Detectability and Annoyance of Synthetic Blocky, Blurry, Noisy, and Ringing Artifacts

Mylène C. Q. Farias, *Member, IEEE*, John M. Foley, and Sanjit K. Mitra, *Life Member, IEEE*

***Abstract*—This paper presents the results of a series of four psychophysical experiments carried out to study the appearance, annoyance, and detectability of common digital video compression artifacts. The approach chosen in this paper was to use synthetic artifacts that look like "real" artifacts, yet are simpler, purer, and easier to describe. This approach allowed the control of the amplitude, distribution, and mixture of different types of artifacts. The algorithms for generating four of the most common types of artifacts: blockiness, blurriness, ringing, and noisiness, are described. The psychophysical experiments performed used video sequences containing different combinations of the generated synthetic artifacts. In these experiments, subjects were asked to detect any impairment and rate its annoyance. With the data gathered, the probability of detection and annoyance values were determined as a function of the total squared error. The results showed that "original video" (content) has a significant effect on both the detection threshold and the mid-annoyance parameter, while the "artifact signal type" does not. It was also found that these two parameters are highly correlated and linearly related.**

***Index Terms*—Artifacts, compression, MPEG, video processing, video quality assessment, video quality metrics.**

I. Introduction

To QUANTIFY the performance of a video communication system, it is important to have fast algorithms (objective video quality metrics) that give an estimate of the quality changes in a video at each of the communication system stages (acquisition, compression, transmission, or display). The measurement of the quality of a video always requires a direct or indirect comparison of the given (test) video with a reference (original) video. The quality of a video may decrease when impairments are introduced during capture, transmission, storage, and/or display, as well as by any image processing algorithm that may be applied along the way (e.g., compression, etc.). Impairments are defined as visible defects (flaws) and can be decomposed into a set of perceptual features called artifacts [1], [2].

Since, in most applications, humans are the ultimate "receivers" of the video material, the estimated quality of a video should correlate with the quality as perceived as human observers. In the past few years, a big effort in the scientific community has been devoted to the development of better video quality metrics that correlate well with the human perception of quality [3]–[6]. Nevertheless, little work has been done on studying and characterizing the individual artifacts (e.g., blockiness, blurriness, ringing, noisiness, etc.). We believe that a good understanding of the characteristics of these artifacts is an important step in the design of better objective video quality metrics. In this paper, we are particularly interested in understanding the visibility and annoyance of a set of important artifacts and their relationship with content. We are also interested in the relationship among different types of artifacts and their individual importance. This information is crucial in the development of post-processing algorithms for digital video receivers where there is limited processing capacity.

Measuring the *visibility* of the changes in a video is considered one way to assess the performance of a communication system. In fact, many of the visual models were developed using error detection threshold experimental data [3]. For a given impairment, there is an associated probability that the test subjects (human observers) will see it. This probability varies as the impairment signal strength varies. The detection threshold is commonly defined as the impairment signal strength (total squared error) that results in a probability of detection of 50%. One application of the threshold data is to provide a measure of visibility of the impairment that can improve the accuracy of a quality measure.

To avoid double-ended scales, many subjective tests measure *annoyance* rather than quality. Annoyance is a measure of how much the impairment bothers the viewer and it has been found to increase as the impairment signal strength increases and quality decreases. If necessary, the annoyance scores can be easily converted to quality scores [7]. Moore *et al.* showed that for compression artifacts annoyance is closely related to visibility [8]. Since many of the metrics estimate quality based on how visible impairments are, the knowledge of how annoyance and visibility threshold are related is very important to their design. Knowing the visibility thresholds of impairments makes it possible to estimate their *perceived* annoyance values [8].

The approach chosen in this paper is to use synthetic artifacts that look like "real" artifacts, yet are simpler, purer, and easier to describe [1], [9]. This approach is promising because of the degree of control it offers with respect to the amplitude, distribution, and mixture of different types of artifacts. In this study, our goals were to describe the appearance, visibility, and annoyance

Manuscript received August 27, 2005; revised August 23, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ioan Tabus. This work was supported in part by CAPES-Brazil, by the National Science Foundation under Grant CCR-0105404, and by the University of California MICRO under a grant, with matching supports from Philips Research Laboratories and Intel Corporation.

M. C. Q. Farias is with the Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, CEP 58109-970, PB, Brazil (e-mail: mylene@ieee.org).

J. M. Foley is with the Department of Psychology, University of California, Santa Barbara, CA 93106 USA (e-mail: foley@psych.ucsb.edu).

S. K. Mitra is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: mitra@ece.ucsb.edu).

Digital Object Identifier 10.1109/TSP.2007.893963

of four artifacts (blockiness, blurriness, ringing, and noisiness) that are the most salient artifacts found in digital videos. To this end, we performed four psychophysical experiments with video sequences containing synthetically generated artifacts. In these experiments, subjects were asked if they detected any impairment and, if so, to rate its annoyance.

This paper is divided as follows. In Section II, we describe the algorithms used to generate the synthetic artifacts. In Section III, we give the details about the experimental methodology. In Sections IV–VII, we present the performed experiments. Finally, in Section VIII, we present the conclusions.

II. Generation of Synthetic Artifacts

The ITU-T Recommendation P.930 proposes an adjustable video reference system that can be used in psychophysical experiments to measure the subjective quality of video [1]. This document gives definitions of different types of artifacts and descriptions of algorithms for generating them synthetically. According to it, the created synthetic artifacts must be relatively pure, and easily adjusted and combined to match the appearance of the full range of compression impairments. Also, the algorithms for generating them must be well defined in a way that the artifacts can be easily reproduced. We add the condition that the synthetic artifacts must produce psychometric and annoyance functions that are similar to those for compression artifacts.

In this paper, we propose an alternative system for generating synthetic artifacts based on the algorithms described in Recommendation P.930. Our algorithms satisfy all of the previous conditions and the created artifacts look like real artifacts. Our set of artifacts is composed of four artifacts (blurriness, noisiness, blockiness, and ringing) considered the most salient present in digital videos [2]. The set is not extensive and, in practice, further variations of each type of artifact may occur. Restricting the number of artifacts to four was necessary because experiments that estimate annoyance and visibility of artifacts require a large amount of data, a reasonable number of originals and about six strength levels for each artifact. In this section, we describe the proposed algorithms for the creation of synthetic blockiness, blurriness, ringing, and noisiness.

A. Blockiness

Recommendation P.930 suggests introducing blockiness only in regions where these artifacts would be more visible, i.e., smooth and moving areas [1]. Once these areas are identified, the artifacts are introduced by changing the luminance of the pertinent blocks. The algorithm proposed in this paper is simpler than the algorithm described in Recommendation P.930 since it does not require that specific areas of the video be identified. The algorithm is applied to the video uniformly and the visibility of the inserted artifacts depends on the characteristics of the area as occurs in compression.

The proposed algorithm takes into account not only the average of the block, but also the average of the surrounding blocks. The first step of the algorithm is to calculate the average of each 8×8 block of the frame and of the 24×24 surrounding block, with the current 8×8 block as its center. Next, the difference, $D(i,j)$, between these two averages is calculated. Then,

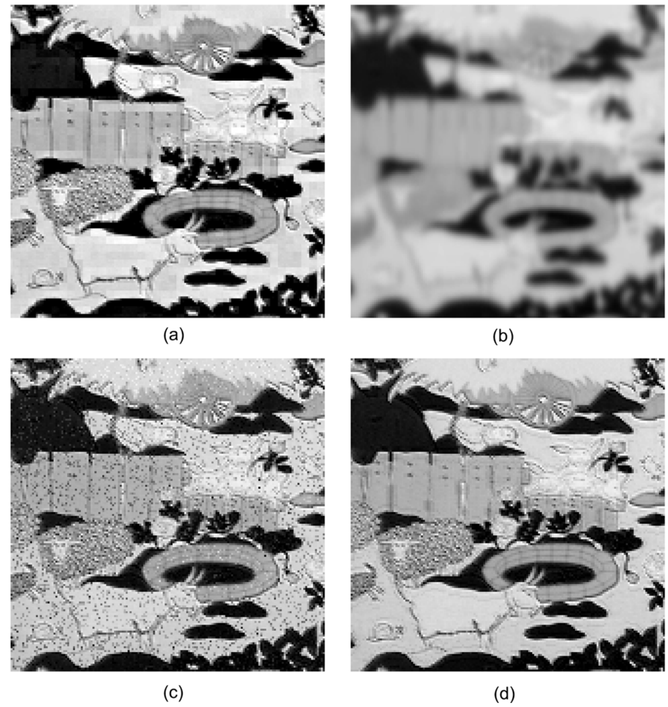


Fig. 1. Sample of artifacts generated synthetically in an area of a frame belonging to the video "Calendar": (a) blockiness; (b) blurriness; (c) noise; and (d) ringing.

to each block of the original frame we added the corresponding element of the difference matrix $D(i,j)$

$$Y(i,j) = X_0(i,j) + n \cdot D(i,j) \tag{1}$$

where X_0 is the original frame, Y is the frame with blockiness, n is a constant number, and i and j denote the spatial position of the pixel in the frame. The values of $D(i,j)$ were limited to avoid that the pixels become saturated. Before adding the blockiness, the average of the frame was corrected to avoid the artifacts becoming more visible than intended. To correct the average, we first calculated the average of the frame before introducing the artifacts and the average after introducing them. Then, we subtracted the latter from the former and added the difference to all pixels in the frame. The algorithm for generating synthetic blockiness can be easily modified to use different block sizes and to include spatial shifts. In Fig. 1(a), a sample of the synthetic blocky artifacts is shown in an area of the video "Calendar."

B. Blurriness

Recommendation P.930 suggests the generation of blurriness with the use of a simple low-pass filter [1]. To control the amount of blurriness, we can use different sizes of filters with different cutoff frequencies. Using a big range of filters increases dramatically the number of "types" of blurriness. Since we want to study four types of artifacts, it is not possible to also study different types of each artifact due to the limit on the number of videos that can be shown in a single experiment. For this reason, in this paper, we used only a simple 5×5 moving average filter to generate blurriness. In Fig. 1(b), a sample of the synthetic blurry artifacts is shown in an area of the video "Calendar."

C. Noisiness

The algorithm used here for creating noisiness is similar to the one proposed in Recommendation P.930 and consists of replacing the luminance value of pixels at random locations with a constrained random value (Gaussian distribution, zero mean, and variance equal to one) [1]. We changed the range of luminance values used by the recommendation to the range [10,120] in order to avoid making the artifact more visible than intended. We used a ratio of impaired/nonimpaired number of pixels in the frame equal to 0.1. The bigger this ratio is, the higher the level of noisiness present in the video frame. In Fig. 1(c), a sample of the synthetic noisy artifacts is shown in an area of the video "Calendar."

D. Ringing

Recommendation P.930 suggests generating ringing using a filter with ripples in the passband amplitude response [1]. The problem with this approach is that besides ringing, it also introduces blurriness and possibly noisiness. Our algorithm for generating ringing consists of a pair of delay complementary filters related through

$$H(z) + G(z) = \rho \cdot z^{-n_0} \tag{2}$$

where $H(z)$ and $G(z)$ are the transfer functions of N-tap high-pass and low-pass finite-impulse response (FIR) filters, respectively, [11]. For $\rho = 1$ and $n_0 = 0$ the output of our system in the z-domain is given by the following equation:

$$Y(z) = [H(z) + G(z)] \cdot X_0(z). \tag{3}$$

Thus, except for a shift, $Y(z)$ is equal to $X_0(z)$ given that the initial conditions of both filters are exactly the same. If we make the initial conditions different, a decaying noise is introduced in the first $N/2$ samples. Since ringing is only visible around edges, the algorithm is only applied to the pixels of the video corresponding to edges in both horizontal and vertical directions. The resulting effect is very similar to the ringing artifact found in compressed images, but without any blurriness or noisiness. In Fig. 1(d), a sample of the synthetic ringing artifacts is shown in an area of the video "Calendar."

III. Psychophysical Experiment Methodology

Four experiments (Experiments I–IV) were performed to gather information about the detection threshold, annoyance, and appearance of blockiness, blurriness, ringing, and noisiness. The methodologies used in these experiments are described in the following sections.

A. Test Sequences Generation

To generate the test sequences, we started with a set of six original video sequences of assumed high quality: "Bus," "Calendar," "Cheerleaders," "Football," "Flower," and "Hockey." They are all 5 s long and are in ITU-R BT.601 format (60 Hz, interlaced, 4:2:0 YCrCb format, 486×720). These videos are commonly used for video compression experiments and are publicly available. Representative frames of these originals are shown in Fig. 2.

Fig. 2. Sample frame of original videos "Bus," "Calendar," "Cheerleaders," "Flower," "Football," and "Hockey."

In this paper, we used an experimental paradigm in which impairments are restricted to an isolated region (defect zone) of the video clip for a short time interval [8], [9]. Each defect zone was 1 s long and did not occur during the first and last seconds of the video to avoid onsets and end effects. For the video "Flower," the defect zones corresponded to the areas containing the "Sky," "Houses," and "Garden." For the other videos, defect zones were rectangular strips (horizontal or vertical) taking approximately one-third of the frame. The advantage of isolating the impairments in defect zones is that the experimental task is greatly simplified. When the impairments are over all frames, it is frequently difficult for the subject to give one single annoyance score because the interactions between impairments and content frequently result in different levels of annoyances for different regions of the video. Also, since the content may vary considerably for different areas of the video, the isolation of the impairments makes it possible to analyze the effect of content on the visibility and annoyance of impairments.

The test sequences of each one of the four experiments had different types of impairments that consisted of combinations of up to four artifacts (blockiness, blurriness, ringing, and noisiness). The individual artifacts were generated as described in Section II and linearly combined to generate the resulting impairments. To add the resulting impairments to the defect zones we used a binary mask that had values equal to "1" for pixels inside the defect zone and "0" for pixels outside it, i.e., impairments were only added to pixels of the video where the mask M was equal to "1." The strength of the impairment is controlled by scaling the pixel-by-pixel differences between the video with impairments and the original video. To create a test sequence Y with up to L artifacts, we used the following expression:

$$Y(i,j,k) = X_0(i,j,k) + \sum_{l=1}^{L}[X_0(i,j,k) - X_l(i,j,k)]r_l \cdot M(i,j,k) \quad (4)$$

where i, j, are the spatial coordinates, k is the frame number, M is the defect zone binary mask, X_0 is the original video, X_l is the video containing only the ℓth artifact, and r_l is the relative strength parameter. We define the strength of the ℓth artifact as the difference between the video with this artifact (X_l) and the original video (X_0). To create a set of videos with varying strengths of this artifact, we multiplied this difference by different values of the parameter r_l, which is defined as the relative strength of the artifact $(0 < r_l \leq 1)$. With the help of a pilot study with a reduced number of subjects, the values of r_l were picked such that the produced impairments vary from hardly visible to very annoying. In general, $0 < \sum r_l < 1$, but, in some cases, we allowed $\sum r_l > 1$ making the impairments stronger. After the impairments are added to the originals, the borders of the defect zones are faded to avoid increasing the visibility.

Although the majority of psychophysical studies vary the defect strength by changing the bit rate and/or the codec implementation [10]–[12], some studies have controlled the defect strength by linearly changing its amplitude [8], [13]. In [13], the two methods were compared and the authors concluded that linear scaling can validly approximate the changes produced by varying the MPEG-2 bit-rate goal. We used linear scaling to control the strength of our impairments because it allowed us to combine as many or as few artifacts as needed at several strengths in order to obtain an estimate of their visibility threshold and mid-annoyance values. In other words, this method allows control of both the appearance and the strength of the impairment making it possible to measure the psychophysical characteristics of each type of artifact separately or combined [8], [9].

B. Apparatus

The videos are displayed using a subset of the PC cards normally provided with the Tektronix PQA-200 picture quality analyzer. The analog output is displayed on a 14-in Sony PVM-1343 monitor. A special-purpose program in Visual C++ is used to run the experiment and record the subjects' data. After each test sequence is shown, this program displays a series of questions on the computer monitor and records the subject's responses in a data file.

The experiments were performed in the Visual Perception Laboratory of the Psychology Department with the lights dimmed. The distance between the subject's eyes and the video monitor is four monitor screen heights. The video monitor is 20 cm tall resulting in a viewing distance of 80 cm. According to the ITU Recommendation BT.500 [12], four heights is the minimal distance allowed between the viewer and the monitor regardless of the size of the monitor.

C. Procedure

Our subjects were drawn from a pool of students in an introductory psychology course in the University of California Santa Barbara (UCSB). They were asked to wear glasses or contact lenses if they need them to watch TV. The experimental sessions lasted about 50 min and were run with one subject at a time. Each subject was seated straight ahead of the monitor, which is centered at or slightly below eye height for most subjects. A test session was broken into five stages. In the first stage, the subject was verbally given *instructions*. In the second stage, examples of original and highly impaired videos were shown to establish the range for the annoyance scale. The subject was instructed to assign an annoyance value of "100" to the most annoying of the example videos. In the third stage, the subject carried out *practice trials* to allow the responses to stabilize before the experimental trials. Then, on the fourth stage, *experimental trials* were performed. In the last stage, the subject was asked to provide qualitative descriptions of the impairments.

The *experiment trials* were performed with the set of test sequences presented in random order. The subject was instructed to search each video for impairments and to perform two or more of the following tasks.

- The *detection* task consisted of detecting a spatially and temporally localized impairment in the test sequence. After each test sequence was shown the subject was asked, "Did you see a defect or an impairment?" She/he was instructed to choose a "yes" or "no" answer.
- The *appearance* task consisted of judging the appearance of the detected impairment by choosing a set of classifiers to describe it. After each test sequence was shown the subject was asked, "How would you describe the impairmen?" To answer this question the subject was presented with a set of classifiers and instructed to pick as many as necessary to describe it.
- The *annoyance* task consisted of giving a numerical judgment of how annoying the detected impairment was. The subject was instructed to enter a positive numerical value indicating how annoying the impairment was after each test sequence was played. Any defect as annoying as the worst impairments in the training stage should be assigned "100," half as annoying "50," 10% as annoying "10," and so forth. Values greater than "100" could be assigned if they judged the impairment to be worse than the most annoying impairment in the training stage.

D. Statistical Data Analysis

Standard methods are used to analyze the data provided by the test subjects [12], [14]. The detection probability of an impairment is estimated by counting the number of subjects who detected this impairment and dividing it by the total number of subjects. The detection probability as a function of the $\log(\mathrm{TSE})$ (*psychometric function*) is fitted using the Weibull function [12], which has an S-shape that fits the data well and is defined as

$$P(E) = 1 - 2^{(E/E_T)^{\kappa}} \quad (5)$$

where $P(E)$ is the detection probability, E is the $\log(\mathrm{TSE})$ of the sequence, E_T is the 50% detection threshold, and κ is a constant that determines the steepness of the function.

The mean annoyance value (MAV) is calculated by averaging the annoyance scores over all observers for each test sequence. The MAV as a function of the $\log(\mathrm{TSE})$ (*annoyance function*) is fitted with the standard logistic function [12]

$$PA = \frac{100}{1 + \exp\left(-\frac{(E - E_{50})}{\eta}\right)} \tag{6}$$

where *PA* is the predicted annoyance. The parameter E_{50} (*mid-annoyance value*) translates the curve in the E-direction and η is inversely related to the steepness of the curve.

To analyze how the subjects described the impairments they saw (appearance task), we first determined the percentage of subjects who described the impairments for each of the classifiers. These percentages were then plotted versus the log total squared error, for all test videos. Since subjects could mark more than one classifier for each sequence, these percentages often added to more than 100.

To test the relationship between the detection threshold and the mid-annoyance values we used the Pearson correlation (r) [14]. To analyze the effects of the different types of impairments and the different videos on the annoyance and psychometric function parameters we have used analysis of variance (ANOVA) tests. The ANOVA test consists of a statistical procedure that can be used to determine significant effects in an experiment [14].

IV. Experiment I: Blocky–Blurry Artifacts

In Experiment I, 30 subjects (24 females and 6 males) performed *detection* and *annoyance* tasks. The goal of this experiment was to make an initial comparison among synthetic impairments and real compression impairments and test the feasibility of using synthetic artifacts. The test sequences consisted of combinations of synthetic blockiness and blurriness signals that were visually very similar to those observed in highly compressed MPEG-2 impairments. The results were compared with the results obtained in a previous experiment performed by Moore *et al.* on detecting MPEG-2 defects inserted in the same defect zones of these videos [15]. The methodology used by Moore is the same as the one used in this paper. The test sequences used in his experiment were generated by linearly combining original videos and MPEG-2 compressed videos at a bit rate of 1 Mb/s.

To generate the test sequences used for Experiment I, we first created videos with only one type of artifact signal using the algorithms described in the Section II. For each original, we created a sequence with blockiness $X_1 = X_{\mathrm{blocky}}$ and a sequence with blurriness $X_2 = X_{\mathrm{blurry}}$ Then, we created a video with equal proportions of blocky and blurry artifacts using (4) with $L = 2$ and $r_1 = r_2 = \rho (0.15 \leq \rho \leq 0.5)$. We selected a range of defect strength parameters so that the impairments looked to the experimenter and subjects the same as those in the videos used by Moore. The TSE values of these synthetic videos were lower than those in the matching MPEG-2 videos. We attribute this to the presence of other artifacts in the MPEG-2 videos [2]. In this experiment, we used five originals ("Bus," "Cheerleaders," "Football," "Flower," and "Hockey"), three defect zones, and six strength parameter values (ρ) resulting in a total of 95 test sequences (5 originals $\times$ 3 defect zones $\times$ 6 strength parameters $+$ 5 originals).

TABLE I
EXPERIMENT I—PARAMETERS FOR ANNOYANCE AND PSYCHOMETRIC FUNCTIONS DETERMINED BY THE METHOD OF LEAST SQUARES. COMPARISON OF PARAMETERS FROM EXPERIMENT I (SYNTHETIC) AND MOORE'S *et al.*'s EXPERIMENT (MPEG) [15]

	E_T		κ		E_{50}		η	
Sequence	SYNTH	MPEG	SYNTH	MPEG	SYNTH	MPEG	SYNTH	MPEG
Bus-Bottom	3.85	3.70	19.15	9.60	4.39	4.08	0.27	0.45
Bus-Middle	3.70	3.45	14.96	26.28	4.18	3.93	0.22	0.25
Bus-Top	3.27	3.13	10.18	12.71	-	-	-	-
Cheer-Bottom	3.27	3.70	19.71	13.03	4.18	4.11	0.28	0.33
Cheer-Middle	3.33	3.27	13.12	18.3	3.92	4.02	0.27	0.45
Cheer-Top	3.27	3.70	10.57	19.27	4.42	4.39	0.34	0.46
Flower-Garden	4.17	4.00	17.09	14.98	4.62	4.59	0.29	0.37
Flower-Houses	3.45	3.27	12.18	15.71	3.99	4.03	0.26	0.29
Flower-Sky	3.27	3.27	16.44	11.65	-	-	-	-
Foot-Left	-	-	-	-	3.87	3.79	0.38	0.28
Foot-Middle	3.13	3.13	15.77	22.02	3.68	3.42	0.25	0.25
Foot-Right	-	-	-	-	3.94	3.53	0.40	0.36
Hockey-Left	3.13	3.03	11.11	18.01	3.58	3.46	0.30	0.39
Hockey-Middle	3.03	2.78	19.79	17.78	3.55	3.37	0.25	0.30
Hockey-Right	-	-	-	-	3.55	3.45	0.28	0.27
Mean	3.47	3.41	15.01	16.61	3.99	3.86	0.29	0.34
t-test, P	0.281		0.414		0.007		0.043	
Correlation, r	0.831		-0.252		0.93		0.262	

TABLE II
EXPERIMENT II—BEST PARAMETERS FOR ANNOYANCE AND PSYCHOMETRIC FUNCTIONS DETERMINED BY THE METHOD OF LEAST SQUARES

	blocky				blurry				combined			
Original	E_T	κ	E_{50}	η	E_T	κ	E_{50}	η	E_T	κ	E_{50}	η
Bus	2.50	6.91	3.48	0.50	-	-	4.15	0.36	-	-	4.06	0.35
Cheer	3.03	10.01	3.92	0.51	-	-	3.83	0.50	3.03	9.03	3.83	0.37
Flower	-	-	3.38	0.47	-	-	4.15	0.28	3.13	8.99	3.91	0.32
Football	0.42	7.86	3.17	0.35	-	-	3.50	0.37	2.70	10.96	3.44	0.34
Hockey	-	-	-	-	-	-	3.45	0.45	2.50	5.75	3.30	0.43
Mean	$\overline{E_T} = 2.47;\ \overline{\kappa} = 8.50$						$\overline{E_{50}} = 3.68;\ \overline{\eta} = 0.40$					

Table I shows the parameters of best fit of the psychometric and annoyance functions from Experiment I. For comparison, we show the same parameters for MPEG impairments (data of Moore *et al.* [15]). The spaces with "-" in Table II correspond to test groups where a fit was not possible because more than 50% of subjects detected the weakest impairment. The correlation between the parameter values from the two experiments and the probability of the t-statistic under the null hypothesis P are shown in the last two lines of Table I. The mid-annoyance values and detection thresholds for both types of impairments are highly correlated. Regarding the small differences in the parameter values for both E_T and κ, we found that they are not statistically significant (t-test, two-tailed, $P > 0.05$). The small differences in parameters of the annoyance functions (η and E_{50}) for the synthetic and MPEG-2 artifacts are statistically significant (t-test, two-tailed, $P > 0.05$). We found that E_T and E_{50} are well correlated ($r = 0.6541$) and are related by the following equation: $E_{50} = 0.823 \cdot E_T + 1.23$. Figs. 3 and 4 show pairs of psychometric and annoyance functions for the videos "Cheerleaders" and "Football" (with defect zones in the bottom and middle areas, respectively). Each graph shows two

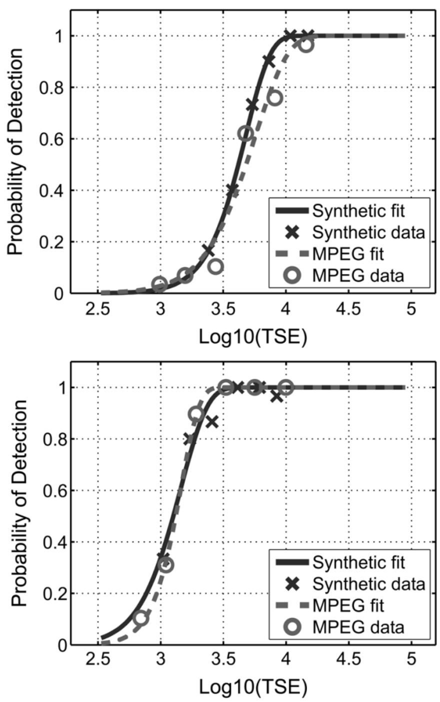


Fig. 3. Psychometric functions of synthetic and MPEG-2 impairments for the sequences "Cheerleaders" (top) and "Football" (bottom), with defect zones in the bottom and middle areas, respectively.

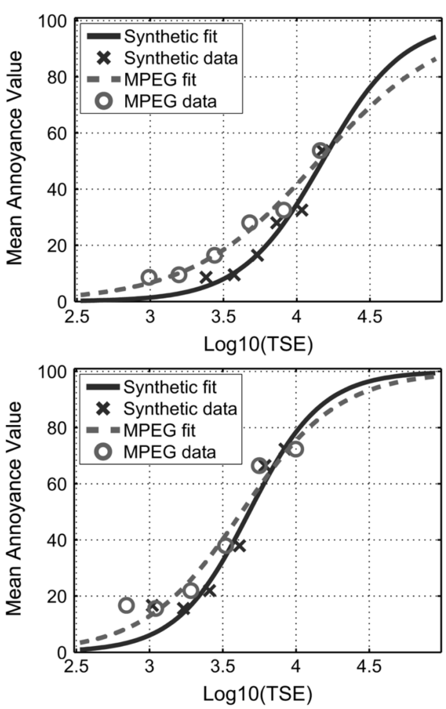


Fig. 4. Annoyance functions of synthetic and MPEG-2 impairments for the sequences "Cheerleaders" (top) and "Football" (bottom), with defect zones in the bottom and middle areas, respectively.

fitted curves, one corresponding to the synthetic impairments and the other to the MPEG-2 impairments.

As can be seen from these figures, the synthetic and real impairments have similar annoyance and psychometric curves. Although not apparent in Fig. 3, most of the annoyance functions for the synthetic impairments are shifted slightly to the right implying (see values in Table I) that at the same TSE synthetic impairments produce slightly less annoyance. Also, most of these functions are steeper for the synthetic impairments, implying that annoyance grows faster with $\log(\mathrm{TSE})$ for these impairments. This is in agreement with the t-test results that show that there are significant differences between the annoyance parameters.

In summary, not only are the generated synthetic impairments visually similar to the real compression impairments, but also they have similar visibility and annoyance properties.

V. Experiment II: Blocky and Blurry Artifacts

The main goal of Experiment II was to estimate the annoyance and visibility of blockiness and blurriness and determine if the synthetic blocky and blurry artifact signals were being perceived as such. The test videos used in this experiment contained either synthetic blocky or blurry artifact signals or a combination of these two types of artifact signals. In this experiment, 44 subjects (21 females and 23 males) performed detection, annoyance, and appearance tasks.

To generate the test sequences, we first created the videos with blockiness X_{blocky} and with blurriness X_{blurry}. Then, we created three different sets of test sequences using (4) with $L = 2$. The first set (combined) consisted of sequences with equal proportions of blocky and blurry artifacts and $r_1 = r_2 = \rho(0.15 \leq \rho \leq 0.5)$. The second set (blocky) consisted of sequences with only blockiness $r_1 = \rho(0.3 \leq \rho \leq 1.0)$ and $r_2 = 0$, while the third set (blurry) consisted of sequences with only blurriness $r_1 = 0$ and $r_2 = \rho(0.3 \leq \rho \leq 1.0)$. In this experiment, we used five originals ("Bus," "Cheerleaders," "Football," "Flower," and "Hockey"), one defect zone, three sets of different types of impairments, and six strength parameter values (ρ) resulting in a total of 95 test sequences (5 originals $\times$ 3 impairment types $\times$ 1 defect zone $\times$ 6 strength parameters $+$ 5 originals).

The best fit parameters for the psychometric and annoyance functions for all test groups are presented in Table II. The fitting parameters varied across originals and impairment signal types. Unfortunately, it was not possible to estimate detection parameters for sequences with only blurriness because for these sequences more than 50% of the subjects detected even the weakest impairment. Therefore, it was not possible to compare the psychometric parameters for these three types of impairments. Fig. 5 depicts the annoyance functions for two sample videos "Bus" and "Cheerleaders." For the same value of $\log(\mathrm{TSE})$ the annoyance for the blocky artifact signal is farthest to the left for three of the videos ("Bus," "Flower," and "Football"). For these three videos, functions for blurry were the farthest to the right and functions for combined, intermediate. For the two other videos ("Cheerleaders" and "Hockey") the annoyance functions are very similar for the three artifact types.

We used ANOVA to test if either the "original" or the "impairment signal type" had a significant effect on the annoyance function parameters. The results show that the "original" had a

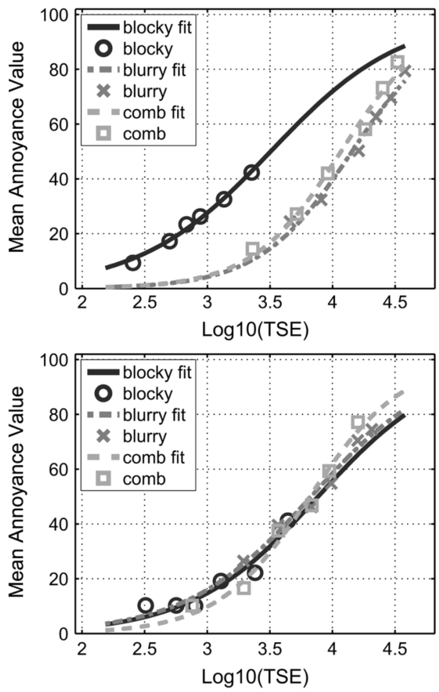


Fig. 5. Annoyance functions for blocky, blurry and combined artifacts for the sequences "Bus" (top) and "Cheerleaders" (bottom), with defect zones in the top and bottom areas, respectively.

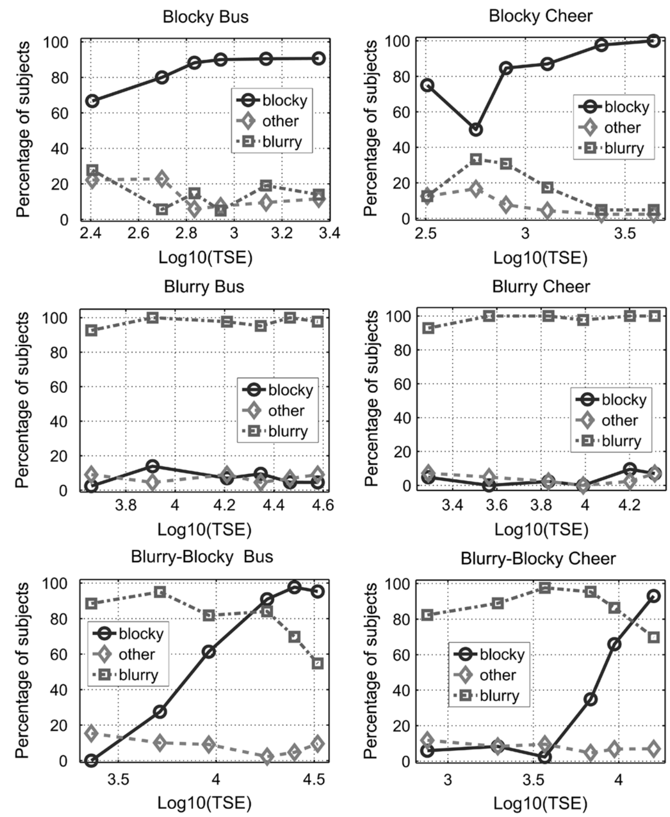


Fig. 6. Percentage of subjects who judged the synthetic blockiness, blurriness, and combined (blocky–blurry) as "blocky," "blurry," and "other."

significant effect on the mid-annoyance value, but the "impairment signal type" effect (illustrated in the top of Fig. 5) was not statistically significant for either of the two parameters.

To analyze how the subjects described the appearance of impairments they saw, we first determined the percentage of subjects who described the impairments as "blocky," "blurry," or "other" for all sequences. Fig. 6 shows the graphs of these percentages versus the $\log(\mathrm{TSE})$, for the videos "Bus" and "Cheerleaders." Each graph contains three curves corresponding to the percentage of "blocky," "blurry," and "other" responses. Since subjects were allowed to mark more than one feature for each sequence, these percentages often added to more than 100. Almost all subjects in most conditions judged the artifacts designed to be blocky and blurry as being "blocky" and "blurry," respectively.

Some subjects reported more than one artifact in cases where only one artifact signal was in the video (see the top and middle sections of Fig. 6). The most common second response was to describe a blocky artifact as also appearing blurry. For all videos, the percentage of subjects that judged the combined artifact signals as "blocky" increased with the TSE, although the proportions of the two artifact signals in the combined video were constant for all TSEs. For higher TSE levels the blockiness seemed to become more salient than the blurriness and some of the subjects classified it only as "blocky." Thus, the appearance of the blocky and combined artifact signals depended in a complex way on the artifact strength.

In summary, although we were not able to compare the psychometric function parameters, a statistical test showed that the annoyance parameters were very similar for the three types of impairments, i.e., even though the three impairment types had a different appearance, their TSE was related to annoyance by similar functions. In spite of the fact that there was a clear effect of the artifact type in three out of the five videos, the ANOVA did not show this effect to be statistically significant. It could be that there is no main effect, but there is a significant interaction. The experiment showed that the synthetic blocky and blurry artifacts were being perceived as such, especially for stronger impairments. It also showed an interaction in the way subjects perceive blockiness and blurriness.

VI. Experiment III: Blurry and Ringing Artifacts

In Experiment III, 40 subjects (19 males and 21 females) performed *detection* and *annoyance* tasks. The main goal of this experiment was to estimate the annoyance and visibility of ringing and blurriness. Although ringing is a fairly commonly occurring artifact, to our knowledge, no work is available on the visibility and annoyance of this type of artifact in videos. Since in compressed videos ringing is frequently accompanied by blurriness, the test sequences used in this experiment contained synthetic ringing and blurry artifacts and a combination of these artifacts.

To generate the test sequences, we first generated videos with only blurriness X_{blurry} and with only ringing X_{ringy}. Then, we created three different sets of test sequences using (4) with $L = 2$. The first set (combined) consisted of sequences with equal proportions of blurry and ringy artifacts and $r_1 = r_2 = \rho\,(0.1 \leq \rho \leq 0.5)$. The second set (blurry) consisted of sequences with only blurriness $r_1 = \rho\,(0.3 \leq \rho \leq 1.0)$ and $r_2 = 0$, while the third set (ringy) consisted of sequences with only ringing $r_1 = 0$ and $r_2 = \rho\,(0.25 \leq \rho \leq 1.25)$. In this

TABLE III
EXPERIMENT III—BEST PARAMETERS FOR ANNOYANCE AND PSYCHOMETRIC FUNCTIONS DETERMINED BY THE METHOD OF LEAST SQUARES

Original	**Impairment**	E_T	κ	E_{50}	η
Bus	Blurry	4.17	11.25	4.77	0.23
Calendar	Blurry	4.17	15.63	4.86	0.20
Cheerleaders	Blurry	3.45	7.13	4.49	0.37
Flower	Blurry	4.00	8.82	4.76	0.23
Hockey	Blurry	2.38	3.41	3.48	0.28
Bus	Comb	4.17	15.94	4.82	0.23
Calendar	Comb	4.17	16.52	4.84	0.21
Cheerleaders	Comb	3.57	6.73	4.55	0.31
Flower	Comb	4.00	14.09	4.71	0.21
Hockey	Comb	-	-	3.56	0.36
Bus	Ringy	3.70	8.44	4.40	0.32
Calendar	Ringy	3.45	9.11	4.25	0.34
Cheerleaders	Ringy	4.00	9.15	4.65	0.29
Flower	Ringy	3.45	14.81	4.11	0.23
Hockey	Ringy	2.70	3.63	3.78	0.50
Mean	-	3.67	10.33	4.40	0.29

experiment, we used five originals ("Bus," "Calendar," "Cheerleaders," "Flower," and "Hockey"), one defect zone, three impairment types (blurry, ringy, and combined), and six strength parameter values (ρ) resulting in a total of 95 test sequences (5 originals $\times$ 1 defect zone $\times$ 3 impairment types $\times$ 6 strength parameters $+$ 5 originals).

With the help of a pilot study with four subjects, we picked values of ρ that produced impairments at strengths varying from hardly visible to very annoying. The values picked for ringy artifacts were higher than the ones for the blurry ones. Even so, the highest strengths of ringing had relatively low TSE and were not highly annoying. Consequently, we were often not able to measure the upper part of the annoyance function for ringing. Nevertheless, we were able to measure enough of this function to get a reasonably good fit to (6). Table III shows the best fit parameters. Figs. 7 and 8 show the pairs of psychometric and annoyance functions for the videos "Bus" and "Cheerleaders."

We performed an ANOVA to test if the "original" and "impairment signal type" had a significant effect on the fitting parameters for both functions. The "original" had a significant effect on the detection threshold $(P = 0.002)$ and mid-annoyance parameters $(P = 0.0249)$, while the "impairment signal type" (blurry, ringing, or combined) did not have a significant effect on any of the fitting parameters $(P > 0.05)$. What this means is that, even though ringing occurs only near edges and blurriness occurs over wide areas of the images, there is no consistent difference between either their thresholds or mid-annoyance values. In other words, although high strengths of ringing rarely occur in practice, annoyance depends on ringing signal strength in essentially the same way that it depends on the blurriness signal strength.

The data of this experiment allowed us to calculated Pearson correlation between E_T and E_{50}. We found that these parameters are highly correlated $(r = 0.948)$ and are related by the following expression: $E_{50} = 0.73 \cdot E_T + 1.77$. This linear expression is similar to the one found for Experiment I, but the coefficient of correlation is higher. We performed a linear regression analysis on the annoyance data to determine if we could predict the MAVs of combined ringy–blurry artifacts from MAVs of the ringy and blurry artifacts by themselves. The results show

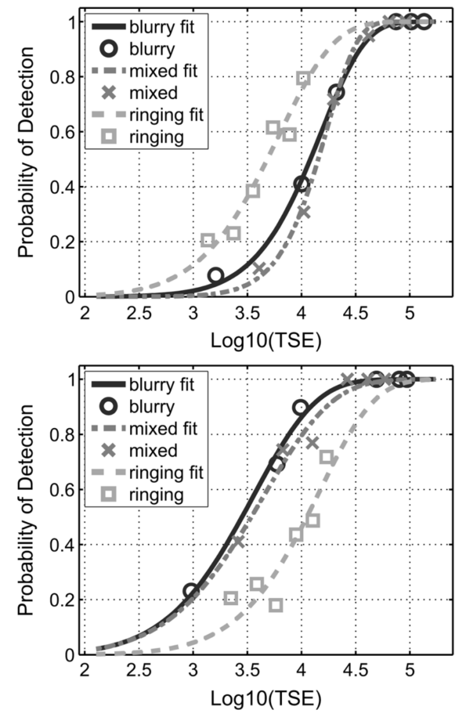


Fig. 7. Psychometric functions for blurry, ringy, and combined artifacts for the sequences "Bus" (top) and "Cheerleaders" (bottom), with defect zones in the middle and bottom areas, respectively.

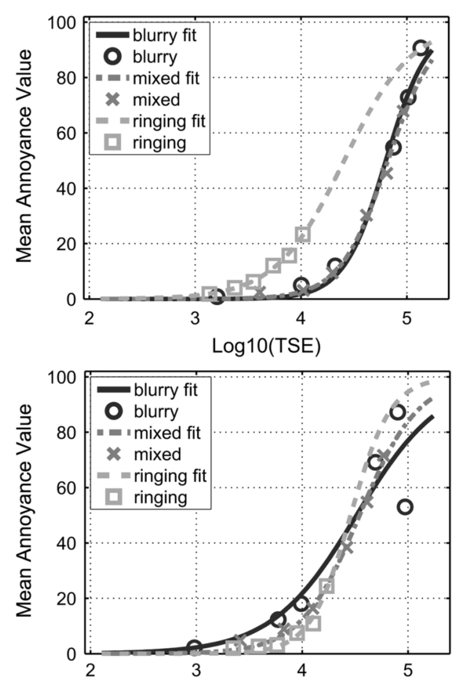


Fig. 8. Annoyance functions for blurry, ringy and combined artifacts for the sequences "Bus" (top) and "Cheerleaders" (bottom), with defect zones in the middle and bottom areas, respectively.

a very significant correlation ($r = 0.96$; t-test $P = 1.8 \cdot 10^{-19}$) among these values. We found that the predicted MAV (PMAV) is given by $\mathrm{PMAV} = 0.63 \cdot \mathrm{MSV}_{\mathrm{blurry}} + 0.41 \cdot \mathrm{MSV}_{\mathrm{ringy}}$.

In summary, the results of this experiment are in agreement with the previous experiments. The "impairment type" did not have a statistically significant effect on neither the annoyance nor psychometric function parameters, implying that the TSE of different artifacts are related in the same way to annoyance and probability of detection. A high correlation was found between E_T and E_{50} and the results are in agreement with the results of Moore for MPEG-2 impairments.

VII. Experiment IV: Blocky, Blurry, Noisy, and Ringy Artifacts

The goal of our last experiment was to compare the visibility and annoyance of all the four artifacts: blockiness, blurriness, ringing, and noisiness. We also wanted to investigate the importance of each type of artifact while determining the overall annoyance. In this experiment, 42 subjects (17 males and 25 females) performed *detection* and *annoyance* tasks. To avoid context effects, we presented the test sequences with the four types of synthetic artifacts (by themselves) at different levels of strength in the same experimental session. This allowed for a more robust test of the visibility thresholds and mid-annoyance values of these artifacts.

The test sequences in this experiment were generated in a similar way to previous experiments. We first generated videos with only blockiness X_{blocky}, with only blurriness X_{blurry}, with only noisiness X_{noisy}, and with only ringing X_{ringy}. To obtain the test sequences with several levels of annoyance, we used (4) with $L = 4$. But, unlike other experiments, the videos with different artifacts were not combined in this experiment and four sets of sequences were generated with only one type of artifact: blocky, blurry, noisy, and ringy. So, for the first set (blocky) we had $r_1 = \rho$ and $r_2 = r_3 = r_4 = 0$, for the second set (blurry) $r_2 = \rho$ and $r_1 = r_3 = r_4 = 0$, for the third set (noisy) $r_3 = \rho$ and$r_1 = r_2 = r_4 = 0$, and for the fourth set (ringy) $r_4 = \rho$ and$r_1 = r_2 = r_3 = 0$. We used five originals ("Bus," "Calendar", "Cheerleaders," "Flower," and "Hockey"), three defect zones, four artifact types (blocky, blurry, noisy, and ringy), and six strength parameter values (ρ) resulting in a total of 365 test sequences (5 originals $\times$ 3 defect zones $\times$ 4 impairment types $\times$ 6 strength parameters + 5 originals). Since the number of test sequences was too large to be performed in only one session, we divided the subjects into three groups and each group viewed a subset of test sequences. Each subset contained all originals, all types of artifact signals, but different groups of defect zones.

Table IV shows the best annoyance and psychometric function fitting parameters grouped by type of artifact. Figs. 9 and 10 show the pairs of psychometric and annoyance functions for the videos "Bus" and "Cheerleaders." Each plot corresponds to one defect zone and one original and contains four different curves, one for each type of artifact signal. As can be seen from these figures, the fits of both functions were generally good.

In order illustrate the variations of the parameters for the different artifacts an originals, in Fig. 11, we show bar plots of the average values of E_T and E_{50} (over defect zones), respectively. As can be seen from these bar plots, the original video seemed to have a larger effect on the detection threshold and mid-annoyance than the type of artifact signal. The result of an ANOVA test showed that "artifact signal type" and "defect zone" did not have a significant effect on any of the fitting parameters. On the

TABLE IV
Experiment IV— Best Parameters for Annoyance and Determined by the Method of Least Squares

	blocky				noisy			
Sequence	E_T	κ	E_{50}	η	E_T	κ	E_{50}	η
Bus-Top	3.13	4.94	3.96	0.50	2.94	19.37	3.35	0.27
Bus-Middle	2.94	14.97	3.40	0.19	-	-	3.72	0.83
Bus-Bottom	3.85	6.04	4.10	0.22	3.13	15.97	3.54	0.26
Calendar-Left	2.38	4.60	3.56	0.46	2.86	8.90	3.60	0.37
Calendar-Middle	2.70	14.83	3.84	0.43	2.44	5.63	3.87	0.44
Calendar-Right	2.63	9.18	3.32	0.31	3.13	9.12	3.76	0.28
Cheer-Top	3.45	17.16	3.77	0.18	3.03	6.25	3.87	0.33
Cheer-Middle	3.23	22.71	3.64	0.18	3.13	12.71	4.09	0.34
Cheer-Bottom	3.45	12.01	3.95	0.33	3.70	4.04	4.67	0.43
Flower-Sky	2.63	6.29	3.22	0.27	-	-	3.96	0.70
Flower-Houses	-	-	3.26	0.30	2.86	14.76	3.74	0.41
Flower-Flowers	3.03	13.98	4.00	0.40	3.85	33.58	4.35	0.26
Hockey-Left	2.70	3.82	2.85	0.50	2.86	6.54	3.61	0.23
Hockey-Middle	2.70	10.94	3.18	0.26	2.86	7.71	3.61	0.29
Hockey-Right	2.56	7.13	3.13	0.19	2.22	4.68	3.48	0.47
	blurry				**ringy**			
Sequence	E_T	κ	E_{50}	η	E_T	κ	E_{50}	η
Bus-Top	2.86	11.32	3.25	0.15	-	-	4.03	0.50
Bus-Middle	3.70	31.34	4.16	0.28	2.70	4.70	3.81	0.40
Bus-Bottom	2.78	11.57	3.21	0.17	3.57	11.56	3.68	0.06
Calendar-Left	3.70	16.47	3.92	0.50	3.03	9.25	3.78	0.31
Calendar-Middle	3.85	12.67	4.26	0.50	3.03	12.53	3.64	0.24
Calendar-Right	3.70	16.83	4.16	0.21	3.03	15.38	3.78	0.26
Cheer-Top	3.23	9.08	3.77	0.22	3.23	7.15	3.75	0.18
Cheer-Middle	3.33	8.68	3.76	0.20	3.45	12.79	3.95	0.18
Cheer-Bottom	3.23	16.59	3.60	0.26	3.57	7.27	4.26	0.40
Flower-Sky	2.94	9.40	3.56	0.26	2.70	14.10	3.74	0.34
Flower-Houses	3.57	12.32	4.14	0.19	3.13	12.07	3.65	0.17
Flower-Flowers	3.85	19.62	4.24	0.15	3.57	16.34	4.13	0.21
Hockey-Left	2.63	13.89	3.00	0.18	-	-	3.80	0.50
Hockey-Middle	2.63	7.73	3.10	0.18	2.78	8.49	3.31	0.20
Hockey-Right	2.50	5.80	2.96	0.18	2.13	4.17	2.80	0.33
Mean	$\overline{E_T} = 3.07$; $\overline{\kappa} = 6.82$				$\overline{E_{50}} = 3.69$; $\overline{\eta} = 0.31$			

other hand, "original video" had a significant effect on the parameters E_T and E_{50}. These results are in agreement with our previous results.

Although the statistical analysis showed that the "artifact type" did not have a significant effect on determining the parameters of the psychometric and annoyance functions, some of the graphs of these functions show that they are displaced from one another on the $\log(\mathrm{TSE})$ axis. This indicates that artifact type might have an effect on both detection and annoyance for individual combinations of video and defect region. However, this effect can be opposite for different videos and for different regions in the same video. As a consequence, the main effect of impairment type is small and usually not statistically significant for this experiment or the previous experiments. It appears that there is an interaction between video, defect zone, and impairment type. However, this interaction has been found not to be statistically significant.

We calculated the Pearson correlation between E_T and E_{50} for the set of all test sequences containing blocky, blurry, ringing, and noisy artifact signals. The parameters have a positive Pearson correlation $r = 0.783$ and are related by the following expression: $E_{50} = 0.71 \cdot E_T + 1.51$. The linear relationship found for this experiment is similar to the one obtained for Experiment III ($E_{50} = 0.73 \cdot E_T + 1.77$) and slightly different from the expression found for Experiment

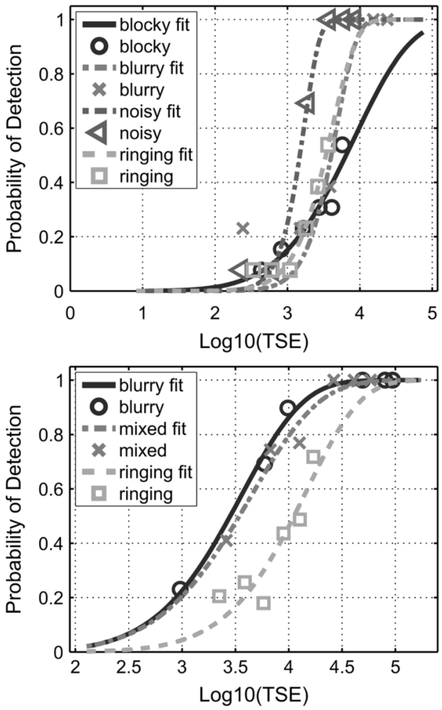


Fig. 9. Psychometric functions for blocky, blurry, ringy, and noisy artifacts for the sequences "Bus" (top) and "Cheerleaders" (bottom), with defect zones in the top areas.

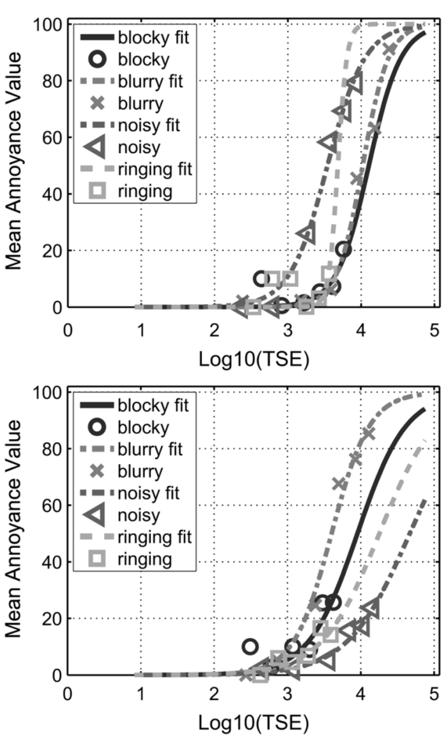


Fig. 10. Annoyance functions for blocky, blurry, ringy, and noisy artifacts for the sequences "Bus" (top) and "Cheerleaders" (bottom), with defect zones in the top areas.

I ($E_{50} = 0.8228 \cdot E_T + 1.2272$). The data from Experiment II was not sufficient analyze the correlation between E_T and E_{50}. Combining the data gathered from Experiments I, III, and IV, we obtained the following linear relation: $E_{50} = 0.8304 \cdot E_T + 1.1962$ with correlation $r = 0.7283$.

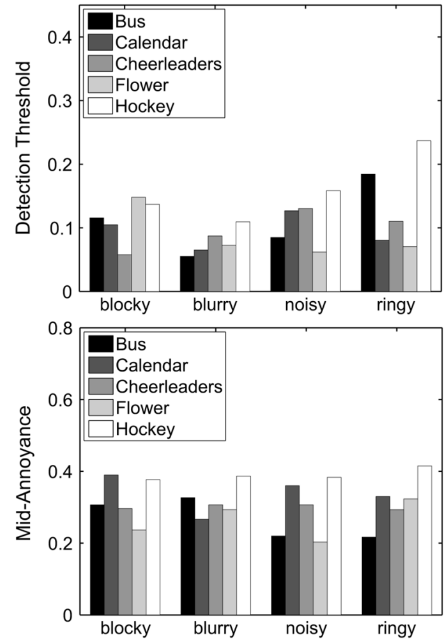


Fig. 11. Bar plot of the detection threshold (top) and mid-annoyance (bottom) values averaged over defect zone.

Moore found a high correlation ($r = 0.971$) between the E_T and E_{50} values for MPEG blocky, blurry, and fuzzy artifacts and the relation $E_{50} = 1.15 \cdot E_T$ [8]. Therefore, the linear relation found in this paper differs from the results found by Moore by a constant. Otherwise, both results are in agreement and show that indeed detection and annoyance are correlated and can be expressed by a linear relationship.

In summary, the results of this experiment showed that the annoyance and visibility of different "artifact types" shared a similar relationship with TSE. On the other hand, content ("original") did have a significant impact on them. We also confirmed results found previously that annoyance and visibility are correlated.

VIII. Conclusion

In this paper, we presented the results of a series of four psychophysical experiments with the goal of studying the appearance, annoyance, and detectability of common compression artifacts. Our approach consisted of generating the artifacts synthetically and studying them at different strengths and combinations. Our generated synthetic impairments looked visually similar to the real compression impairments. In Experiment I, we compared them with real compression artifacts and found that they had similar visibility and annoyance functions. In Experiment II, subjects correctly identified the artifacts. This experiment also showed that blocky–blurry artifacts appeared blurry at low strength and blocky at high strength. Experiments II-IV showed that the "artifact signal type" (blockiness, blurriness, ringing, or noisiness) did not have a significant main effect on the fitting parameters. The "original," on the other hand, had a significant effect on both the detection threshold and the mid-annoyance parameters. Finally, the results of Experiments I, III, and IV confirmed previous results that annoyance and visibility parameters E_T and E_{50} are positively correlated and linearly related.

REFERENCES

[1] Int. Telecomm. Union, Geneva, Switzerland, "ITU-T Recommendation P.930: Principles of a reference impairment system for video," 1996.

[2] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Process.*, vol. 70, no. 3, pp. 247–78, 1998.

[3] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, pp. 179–206.

[4] A. B. Watson, "Towards a visual quality metric for digital video," in *Proc. Eur. Signal Process. Conf.*, Island of Rhodes, 1998, pp. 1057–1060.

[5] S. Winkler, "Visual quality assessment using a contrast gain control model," in *Proc. IEEE 3rd Workshop Multimedia Signal Process.*, 1999, pp. 527–532.

[6] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," in *Proc. SPIE Human Vision, Visual Process., Digit. Display IV*, 1993, pp. 15–26.

[7] H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments," in *Proc. SPIE Human Vision, Visual Process. Digit. Display III*, 1992, pp. 16–26.

[8] M. Moore, J. Foley, and S. Mitra, "Defect visibility and content importance: Effects on perceived impairment," *Image Commun.*, vol. 3, pp. 185–203, 2004.

[9] M. C. Q. Farias, "No-reference and reduced reference video quality metrics: New contributions," Ph.D. dissertation, Dept. Electr. Comput. Eng., Univ. California Santa Barbara, Santa Barbara, 2004.

[10] G. W. Cermak, S. Wolf, E. P. Tweedy, M. H. Pinson, and A. A. Webster, "Validating objective measures of MPEG video quality," *SMPTE J. Soc. Motion Picture Television Eng.*, vol. 107, no. 4, pp. 226–35, 1998.

[11] Video Quality Experts Group, "VQEG subjective test plan (Phase 1)," (1999). [Online]. Available: ftp://ftp.crc.ca/crc/vqeg/phasel-docs

[12] Int. Telecommun. Union, Geneva, Switzerland, "ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures," 1998.

[13] J. M. Libert, C. P. Fenimore, and P. Roitman, "Simulation of graded video impairment by weighted summation: Validation of the methodology," in *Proc. SPIE Multimedia Syst. Appl. II*, 1999, pp. 254–65.

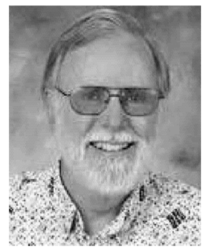
[14] G. W. Snedecor and W. G. Cochran*, Statistical Methods*, 8th ed. Ames, IO: Iowa State Univ. Press, 1989.

[15] M. S. Moore, J. M. Foley, and S. K. Mitra, "Detectability and annoyance value of MPEG-2 artifacts inserted into uncompressed video sequences," in *Proc. SPIE Human Vision Electron. Imaging V*, 2000, pp. 99–110.

Mylène C. Q. Farias (M'02) received the B.Sc. degree in electrical engineering from Universidade Federal de Pernambuco (UFPE), Recife, Brazil, in 1995, the M.Sc. degree in electrical engineering from the Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil, in 1998, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2004, for work in no-reference video quality metrics.

She is currently a Postdoctoral Researcher and a Visiting Professor at Universidade Federal de Campina Grande (UFCG), Campina, Brazil. She has worked as a Research Engineer in video quality assessment and validation of video quality metrics at Fundação CPqD Centro de Pesquisa e Desenvolvimemto em Telecomunicações, Campinas, Brazil. She has also worked as an intern in video quality assessment of sharpness algorithms for Philips Research Laboratories, Eindhoven, The Netherlands, and as a Consultant developing no-reference video quality metrics for Intel Corporation, Phoenix, AZ. Her current research interests include video quality metrics, video processing, multimedia, watermarking, and information theory.

Dr. Farias is a member of SPIE and the video quality experts group (VQEG).



John M. Foley was an undergraduate at the University of Notre Dame, Notre Dame, IN, where he majored in physics. He did graduate studies at Columbia University, New York, where he received the Ph.D. degree in psychology with a specialty in visual perception.

Shortly after that, he joined the faculty of the University of California, Santa Barbara, where he is currently a Research Professor of Psychology. He works on the experimental and theoretical analysis of pattern vision, space perception, and visual-motor control. He was a Fulbright Scholar at the University of Louvain, Belgium, an NIH Fellow at the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, a James McKeen Cattell Fellow at the University of Cambridge in England, a Visiting Lecturer at the University of Tokyo, Tokyo, Japan, and a Hoffman–LaRoche Fellow at the Institute for Biophysics at the University of Freiburg in Germany.

Sanjit K. Mitra (S'59–M'63–SM'69–F'74–LF'01) is the Stephen and Etta Varra Professor in the Department of Electrical Engineering Systems, University of Southern California, Los Angeles, and a Research Professor at the University of California, Santa Barbara. He has held visiting appointments in Australia, Austria, Brazil, Croatia, Finland, Germany, India, Japan, Singapore, Turkey, and the U.K.

Dr. Mitra has served IEEE in various capacities including service as the President of the IEEE Circuits and Systems Society in 1986. He was a recipient of several awards including the 1989 Education Award and the 2000 Mac Van Valkenburg Society Award of the IEEE Circuits and Systems Society, the 1996 Technical Achievement Award, the 2001 Society Award and the 2006 Education Award of the IEEE Signal Processing Society, the 2001 McGraw-Hill/Jacob Millman Award of the IEEE Education Society, the 2002 Technical Achievement Award of the European Association for Signal Processing (EURASIP), the 2005 SPIE Technical Achievement Award of the International Society for Optical Engineering, the 2006 IEEE James H. Mulligan Jr. Education Medal, and a co-recipient of the 2000 Blumlein-Browne-Willans Premium of the Institution of Electrical Engineers (London) and the 2001 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award. He has been awarded honorary doctorate degrees from the Tampere University of Technology, Finland, and the Technical University of Bucharest, Romania, and the University Medal of the Technical University of Slovakia at Bratislava. He is an Academician of the Academy of Finland, a Member of the U.S. National Academy of Engineering, a Member of the Norwegian Academy of Technological Sciences, a Foreign Member of the Croatian Academy of Sciences and Arts, and a Foreign Member of the Academy of Engineering of Mexico. He is a Fellow of the AAAS and SPIE.