Perceptual Analysis of Video Impairments that Combine Blocky, Blurry, Noisy, and Ringing Synthetic Artifacts

Mylène C.Q. Farias,[a] John M. Foley,[b] and Sanjit K. Mitra [a†]

[a] Department of Electrical and Computer Engineering,
[b] Department of Psychology,
University of California, Santa Barbara, CA 93106 USA

ABSTRACT

In this paper we present the results of a psychophysical experiment which measured the overall annoyance and artifact strengths of videos with different combinations of blocky, blurry, noisy, and ringing synthetic artifacts inserted in limited spatio-temporal regions. The test subjects were divided into two groups, which performed different tasks – 'Annoyance Judgment' and 'Strength Judgment'. The 'Annoyance' group was instructed to search each video for impairments and make an overall judgment of their annoyance. The 'Strength' group was instructed to search each video for impairments, analyze the impairments into individual features (artifacts), and rate the strength of each artifact using a scale bar. An ANOVA of the overall annoyance judgments showed that the artifact physical strengths had a significant effect on the mean annoyance value. It also showed interactions between the video content (original) and 'noisiness strength', 'original' and 'blurriness strength', 'blockiness strength' and 'noisiness strength', and 'blurriness strength' and 'noisiness strength'. In spite of these interactions, a weighted Minkowski metric was found to provide a reasonably good description of the relation between individual defect strengths and overall annoyance. The optimal value found for the Minkowski exponent was 1.03 and the best coefficients were 5.48 (blockiness), 5.07 (blurriness), 6.08 (noisiness), and 0.84 (ringing). We also fitted a linear model to the data and found coefficients equal to 5.10, 4.75, 5.67, and 0.68, respectively.

Keywords: artifacts, perceptual video quality, video, blockiness, blurriness, noisiness, ringing.

1. INTRODUCTION

Impairments can be introduced during capture, transmission, storage, and/or display, as well as by any image processing algorithm (e.g. compression) that may be applied along the way. They can be very complex in their physical descriptions and also in their perceptual descriptions. Most of them have more than one perceptual feature, but it is possible to produce impairments that are relatively pure. To differentiate impairments from their perceptual features, we use the term artifact to refer to the perceptual features of impairments and artifact signal to refer to the physical signal that produces the artifact. Examples of artifacts introduced by digital video systems are blurriness, noisiness, ringing, and blockiness.[1,2]

Designing a video quality metric, especially a no-reference metric, is not an easy task. One approach consists of using a multidimensional feature extraction, i.e., to recognize that the perceived quality of a video can be affected by a variety of artifacts and that the strengths of these artifacts contribute to the overall annoyance [3]. This approach requires a good knowledge of the types of artifacts present in digital videos. Although many video quality models have been proposed, little work has been done on studying and characterizing the individual artifacts found in digital video applications. An extensive study of the most relevant artifacts is necessary, since we still do not have a good understanding of how artifacts depend on the physical properties of the video and how they combine to produce the overall annoyance.

The approach taken in this work for studying individual artifacts has been to work with synthetic artifacts that look like "real" artifacts, yet are simpler, purer, and easier to describe.[2] 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. Synthetic artifacts

[*] Further author information: (Send correspondence to M.C.Q.F.)
M.C.Q.F.: E-mail: mylene@ece.ucsb.edu, J.M.F.: E-mail: foley@psych.ucsb.edu, S.K.M.: E-mail: mitra@ece.ucsb.edu.

make it possible, for example, to study the importance of each artifact to human observers. Such artifacts are necessary components of the kind of reference impairment system recommended by the ITU-T for the measurement of image quality [2] and offer advantages for experimental research on video quality.

There are several properties that are desirable in synthetic artifacts, if they are to be useful for these purposes. According to ITU-T [2], the synthetic artifacts should:

- be generated by a precisely defined and easily replicated algorithm,
- be relatively pure and easily adjusted and combined to match the appearance of the full range of compression impairments, and
- produce psychometric functions and annoyance functions that are similar to those for compression artifacts.

In this work, we created four types of synthetic artifacts; blockiness, blurriness, noisiness, and ringing. We generated test sequences by combining blockiness, blurriness, ringing, and noisiness signals and different subsets of these four. Each signal was either present at full strength or absent. Then, we performed a psychophysical experiment in which human subjects detected these impairments, judged their overall annoyance, analyzed them into artifacts and rated the strengths of the individual artifacts. The main goal of this work was to determine how the strengths of blocky, blurry, ringing, and noisy artifacts combine to determine the overall annoyance and to express this in a model that shows the relative importance of the different artifacts in determining overall annoyance.

2. GENERATION OF SYNTHETIC ARTIFACTS

In this section we describe the algorithms for the creation of synthetic blockiness, blurriness, ringing, and noisiness. The proposed algorithms satisfy the conditions recommended by ITU-T and are simpler than the algorithms described in the ITU-T recommendation. Further, the algorithms have the advantage of producing relatively pure artifacts that are a good approximation of the artifacts generated by digital video coding systems and can be combined in different strengths and proportions.

Blockiness (also known as blocking) is a distortion of the image/frame characterized by the appearance of the underlying block encoding structure.[2] Blockiness is often caused by coarse quantization of the spatial frequency components during the encoding process. We produced blockiness by using the difference between the average of each block and the average of the surrounding area to make each block stand out. Since many compression algorithms use 8×8 blocks, this was the size of the blocks that were used by the algorithm. The algorithm for generating blockiness was applied separately to the Chrominance (Cb and Cr) and Luminance (Y) components of the video. The algorithm can be easily modified to use different block sizes and to include spatial shifts frequently introduced by compression algorithms.[4] To generate blockiness, we first calculated the average of each 8×8 block of the frame and of the 24×24-surrounding block, which had the current 8×8 block at its center. Then, we calculated the difference, $D(i,j)$, between these two averages for each block of the frame. The values of $D(i,j)$ were the same for all pixels inside the same 8×8 block. To each block of the original frame, we added the corresponding element of the difference matrix $D(i,j)$:

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

where X_0 is the original frame and Y is the frame with blockiness and i and j refer to spatial position of the pixel in the frame. While adding D to the frame it was important to make sure that none of the pixels become too saturated, i.e., either they were too negative (look much darker than the surrounding area), or they were too positive (look much brighter than the surrounding area). The values of D were limited to avoid this problem. Before adding the blockiness to the defect zones, the average of the frame was corrected to avoid the borders around the defect zones becoming more visible than intended. To correct the average we first calculated the average of the frame, μ_0, before introducing the artifacts, and the average, μ_f, after introducing them. Then, we added the average difference $\mu_0 - \mu_f$ to all pixels in the frame.

Blurriness is defined as a loss of spatial details and a reduction in the sharpness of edges in moderate to high frequency regions of the image or video frame, such as in roughly textured areas or around scene objects.[2] Blurriness presents itself in almost all processing stages of a communications system; in acquisition, where it is introduced by both the camera lens and camera motion, during pre- and post-processing, and display, where it shows up in monitors with low resolution. In compressed videos, blurriness is often caused by trading off bits to code resolution and motion.

Blurriness can be easily simulated by applying a symmetric, two-dimensional FIR (finite duration impulse response) low-pass filter to the frame array.[2] Several filters with varying cut-off frequencies can be used to allow control over the amount of blurriness introduced. In this work, we used a simple 5×5 average filter to generate blurriness. Varying the size of the filter increases the spread of the blur, making it stronger and, consequently, more annoying.

Physically noise (noisiness signal) is defined as an uncontrolled or unpredicted pattern of intensity fluctuations that is unwanted and does not contribute to the quality of a video image.[1,2] There are many types of noise present in compressed digital videos and two of the most common are mosquito noise and quantization noise. We created synthetic noisiness by replacing the luminance value of pixels at random locations with a constrained random value. The color components were left untouched. The random location of the pixels to change was determined by drawing two random numbers, corresponding to the coordinates of the pixel. After a pixel location was determined, the pixel value was replaced by a random value in the range 10 to 120 to avoid saturation. Additional pixel locations were selected until the desired ratio of impaired/non-impaired number of pixels was obtained. This ratio is an indication of the level of noisiness present in the video. The ratio used for this work was 10%.

Ringing is fundamentally related to the Gibb's phenomenon.[5] It occurs when the quantization of individual DCT coefficients results in high frequency irregularities of the reconstructed block. Ringing manifests itself in the form of spurious oscillations of the reconstructed pixel values. It is more evident along high contrast edges, especially if the edges are in the areas of generally smooth textures.[1,2] The ITU-T reference impairment system recommends generating ringing using a filter with ripples in the passband amplitude response, which creates an echo impairment.[2] The problem with this approach is that besides ringing, this procedure also introduces blurriness and possibly noisiness. Since our goal was the generation of artifacts as pure as possible, we propose a new algorithm for synthetically generating ringing that does not introduce other artifacts. Our algorithm consisted of a pair of delay-complementary highpass and lowpass filters, related by the following relationship:

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

where $H(z)$ and $G(z)$ are N-tap highpass and lowpass filters, respectively. We set $\rho = 1$ and $n_0 = 0$. The output of our system was given by the following equation:

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

So, except for a shift, Y was equal to X_0, given that the initial conditions of both filters were exactly the same.[5] If, on the other hand, we made the initial conditions different, a decaying noise was introduced in the first $N/2$ samples that resembled the ringing artifact produced by compression. An example of this effect can be seen on Figure 1, where both input (solid line) and output (dashed line) are plotted. In this example, $N=10$ and the input was $x_0=\cos(0.1t)+\cos(0.8t)$. Since ringing is only visible around edges, the algorithm was only applied to the pixels

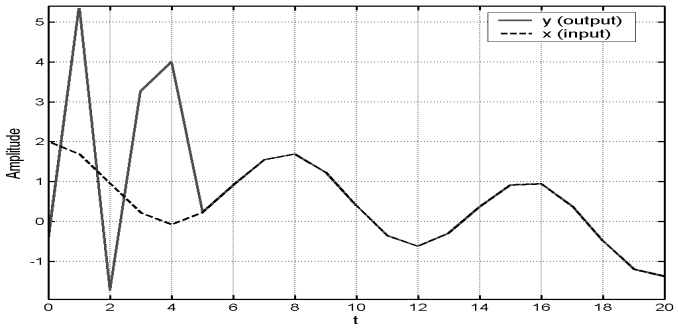


Figure 1 Ringing simulation in a 1-D signal with a sharp edge at time 0. Dashed line is the input signal, while the solid line is the reconstructed signal with shift compensation.

of the video corresponding to edges in both horizontal and vertical directions. We used the Canny algorithm[6] to detect the edges. The resulting effect was very similar to the ringing artifact found in compressed images, but without any blurriness or noisiness.

The Recommendation ITU-T P.930[2] specifies that a system with the purpose of simulating commonly found artifacts must be able to produce them in different proportions and strengths. In this work, we linearly combine the synthetic artifact signals using a combination rule. The main advantage of using this method is that it reduces the possibility of one artifact eliminating or reducing another artifact. For example, if we add blockiness to a video and later filter the video for adding blurriness, the last operation would probably eliminate a good amount of blockiness. Combining artifact signals using a combination rule produces less of this type of interaction. Another advantage is that this method allow us to study each artifact individually.

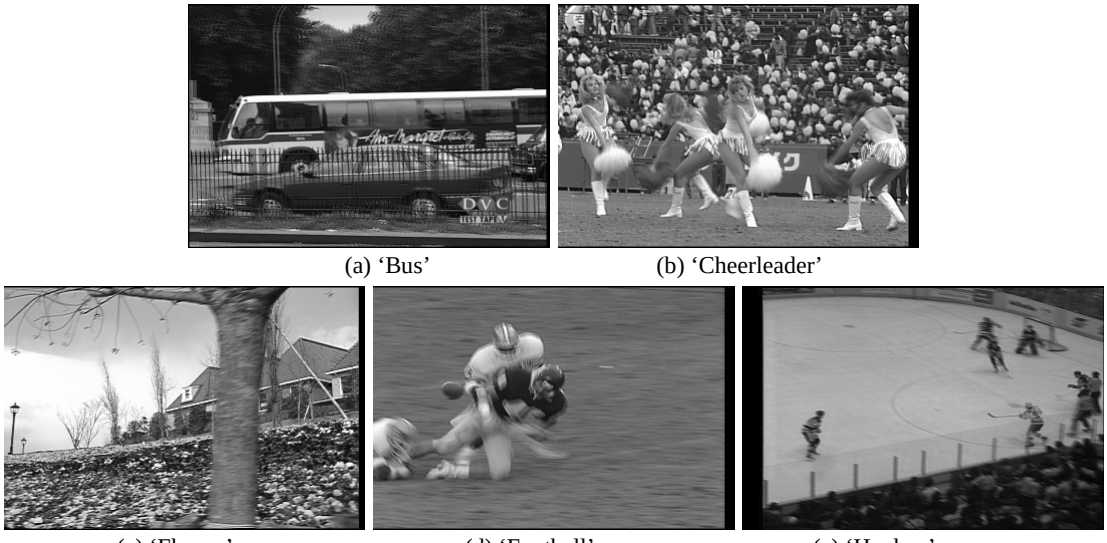
2. GENERATION OF EXPERIMENT TEST SEQUENCES

To generate the test video sequences, we started by choosing a set of five original video sequences of assumed high quality: 'Bus', 'Calendar', 'Cheerleader', 'Flower', and 'Hockey'. These videos are commonly used for video experiments and publicly available.[7] Representative frames of the videos used are shown in Figure 2. The second step was to generate videos in which one type of artifact dominated and produced a relatively high level of annoyance. For each original, 4 new videos were created: X_{blurry}, with only blurriness, X_{blocky}, with only blockiness, X_{ringy}, with only ringing, and X_{noisy}, with only noisiness. These synthetic artifacts were not equal in Total Squared Error (TSE) or in Annoyance; both TSE and annoyance were less for blockiness and ringing than for blurriness and noisiness.

Then, the test sequences (Y) were generated by linearly combining the original video with the video containing the individual artifact (X_{blurry}, X_{blocky}, X_{ringy}, or X_{noisy}) in different proportions, as given by the following equation:

$$Y = a \cdot X_{blocky} + b \cdot X_{blurry} + c \cdot X_{noisy} + d \cdot X_{ringy} + w \cdot X_0 \tag{4}$$

where X_0 is the original video, Y is the impaired video and a, b, c, d, and w are the weights of the blocky, blurry, noisy, ringy, and original videos, respectively ($0 \le a, b, c, d, w \le 1$). By varying these values, we can change the appearance of the overall impairment making it more blocky, blurry, noisy, or ringy, as desired. The 24 combinations of the parameters a, b, c, d, and w used to generate the test sequences are shown in columns 2-5 of Table 1.



(a) 'Bus' (b) 'Cheerleader'

(c) 'Flower' (d) 'Football' (e) 'Hockey'

Figure 2 Sample frames of original videos used in the experiment.

Table 1 Set of values (combinations) for *a*, *b*, *c*, *d*, and *w* for the experiment. Average values of *MSV*s and *MAV*s for each combination over all videos.

Comb	*a*	*b*	*c*	*d*	*w*	MSV_{block}	MSV_{blur}	MSV_{noise}	MSV_{ring}	*MAV*
1	0	0	0	0	1	0.042	0.322	0.035	0.748	0.385
2	0	0	0.67	0	0.33	0.202	0.225	5.9	0.166	35.844
3	0	0.67	0	0	0.33	0.327	5.931	0.053	0.31	29.052
4	0	0.67	0.67	0	0	0.252	4.999	6.408	0.313	62.533
5	1	0	0	0	0	4.292	0.475	0.056	0.276	17.859
6	1	0	0.67	0	0	1.496	0.58	5.97	0.228	43.756
7	1	0.67	0	0	0	6.663	2.491	0.04	0.195	48.607
8	1	0.67	0.67	0	0	4.518	2.819	6.293	0.258	67.837
9	0	0	0	1	0	0.13	0.474	0.093	2.568	3.422
10	0	0	0.67	1	0	0.196	0.392	6.254	0.363	38.978
11	0	0.67	0	1	0	0.211	6.109	0.501	3.181	36.430
12	0	0.67	0.67	1	0	0.171	4.626	6.55	0.672	64.970
13	1	0	0	1	0	4.771	0.609	0.121	1.177	18.111
14	1	0	0.67	1	0	1.508	0.644	6.232	0.235	45.896
15	1	0.67	0	1	0	6.508	2.818	0.275	0.757	56.778
16	1	0.67	0.67	1	0	4.235	2.798	6.239	0.405	70.400
17	1	1	0.33	1	0	6.247	4.296	5.443	0.575	81.830
18	1	0.67	0	1	0	6.608	2.918	0.175	0.967	54.144
19	0.67	0.67	0	0.67	0	5.116	4.052	0.078	0.455	42.926
20	0	0	0	0.33	0.67	0.098	0.433	0.055	0.791	0.852
21	0	0	0	0.67	0.33	0.076	0.369	0.156	1.457	1.385
22	0	0	0.1	0	0.9	0.031	0.326	0.255	0.594	1.659
23	0	0	0.25	0	0.75	0.182	0.391	3.022	0.161	12.022
24	0	0	0.8	0	0.2	0.167	0.262	6.358	0.211	41.281

In most cases, $a + b + c + d \leq 1$ but, for some combinations in this experiment, this sum was greater than 1 to make impairments stronger. Nevertheless, pixel values were limited between 0 and 255 to avoid saturation. Again, we did not use all possible combinations of the four artifact signals because that would have made the experiment too long. The total number of test sequences in this experiment was 125, which included 120 test sequences (5 originals × 24 combinations) plus the five original sequences. The sequences were shown in different random orders for different groups of observers during the main experiment.

In order to be able to identify the major factors and interactions terms affecting the annoyance values, the set of combinations include a *full factorial design*[8] (combinations 1-16) of the four artifact signals. A full factorial design is an experimental design used when the number of factors is limited. In such a design, the levels (or strengths) of the variables are chosen in such a way that they span the complete factor space. Often, only a lower and upper level are chosen. In our case, we have four variables that correspond to the strengths of blocky, blurry, ringy, and noisy artifact signals (*a*, *b*, *c*, *d*, and *w*). As can be seen in Table 1 (combinations 1-16), only two values are possible for each artifact signal strength: 0 and 1.00 for ringing and blockiness, 0 and 0.67 for blurriness and noisiness. Ringing and blockiness are given higher upper values in order to make the artifacts more similar in TSE and annoyance. Combinations 17-19 were added as samples of 'typical' compression combinations. The last five combinations were added to complement data from previous experiments.

3. METHOD

The Image Processing Laboratory at UCSB, in conjunction with the Visual Perception Laboratory, has been performing experiments on video quality for the last three years. Our test subjects were drawn from a pool of students in the introductory psychology class at UCSB. The students are thought to be relatively naive concerning video artifacts and the associated terminology.

The normal approach to subjective quality testing is to degrade a video by a variable amount and ask the test subjects for a quality/impairment rating.[9] The degradation is usually applied to the entire video. In this research we have been using an experiment paradigm that measures the annoyance value of brief, spatially limited artifacts in video.[10] We degrade one specific region of the video for a short time interval. The rest of the video clip is left in its original state. Different regions were used for each original to prevent the test subjects from learning the locations where the defects appear. The regions used in this experiment were centered strips (horizontal or vertical) taking 1/3 of the frame. They were 1 second long and did not occur during the first and last seconds of the video.

For our experiments, the test sequences were stored on the hard disk of an NEC server. Each video was displayed using a subset of the PC cards normally provided with the Tektronix PQA-200 picture quality analyzer. Each test sequence can be loaded and displayed in six to eight seconds. A generator card was used to locally store the video and stream it out in a serial digital (SDI) component format. The test sequence length was limited to five seconds by the generator card. The analog output was then displayed on a Sony PVM-1343 monitor. The result was five seconds of broadcast quality (except for the impairment), full-resolution, NTSC video. In addition to storing the video sequences, the server was also used to run the experiment and collect data. A special-purpose program recorded each subject's name, displayed the video clips, and ran the experiment. After each test sequence was shown, the experiment program displayed a series of questions on a computer monitor and recorded the subject's responses in a subject-specific data file.

The experiments were run with one test subject at a time. The subjects were asked to wear any vision correction devices (glasses or contacts) that they would normally wear to watch television. Each subject was seated in front of the computer keyboard at one end of a table. Directly ahead of the subject was the Sony video monitor, located at or slightly below eye height for most subjects. The subjects were positioned at a distance of four screen heights (80 cm) from the video monitor. The subjects were instructed to keep their heads at this distance during the experiment, and their position was monitored by the experimenter and corrected when needed.

The course of each experimental session went through five stages: instructions, examples, practice, experimental trials, and interview. In the first stage, the subject was verbally given instructions. In the second stage, sample sequences were shown to the subject. The sample sequences represented the impairment extremes for the experiment and were used to establish the annoyance value range. The practice trials were identical to the experimental trials, except that no data were recorded. The practice trials were also used to familiarize the subject with the experiment. Twelve practice trials were included in this session to allow the subjects' responses to stabilize before the experimental trials begin. Subjects in the experiment were divided into two independent groups. The first group was composed of 23 subjects that performed *detection* and *annoyance* tasks. The second group was composed of 30 subjects that performed a *strength* task. Both groups watched and judged the same test sequences which consisted of 24 combinations of blocky, blurry, noisy, and ringing artifact signals at different strengths and proportions. The two groups viewed the same video sequences, but the instructions, training and tasks performed were different for each group.

The 'Annoyance' group was composed of 23 subjects. They were instructed to search each video for defective regions. After each video was presented, subjects were asked two questions. The first question was 'Did you see a defect or impairment?' If the answer was 'no', no further questions were asked. If the answer was 'yes', the subject was asked 'How annoying was the defect?.' To answer this, the subject entered a value between '0' and '100', where '0' meant that the defect was not annoying at all and '100' that is was as annoying as the worst example in the training section. A defect half as annoying should be given 50, and any twice as annoying 200 and so forth. Although we tried to include the worst test sequences in the sample set, we acknowledge the fact that the subjects might find some of the other tests clips to be more annoying and specifically instruct them to go above 100 in that case.

The 'Strength' group was composed of 30 subjects. They were instructed to search each video for impairments that might contain up to four different artifacts – blocking, blurring, noisiness, and ringing. In the sample stage, we showed the original videos and examples of videos with the four artifacts by themselves. After each video was played, the subjects were asked to rate the strength of each artifact using one of four scale bars. Each bar was labeled with a continuous scale (0–10). The subject was never explicitly asked if an impairment was seen. Instead, all four of the scale bars were initialized to zero and subjects were instructed not to enter any value if no defect was seen.

At the end of the experimental trials, we asked the test subjects for qualitative descriptions of the defects that were seen. The qualitative descriptions helped in the design of future experiments.

4. DATA ANALYSIS

We used standard methods[9] for analyzing the annoyance judgments provided by the test subjects. We first computed two measures: the Total Squared Error (TSE) and the Mean Annoyance Value (*MAV*) for each test sequence. The TSE is our objective error measure and is defined as:

$$\text{TSE} = 1/N \sum_{i=1}^{N} (Y_i - X_i)^2 \tag{5}$$

where Y_i is i-th pixel value of the test sequence, X_i is the corresponding pixel of the original sequence, and N is the total number of pixels in the video. The *MOS* is our subjective error measure and is calculated by averaging the annoyance levels over all observers for each video:

$$MOS = 1/M \sum_{i=1}^{M} S(i) \tag{6}$$

where *S (i)* is the annoyance level reported by the *i*-th observer. *M* is the total number of observers. The data gathered from subjects in the 'Annoyance' group, the MOS data gathered provided one *MOS* value for each test sequence - the the Mean Annoyance Values (*MAV*). The data gathered from subjects in the 'Strength' group provided four *MOS* values for each test sequence – the Mean Strength Values (*MSV*s) for blockiness, blurriness, noisiness, and ringing, i.e., MSV_{block}, MSV_{blur}, MSV_{noise}, and MSV_{noise}. The average values of the *MAV* and the *MSV*s for all videos are shown in columns 5-9 of Table 1.

Figures 3 and 4 show the bar plots for MSV_{block}, MSV_{blur}, MSV_{noise}, and MSV_{noise}. Each graph shows the *MSV*s obtained for each of the originals. The coefficients *a*, *b*, *c*, and *d* (see eq.4) corresponding to the physical strengths of blockiness, blurriness, noisiness, and ringing, respectively, are shown over each graph. As can be seen from Figure 3 and 3, for the test sequences with only one type of artifact, the highest *MSV*s were obtained for the corresponding artifact. For example, for the test combinations 2, 3, 5 and 9 (Figures 3 (b), (c), (e), and (i)) corresponding to videos with only one type of synthetic artifact signal (blockiness, blurriness, noisiness, or ringing), the highest *MSV*s were obtained for the corresponding pure artifact, while the other three types of artifact signals received small values. In general, the subjects were able to identify the artifact strength proportion. *MAV*s are the highest for videos that contain noisy artifact signals (see Table 1). Combination number 1 (Figure 3 (a)) corresponds to the original videos. Again, the values for the average of *MAV*s and *MSV*s corresponding to the originals are not zero, indicating that subjects reported that these videos contained some type of impairment and annoyance levels different from zero.

We performed an ANOVA test on the data from combinations 1-16 to investigate the effects of the variables artifact signal strength (*a*, *b*, *c*, and *d*) and 'original' on the MAV. Table 2 shows the ANOVA results for the main effects and interactions among terms (columns 2-5 of Table 1). The results show that all artifact signals have a significant effect on MAV ($P < 0.05$). Regarding the interactions among the artifact signal strengths and originals, the results showed an interaction between 'original' and *c* (noisy), 'original' and *b* (blurry), *a* (blocky) and *c* (noisy), and *b* (blurry) and *c* (noisy).

Our principal interest in measuring the artifacts' strength was to investigate the relationship between the perceptual strengths of each type of artifact and the overall annoyance. In other words, we wanted to predict the *MAV* from the 3 *MSV*s (MSV_{block}, MSV_{blur}, MSV_{noise}, and MSV_{noise}). To verify if it was possible to find such a model, we used a Minkowski metric to model the annoyance of video impairments as a combination rule of blockiness, blurriness, noisiness, and ringing *MSV*s.[11] From previously experiments we found that the perceptual strengths of artifacts are weighted differently in the determination of overall annoyance.[12,13] Therefore, we modified the traditional Minkowski metric expression by adding scaling coefficients to each artifact term:

$$Y_p = \left[\alpha \cdot (MSV_{block})^p + \beta \cdot (MSV_{blur})^p + \gamma \cdot (MSV_{noise})^p + \nu \cdot (MSV_{block})^p \right]^{1/p} \tag{7}$$

where Y_p is the predicted annoyance. Note that the strengths here are perceived strengths, not physical strengths.

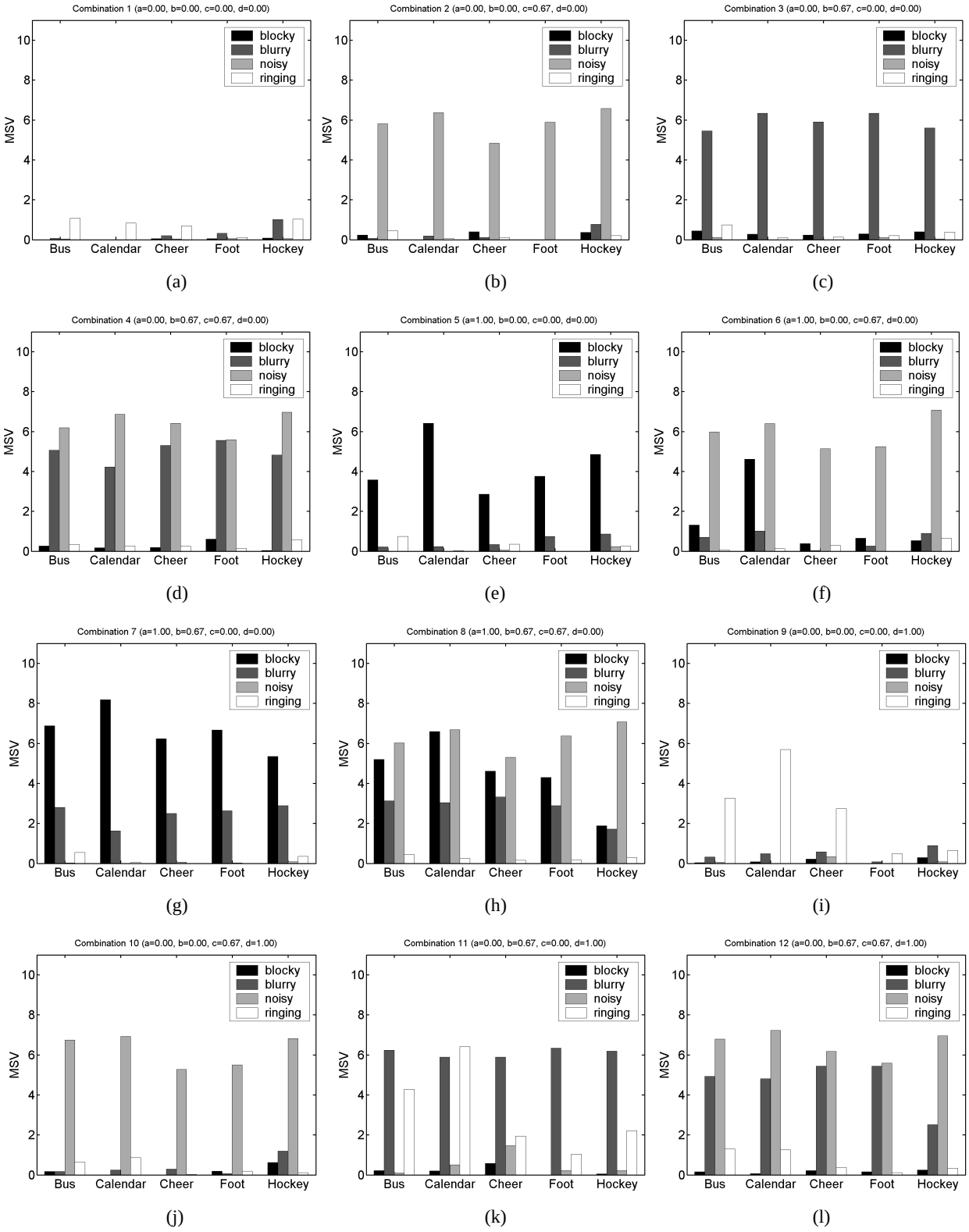


Figure 3 MSVs bar plots for combinations 1-12.

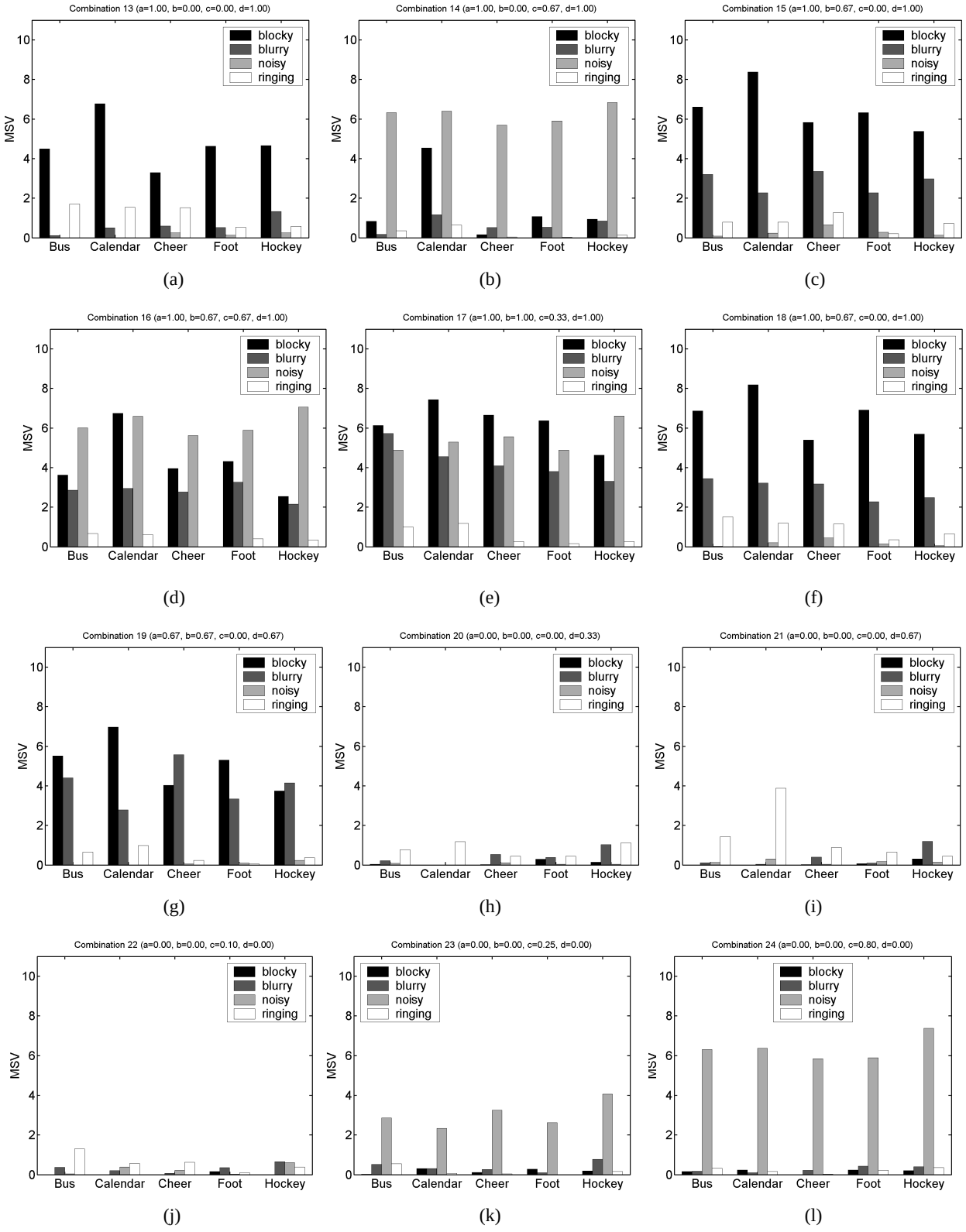


Figure 4 MSVs bar plots for combinations 13-24.

Table 2 ANOVA table for factorial test (combinations 1-16). Statistically significant terms ($P < 0.05$) are marked with a symbol '*'.

Source	Sum. Sq.	d.f.	Mean Sq.	F	Prob > F
c	15066.0	1	15066.0	393.47	0*
b	16871.6	1	16871.6	440.62	0*
a	2978.6	1	2978.6	77.79	0*
d	264.8	1	264.8	6.92	0.0114*
Original	2504.5	4	626.1	16.35	0*
c*a	277.9	1	277.9	7.26	0.0096*
c*b	675.8	1	675.8	17.65	0.0001*
c*d	22.9	1	22.9	0.60	0.4429
c*original	722.8	4	180.7	4.72	0.0027*
b*a	4.2	1	4.2	0.11	0.7434
b*d	44.9	1	44.9	1.17	0.2842
b*original	885.9	4	221.5	5.78	0.0007*
a*d	2.6	1	2.6	0.07	0.7973
b*original	1109.9	4	277.5	7.25	0.0001*
d*original	271.9	4	68.0	1.78	0.1489
Error	1876.2	49	38.3		
Total	43580.5	79			

Using a nonlinear fitting procedure, we fitted the data gathered from the psychophysical experiment in order to obtain a 'predicted' overall annoyance from the perceptual strength measurements MSV_{block} , MSV_{blur} , MSV_{noise} , and MSV_{noise} . The fit procedure returned optimal values for p (Minkowski exponent), and α, β, γ, and ν (Minkowski scaling coefficients corresponding to blockiness, blurriness, noisiness, and ringing, respectively). The advantage of this 'modified' Minkowski metric is that it provides a quantitative measure for the importance of each type of artifact to the overall annoyance.

Tables 3 summarizes the results of the Minkowski fit obtained for all test sequences and the data set containing all test sequences. Figure 5 depicts the plot of the *MAV* (obtained from the subjects) versus Predicted Mean Annoyance Value (*PMAV*) corresponding to the data set containing all test sequences. This fit is good ($r = 0.96$,$P = 0$) and the optimal value found for the Minkowski coefficient (p) is 1.03 and the scaling coefficients are $\alpha = 5.48$, $\beta = 5.07$, $\gamma = 6.08$, and $\nu = 0.84$. It is interesting to notice from Table 3 that the coefficients for ringing (ζ) are all very small ($0 \leq \zeta \leq 1.65$) implying that the ringing artifact is the artifact with smaller weight. This can be observed also in Table 2 that contained the results of the ANOVA test. If we choose a smaller confidence interval for the ANOVA, for example 99% ($P < 0.01$) instead of 95% ($P < 0.01$), ringing would not have a statistically significant effect on *MAV*. However, it should be remembered that the perceptual strengths of our ringing artifacts were relatively low and interactions may have reduced their contribution to annoyance below what it would be if they were high.

In Table 3, values of the Minkowski power (p) are all between 1 and 1.2. Based on these results, we varied the value of p in the range from 0.9 to 1.3 and repeated the fitting procedure for each one of these values. A model comparison test[8] showed that there is no significant statistical difference between the more generic model (Minkowski) and the simpler model with p constant, if p is in the interval [1.00, 1.25]. From this range, we are particularly interested in the results for $p = 1$ (linear model) that are shown in Table 4. Figure 6 depicts the plot of the *MAV* versus *PMAV* obtained from the linear model corresponding to the data set containing all test sequences. The fit is also reasonably good ($r \geq 0.91$ and $P \sim 0$). Again, we notice that the coefficients for ringing (ζ) are very small ($0 \leq \zeta \leq 1.68$). These results are similar to the our previous results[12] that showed annoyance models using linear model and the Minkowski model have the same performance according to a model comparison test.

5. CONCLUSIONS

The results of this experiment showed that the perceptual strengths of blockiness, blurriness, ringing, and noisiness signals were roughly correctly identified. Performing an ANOVA test, we found that, besides the 'original', all artifact

signal strengths (*a*, *b*, *c*, and *d*) had a significant effect on MAV ($P < 0.05$). The ANOVA also indicated that there are interactions among some of the artifact signal strengths and the original. Annoyance models were found by combining the perceptual strengths (*MSV*) of the individual artifacts using a Minkowski metric and a linear model. For the set containing all test sequences, the fit using the Minkowski metric returned a Minkowski exponent (*p*) equal to 1.03 and coefficients 5.48, 5.07, 6.08, and 0.84 corresponding to blockiness, blurriness, noisiness and ringing, respectively. For the linear model, the results were equally good and returned coefficients 5.1, 4.75, 5.67, and 0.86 corresponding to blockiness, blurriness, noisiness, and ringing, respectively. A comparison between the Minkowski metric and linear model showed that there is no statistical difference between these two models. Therefore, in spite of interactions, the linear model provids a reasonably good description of the relation between individual defect strengths and overall annoyance.

Table 3 Results from the Minkowski fit**.**

Videos	*p*	α	β	γ	ν	Residuals	*r*	*t* value	*P* value
Bus	0.85	3.42	3.32	3.77	0.43	25.24	0.96	15.26	0
calendar	1.1	7.79	6.29	7.52	1.65	39.03	0.92	11.14	0
Cheer	0.97	4.19	4.66	4.91	0	20.89	0.96	17.04	0
Foot	1.2	6.15	10.91	7.4	0.02	31.31	0.92	10.69	0
Hockey	1.08	5.93	4.19	7.96	0	24.13	0.95	13.86	0
All	1.03	5.48	5.07	6.08	0.84	79.36	0.95	13.86	0

Table 4 Results from the Minkowski fit – linear case ($p = 1$).

Videos	*p*	α	β	γ	ν	Residuals	*r*	*t* value	*P* value
Bus	1	5.01	4.9	5.43	0	26.84	0.95	14.06	0
calendar	1	6.00	4.98	5.78	1.54	39.69	0.92	11.13	0
Cheer	1	4.59	5.03	5.30	0	20.96	0.96	16.93	0
Foot	1	3.75	6.48	4.68	0	31.2	0.91	10.59	0
Hockey	1	4.74	3.77	6.54	0	24.39	0.95	13.84	0
All	1	5.10	4.75	5.67	0.86	79.44	0.95	13.84	0

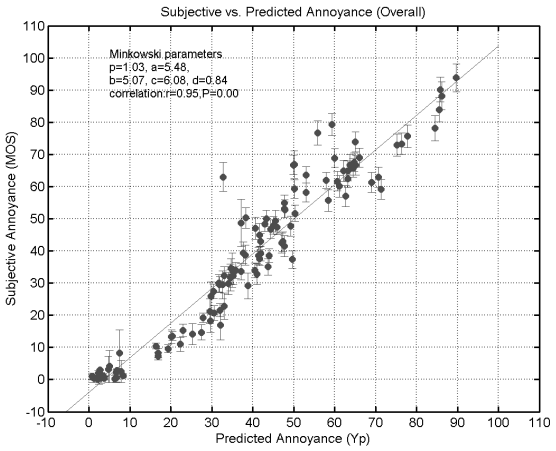


Figure 5: Subjective vs. Predicted Annoyance for videos. Results of Mikowski fitting: $p = 1.03$, $b = 5.07$, $c = 6.08$, $d = 0.84$.

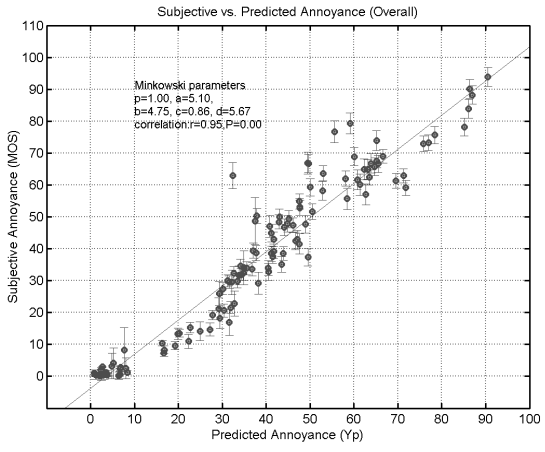


Figure 6: Subjective vs. Predicted Annoyance all for all videos. Results of Mikowski fitting $a = 5.48$, with $p = 1.0$, $a = 5.10$, $b = 4.75$, $c = 5.67$, $d = 0.86$.

ACKNOWLEDGMENTS

This work was supported in part by CAPES-Brazil, in part by a National Science Foundation Grant CCR-0105404, and in part by a University of California MICRO grant with matched supports from Philips Research Laboratories and Microsoft Corporation.

REFERENCES

1. M. Yuen, and H.R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, **70**, pp. 247-278, 1998.
2. Recommendation ITU-R BT.500-930, "Principals of a reference impairment system for video," ITU-T 1996.
3. A. J. Ahumada, Jr. and C. H. Null, "Image quality: A multimensional problem," *Digital Images and Human Vision*, pp. 141-148, 1993.
4. S. Wolf, "Measuring digital video transmission channel gain, level offset, active video shift, and video delay," ANSI T1A1 T1A1.5/96-110, May 31, 1996.
5. S. Mitra, *Digital signal processing: a computer based approach*, 2nd ed. New York, NY, USA: McGraw-Hill, 2001.
6. J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.
7. V. Q. E. Group, "VQEG Subjective Test Plan," http://ftp.crc.ca/test/pub/crc/vqeg/ 1999.
8. W. Hays, *Statistics for the social sciences*, Madison Avenue, New York, N.Y.: LLH Technology Publishing, 1981.
9. ITU Recommendation BT.500-8, "Methodology for subjective assessment of the quality of television pictures," 1998.
10. M.S. Moore, "Psychophysical measurement and prediction of digital video quality," Ph.D. thesis, University of California, Santa Barbara, June 2002.
11. H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments," Proc. of the SPIE, Human Vision and Electronic Imaging III, San Jose, CA, vol. 1666, pp. 16-26, January 1992.
12. M. C. Q. Farias, J. M. Foley, and S. K. Mitra, "Perceptual Contributions of Blocky, Blurry and Noisy Artifacts to Overall Annoyance," IEEE International Conference on Multimedia & Expo, Baltimore, MD, USA, pp. 529-532, 2003.
13. M. C. Q. Farias, M. S. Moore, J. M. Foley, and S. K. Mitra, "Perceptual Contributions of Blocky, Blurry, and Fuzzy Impairments to Overall Annoyance," SPIE Human Vision and Electronic Imaging, San Jose, CA, USA, pp.109-120, 2004.