurriness artifacts. 16 participants scored the annoyance of test sequences containing combinations of blockiness and blurriness at different strengths. We represent hereon the artifact strength combinations as a vector (bloc;blur), where 'bloc' is the blockiness strength and 'blur' is the blurriness strength. Combinations contained artifacts at 3 possible strengths (0.0, 0.4, and 0.6) in a full factorial design ($3^2 = 9$ combinations), including the unimpaired videos (0.0;0.0). Two further combinations, pure blockiness and pure blurriness at strength 0.8, i.e. (0.8;0) and (0;0.8), were also added to the set, resulting in $11 \times 7 = 77$ test sequences.

In **Experiment 3**, 23 participants rated the annoyance of test sequences containing different combinations of blockiness, blurriness, and packet-loss artifacts. Hereafter we represent the strength combinations as a vector (PDP;bloc;blur), with the same notation as above. Blockiness, blurriness, and packet-loss artifacts were combined at 3 different strengths: bloc $\in [0, 0.4, 0.6]$, blur $\in [0, 0.4, 0.6]$, and PDP $\in [0, 0.7\%, 8.1\%]$, resulting in $19 \times 7 + 7 = 140$ test sequences [15]. To avoid fatigue, the experimental session was divided in three sub-sessions, with two 10-minutes breaks in between.

III. Data preparation and alignment

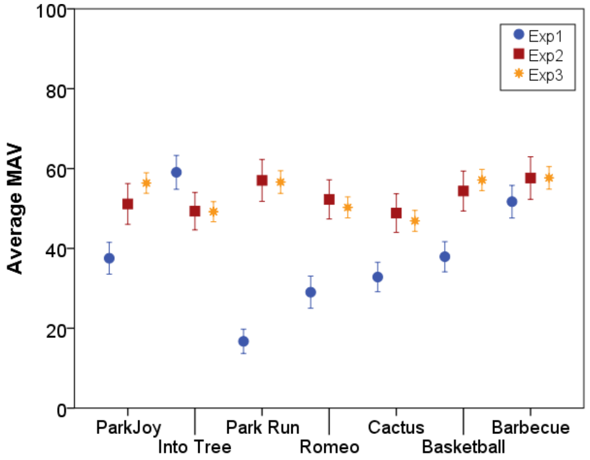
For each sequence in each experiment, we computed a mean annoyance value (MAV) as the average of the annoyance scores gathered from all participants [16]. However, these could not be readily used. Despite the extra care we put in keeping experimental conditions similar, we expected MAVs to be misaligned (i.e., not referring to the same underlying annoyance scale) across our three experiments.

Results gathered from experiments that comply to the same experimental methodology may still differ because of the differences in physical location, viewer expectations, or set of stimuli. [2]. Also, scores may suffer from context effects [17] as a result of the tendency of participants to use the entire scoring scale to evaluate the annoyance of the test stimuli. For example, in an experiment involving unimpaired or just slightly impaired stimulus, the latter may still get high annoyance values, as compared to the rest of the set. Similarly, in an experiment containing mildly to highly impaired stimuli, the mildly impaired ones may get unnaturally low annoyance (high quality) scores. As a result, the MAVs of these two experiments would not be comparable, being expressed on underlying scales covering a different range of annoyance.

Figure 2 shows the average MAVs across all versions of the same original content, for the three experiments. For Experiments 2 and 3, all original contents have similar average MAVs, roughly in the middle of the annoyance scale, as one would expect. For Experiment 1, though, the original contents ‘Park Run’, ‘Cactus’ and ‘Romeo’ have significantly lower MAVs than the rest, and their corresponding videos in Experiments 2 and 3. This suggests a misalignment across the scores given in different experiments. Thus, we opted for re-aligning the MAVs of the three experiments to the same scale before proceeding with the annoyance modeling phase.

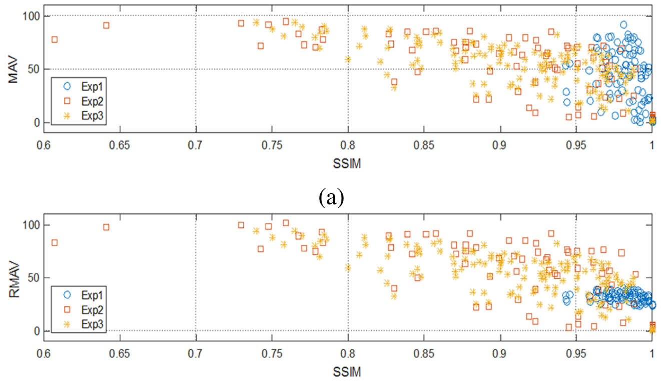
Pinson et al. proposed the iterative nested least squares algorithm (INLSA) to compare subjective scores from different experiments and convert them into a common scale [2], [18]. INLSA makes use of objective video quality metrics to re-align subjective scores from different experiments, by iteratively solving two least squares problems. The solution of the first problem homogenizes the scores from different experiments using a first-order correction. The solution of the second one solves the approximation of the homogenized scores using an objective quality metric. An iteration of these two least-squares problems provides a full mapping of the scores of the different experiments onto a common scale.

We re-aligned the scores (MAVs) of Experiments 1-3 with INLSA using Structural Similarity Index (SSIM, [19]) as the objective metric. First, we applied a linear function to map MAVs across the three experiments [18]. Then, we scaled the scores using Experiment 3 as the reference, since this is the experiment with the highest number of artifact combinations. Figure 3 (a) and (b) show the MAVs for the complete set of experiments before (a) and after (b) applying INLSA. Notice that in Figure 3 (a), MAVs of Experiment 1 seem to be clustered towards the top part of the SSIM scale, while spanning the entire annoyance range. This is not true for MAVs of the other two experiments. As shown in Figure 3 (b), after applying INLSA, the MAVs of Experiment 1 annoyance range is more commensurate to their SSIM range, what suggests that they can be merged with those of the other two experiments to be analyzed as a whole.



Error Bars: 95% CI

Fig. 2: Average values of MAV of all test sequences corresponding to each original video.



(b)

Fig. 3: MAVs versus SSIM for Experiments 1-3: (a) original MAVs and (b) MAVs re-aligned with INLSA (RMAV) [18].

IV. Annoyance Models

The goal of this work was to verify to what extent models combining objective artifact strength values (PDP, bloc, and blur) can predict perceived annoyance (expressed through re-aligned MAVs, hereafter referred to RMAVs) of videos impaired by multiple, overlapping artifacts. In addition, we wanted to study the extent to which each artifact (or combination thereof) would contribute to the overall annoyance perception. To do so, we tested a set of linear and non-linear models of these relationships. For ease of interpretation of the results (i.e., to obtain model coefficients of comparable magnitude), we imposed all artifact strength values to vary within the same range $[0, 1]$. PDP values were re-scaled within such range by assuming that a PDP greater than 10 would be unrealistic in real-world network conditions and considering 10 as the maximum possible value for PDP [20]. The normalized packet-loss strength was given by pdp = PDP$/10$.

Linear Models. We first tested simple linear models without interaction terms, to quantify the individual contributions of the different artifacts to annoyance. The basic, ideal model (eq. 1) assumes the overall annoyance to be the result of the linear combination of the single artifact strengths. According to this model, sequences without the presence of any artifact (i.e. (pdp;bloc;blur)= $(0; 0; 0)$) are perceived as not annoying ($PRA_{L1} = 0$). However, it may be the case that sequences without any blurriness, blockiness, or packet-loss artifacts are perceived as slightly annoying, due for example to impairments already present in the original content. In fact, the average RMAV of the unimpaired sequences was 13.47 in our experiments (2.38 for non-realigned MAVs). To account for this, we also tested a model with the addition of an intercept term δ (eq. 2). Having defined PRA_{L1} and PRA_{L2} as the annoyance predicted by the models without and with an intercept term, and pdp, bloc, and blur as the artifact strength parameters, we have:

$$PRA_{L1} = \alpha \cdot \text{pdp} + \beta \cdot \text{bloc} + \gamma \cdot \text{blur}, \tag{1}$$

$$PRA_{L2} = \alpha \cdot \text{pdp} + \beta \cdot \text{bloc} + \gamma \cdot \text{blur} + \delta. \tag{2}$$

Table I shows the result of the least-squares fit on RMAVs. For both models, the prediction performance, measured in terms of the Pearson correlation coefficient (PCC) and Spearman correlation coefficient (SCC), is also reported.

TABLE I: Fitting of linear models without interaction term to RMAV.

Models	δ	α	β	γ	**PCC**	**SCC**
PRA_{L1}		35.770	78.404	52.602	0.844	0.867
PRA_{L2}	18.170	19.768	61.499	35.698	0.850	0.870

Previous literature (e.g. [7]) showed that to accurately model the annoyance of overlapping artifacts, interaction terms should be taken into account. Artifacts, when combining, may mask or increase the perceptual strength of the others (interaction), thereby impacting on the overall annoyance feeling. To verify this, we tested a new linear model with interactions, (PRA_{L3}), defined as:

$$\begin{aligned} PRA_{L3} = {} & \alpha \cdot \text{pdp} + \beta \cdot \text{bloc} + \gamma \cdot \text{blur} + \\ & \rho_1 \cdot \text{pdp} \cdot \text{bloc} + \rho_2 \cdot \text{pdp} \cdot \text{blur} + \\ & \rho_3 \cdot \text{bloc} \cdot \text{blur} + \rho_4 \cdot \text{pdp} \cdot \text{bloc} \cdot \text{blur}. \end{aligned} \tag{3}$$

We also tested the same model with the addition of an intercept term δ, whose predictions we denote, hereafter, as PRA_{L4}. The results of the least squares fit of both models on RMAVs are presented in Tables II and III, respectively. In both tables, column 2 shows the estimated interaction coefficients and column 5 shows the corresponding p-values (based on t-test, two-tailed, $p < 0.05$ indicates significance of the term).

TABLE II: Fitting of the linear model with interactions (PRA_{L3}) for RMAVs.

Coef.	**Estimate**	**Std. Error**	**t-value**	$Pr(>\|t\|)$
α	57.064	2.784	20.494	$< 2e-16^a$
β	88.685	3.663	24.212	$< 2e-16^a$
γ	61.703	3.663	16.846	$< 2e-16^a$
ρ_1	-69.785	11.217	-6.222	$< 1.65e-09^a$
ρ_2	-63.363	11.217	-5.649	$< 3.74e-08^a$
ρ_3	-10.196	12.416	-0.821	0.4122
ρ_4	55.827	32.768	1.704	0.0895

^{a} Statistically significant at ($P < 0.05$) | PCC = 0.858, SCC = 0.885.

TABLE III: Fitting of the linear model with interactions and with an intercept coefficient (PRA_{L4}) for RMAVs.

Coef.	**Estimate**	**Std. Error**	**t-value**	$Pr(>\|t\|)$
δ	14.420	1.689	8.540	$6.83e-16^a$
α	33.757	3.702	9.118	$< 2e-16^a$
β	64.681	4.328	14.946	$< 2e-16^a$
γ	37.698	4.328	8.711	$< 2e-16^a$
ρ_1	-29.924	11.105	-2.695	0.00744^a
ρ_2	-23.503	11.105	-2.116	0.03514^a
ρ_3	28.800	12.053	2.390	0.01749^a
ρ_4	-11.286	30.470	-0.370	0.71134

^{a} Statistically significant at ($P < 0.05$) | PCC = 0.869, SCC = 0.884.

Non-Linear Models. It is reasonable to hypothesize that the proposed linear models, although fairly accurate, are unable to mimic complex non-linear interactions of the different artifacts [21]. Hence, below we report the performance of two types of non-linear models: a Minkowski metric and a Support Vector Regression (SVR).

Minkowski metrics have been shown to be useful to mimic interactions of spatial artifacts in standard definition videos [7]. We tested two models, again without (Eq. 4) and with (Eq. 5) an intercept term:

$$PA_{M1} = (\text{pdp}^m + \text{bloc}^m + \text{blu}^m)^{\frac{1}{m}}, \tag{4}$$

$$PA_{M2} = (\delta + \text{pdp}^m + \text{bloc}^m + \text{blu}^m)^{\frac{1}{m}}. \tag{5}$$

where m is the Minkowski power. Table 5 show the results of the least squares fit on RMAVs.

TABLE IV: Fitting of Minkowski models without and with intercept to RMAV.

Models	m	δ	**PCC**	**SCC**
PRA_{M1}	0.215		0.562	0.770
PRA_{M2}	0.397	3.424	0.770	0.744

Finally, we also tried a more black-box approach, where the model would not be defined upfront but learned directly from the data (i.e. our dataset of 308 videos). We used SVR to predict annoyance from the artifact strength data, as similar machine learning-based approaches have been shown to be suitable to model complex non-linear perceptual processes related to artifact annoyance [21]. We use 10-fold cross validation technique on the training set from SVM function in the R software. For that, dataset was randomly split into two non-overlapping sets: a training set (80%) and a testing set (20%). Our tests showed that using a radial kernel for the SVR was best performing to predict RMAVs. PCC and

SCC values obtained from the trained SVR were 0.9487 and 0.9514, respectively, which were the highest correlation values obtained in this work.

A. Discussion

For all linear models (PRA_{L1} to PRA_{L4}), the coefficients relative to the blockiness term (β) were found have the highest magnitude, implying that the presence and strength of blockiness has the biggest impact on the impairment annoyance. Blurriness was found to have the second highest impact and packet-loss the third. This ordering was maintained also when an intercept term was added, although the magnitudes of all coefficients were considerably reduced.

The majority of the interaction effects were statistically significant, except for the term relative to the interaction of blockiness and blurriness (ρ_3) in the PRA_{L3} model (without intercept). The second order coefficients (for the pairwise interaction terms) were mostly negative, what suggests that the perceptual effect of the combination of two artifacts was not a simple addition of the annoyance generated by the artifacts in isolation and, resulting in a somewhat lower annoyance. This may indicate that there are masking effects among artifacts, with artifacts mutually attenuating each other's strength and annoyance. The interaction coefficients with higher magnitude were those corresponding to the $pdp \cdot bloc$ and $pdp \cdot blur$ terms. This may indicate that packet-loss affected how blockiness and blurriness was perceived, somehow diminishing their visual impact. The third order coefficient, associated to the interaction of all artifacts (ρ_4), was non-significant, indicating that combinations of three artifacts did not contribute to the estimation of the overall annoyance (which was fully explained by the main effects and pairwise interaction terms). It should also be noted that the inclusion of these interaction terms in the models improved their predictive power (although not dramatically)

When an intercept constant was added to the linear models, their prediction power (measured by the PCC and SCC values between predicted annoyance scores and RMAVs) increased. This supports our hypothesis that, although unimpaired, the original contents (i.e. (pdp;bloc;blur)= $(0;0;0)$) showed in the experiments might have contained slight, pre-existing artifacts, judged as annoying by our participants. The value of δ obtained for the PRA_{L4} model, the most accurate of the linear ones was indeed 14.42, which is quite close the average RMAV of the unimpaired sequences (13.47).

For the non-linear Minkowski models, the predicting power of the fit models was lowest among all models. The addition of an intercept coefficient was beneficial also in this case, improving the predictive performance. The Minkowski power found ($m < 0.397$) was considerably lower than the values found by other authors [7], what indicates that the model is sensitive to small changes in artifact strengths.

Finally, the SVR-based approach has the best performance, in terms of prediction of RMAVs. This indicates that, although the linear models were quite informative with respect to the interactions among artifacts in generating visual annoyance, they might not be able to capture the complex non-linear processes that underlie human perception, in line with what was already proposed in literature [21]. Nevertheless, the SVR-based model lacks interpretability and further work is needed to unveil the relative importance of the different artifact strengths in determining the predicted annoyance.

Added value of re-aligning MAVs. It is worthwhile at this point to wonder whether the INLSA-based MAV realignment was beneficial. To verify this, we fit again all models, but on the non-realigned MAVs. For the sake of brevity, we report here only the main findings. Table V summarizes the accuracy of all models in predicting RMAVs and MAVs in terms of PCC and SCC values between subjective annoyance scores and their predictions.

TABLE V: Pearson (PCC) and Spearman (SCC) correlation coefficients between the different model predictions and the annoyance scores, when fit on re-aligned (RMAV) and non-realigned (MAV) scores

	prediction on RMAV		prediction on MAV	
Models	**PCC**	**SCC**	**PCC**	**SCC**
PRA_{L1}	0.844	0.867	0.726	0.721
PRA_{L2}	0.850	0.870	0.730	0.727
PRA_{L3}	0.858	0.885	0.797	0.771
PRA_{L4}	0.869	0.884	0.803	0.787
PRA_{M1}	0.562	0.770	0.472	0.652
PRA_{M2}	0.770	0.744	0.660	0.654
PRA_{SVR}	0.949	0.951	0.868	0.817

In all cases, the models fit on RMAVs obtained better performance than the models fit on MAVs. This supports our choice of re-aligning the data before fitting the models. It is worth mentioning that, whereas for the Minkowski models the coefficient estimates were mostly similar when fitting them on MAVs or RMAVs, for the linear models this was not the case. When fitting on MAVs, in all linear models the pdp coefficient α was higher than the blur one (γ), contrary to what we found when fitting to RMAVs. This may be due to the fact that, as shown in Figure 3, the MAVs of Experiment 1 sequences, which contained only packet-loss artifacts, were overestimated. As a consequence, the impact of packet-loss artifacts on annoyance was exaggerated.

V. Conclusions

In this paper, we evaluated the perceptual impact of combinations of blockiness, blurriness, and packet-loss on video quality. We performed three experiments in which videos impaired with these artifacts, in isolation and in combinations, were evaluated by a pool of subjects. We then analyzed how the different artifact strengths combine to produce annoyance and proposed several annoyance models, including linear models with and without interactions, Minkowski models, and a non-linear model based on SVR. The SVR-based model had the best performance, indicating that complex non-linear interactions between artifact appearances underlie below annoyance perception. Interactions were also observed in the linear models, notably suggesting that the overlap of multiple artifacts may generate masking effects, decreasing the overall annoyance perception. It should be noted that all models in this work have been based on known (and controlled) values of artifact strengths. Further work is needed to determine to what extent the annoyance scores of the videos are predictable when these strengths are unknown and their values need to be estimated using an artifact metric or an analysis of the (encoded or decoded) bitstream.

Acknowledgment

This work is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), by Fundação de Empreendimentos Científicos e Tecnológicos (Finatec) and by Nuffic-Netherlands.

References

[1] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, 2011.

[2] M. H. Pinson and S. Wolf, "An objective method for combining multiple subjective data sets," in *Visual Communications and Image Processing*. International Society for Optics and Photonics, 2003, pp. 583–592.

[3] H. Liu and I. Heynderickx, "A perceptually relevant no-reference blockiness metric based on local image characteristics," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 2, 2009.

[4] R. V. Babu, A. Perkis, and O. I. Hillestad, "Evaluation and monitoring of video quality for uma enabled video streaming systems," *Multimedia Tools and Applications*, vol. 37, no. 2, pp. 211–231, 2008.

[5] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: what does the future hold?" *Multimedia Tools and Applications*, vol. 51, no. 2, 2011.

[6] M. C. Q. Farias, J. M. Foley, and S. K. Mitra, "Detectability and annoyance of synthetic blocky, blurry, noisy, and ringing artifacts," *IEEE Trans. on Signal Processing*, vol. 55, pp. 2954–2964, 2007.

[7] M. C. Farias and S. K. Mitra, "Perceptual contributions of blocky, blurry, noisy, and ringing synthetic artifacts to overall annoyance," *Journal of Electronic Imaging*, vol. 21, no. 4, pp. 043 013–043 013, 2012.

[8] M. S. Moore, J. M. Foley, and S. K. Mitra, "Defect visibility and content importance: Effects on perceived impairment," *Image Communication*, vol. 19, pp. 185–203, Feb. 2004.

[9] Q. Huynh-Thu and M. Ghanbari, "Modelling of spatio-temporal interaction for video quality assessment," *Image Commun.*, vol. 25, no. 7, pp. 535–546, Aug. 2010.

[10] "Varium project video database," http://www.ene.unb.br/mylene/databases.htm, accessed: 2016-01-30.

[11] "Itu-t recommendation p.930: Principles of a reference impairment system for video," 1996.

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

[13] M. C. Q. Farias, I. Heynderickx, B. L. M. Espinoza, and J. A. Redi, "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.

[14] J. Redi, I. Heynderickx, M. C. Q. Farias, and B. Macchiavello, "On the impact of packet-loss impairments on visual attention mechanisms," 2013, pp. 1107–1110.

[15] A. F. Silva, M. C. Q. Farias, and J. A. Redi, "Assessing the influence of combinations of blockiness, blurriness, and packet loss impairments on visual attention deployment," in *IS&T/SPIE Electronic Imaging*, vol. 9394, 2015, pp. 93 940Z–93 940Z–11.

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

[17] P. Corriveau, C. Gojmerac, B. Hughes, and L. Stelmach, "All subjective scales are not created equal: The effects of context on different scales," *Signal processing*, vol. 77, no. 1, pp. 1–9, 1999.

[18] S. D. Voran, "Iterated nested least-squares algorithm for fitting multiple data sets," *NASA STI/Recon Technical Report N*, vol. 3, p. 12919, 2002.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[20] R. Haywood and X.-H. Peng, "On packet loss performance under varying network conditions with path diversity," in *Proceedings of the 2008 International Conference on Advanced Infocomm Technology*. ACM, 2008, p. 106.

[21] P. Gastaldo, R. Zunino, and J. Redi, "Supporting visual quality assessment with machine learning," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–15, 2013.