

# DETECTABILITY AND ANNOYANCE OF SYNTHETIC BLURRING AND RINGING IN VIDEO SEQUENCES\*

*Mylène C. Q. Farias<sup>a</sup>, Sanjit K. Mitra<sup>a</sup>, and John M. Foley<sup>b</sup>*

<sup>a</sup>Department of Electrical and Computer Engineering

<sup>b</sup>Department of Psychology

University of California, Santa Barbara

Santa Barbara, CA 93106 USA

## ABSTRACT

One approach for studying digital videos impairments is to work with synthetic artifacts which look like real impairments, yet are simpler, purer and easier to describe. In this paper, we created synthetic ringing and blurring and inserted them in short video sequences. In a psychophysical experiment we measured the probability of detection and the annoyance value of these artifacts as a function of their total squared error. Although ringing occurs only near edges and blurring can occur over wide areas of the images, there is no consistent difference between either the thresholds or mid-annoyance strengths. There are, on the other hand, interactions between the specific video and artifact type in the determination of these values. Mid-annoyance strength was found to be highly correlated with threshold. Also, we combined ringing and blurring to produce mixed artifacts. Their thresholds and mid-annoyance strengths tend to be intermediate between those of the individual artifacts. Their annoyance value is well predicted by a weighted sum of the annoyance values for blurring and ringing with weights of approximately 0.6 and 0.4, respectively.

## 1. INTRODUCTION

An impairment or a defect is defined as a perceived flaw introduced into an image or video during capture, transmission, storage, and/or display, as well as by any image processing algorithm (e.g. enhancement, compression) that may be applied to the images. Impairments can be very complex in their physical description and also in their perceptual description. Most of them have more than one perceptual feature or perceived artifact. Examples of perceived artifacts introduced by digital systems are blurriness, noisiness, ringing, and blockiness [1].

Many video quality models have been proposed, but little work has been done on studying and characterizing the individual artifacts found in digital video applications. A study of the individual perceived artifacts is necessary since their relationship with the overall quality is not known. A characterization of the most common artifacts is also an important step in the design of a multi-metric quality measurement system [2].

One approach for studying impairments is to work with synthetic artifacts that look like “real” artifacts, yet are simpler, purer, and easier to describe. Such artifacts are necessary components of the kind of reference impairment system recommended by the ITU-T for the measurement of image quality [3] and offer advantages for experimental research on video quality. 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. This control makes it possible, for example, to study the importance of each type of artifact for human observers.

The goal of this study was to examine properties of two synthetic artifacts - blurring and ringing. To this end, we inserted ringing and blurring artifacts in short video sequences and performed a psychophysical experiment to estimate their visibility and annoyance.

## 2. GENERATION OF SYNTHETIC ARTIFACTS

Ringing is fundamentally related to the Gibb’s phenomenon. 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 texture [1].

The ITU-T reference impairment system recommends generating ringing using a filter with ripples in the passband amplitude response, which creates an echo impairment [3]. The problem with this approach is that besides ringing, this procedure also introduces blurring and possibly noise. Since our goal is the generation of artifacts as pure as possible, we developed an algorithm for synthetically generating ringing. Our algorithm consisted of a pair of delay-complementary highpass and lowpass filters, related by the following relationship:

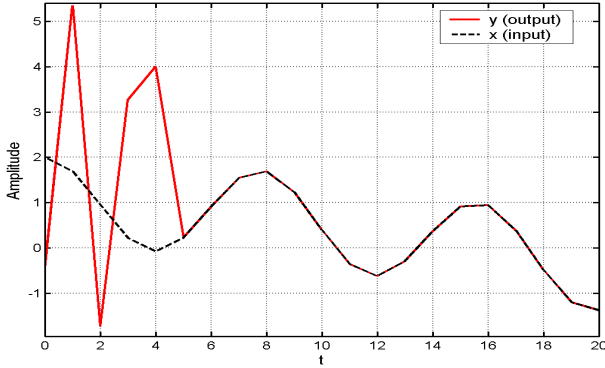
$$H(z) + G(z) = \beta \cdot z^{-n_0}, \quad (1)$$

where  $H(z)$  and  $G(z)$  are  $N$ -tap highpass and lowpass filters, respectively. We set  $\beta = 1$  and  $n_0 = N/2$  for this work. The output of our system is given by the following equation:

$$Y(z) = [H(z) + G(z)] \cdot X(z), \quad (2)$$

So, except for a shift,  $Y$  is equal to  $X$ , given that the initial conditions of both filters are exactly the same [4]. If, on the other hand, we make the initial conditions different, a decaying “noise” is introduced in the first  $N/2$  samples that resembles

\* 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 matching support from Philips Research Laboratories.



**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.

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 is:  $x = \cos(0.1 \cdot t) + \cos(0.8 \cdot t)$ .

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. We use the Canny algorithm [5] to detect the edges. The resulting effect is very similar to the ringing artifact found in compressed images, but without any blurring or noise.

Blurring is the reduction in sharpness of edges and spatial detail [1]. In compressed images blurring is often caused by trading off bits to code resolution and motion. Blurring artifacts can be easily generated by applying a symmetric, two-dimensional low-pass FIR (Finite Impulse Response) filter to the digital image array [3]. We used a  $5 \times 5$  mean filter in this experiment. Different filters with varying cut-off frequencies could be used to control the amount of blurriness introduced.

## 2. PSYCHOPHYSICAL EXPERIMENT

With the goal of estimating the visibility and annoyance of blurring and ringing, a psychophysical experiment was performed. 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-garden, and Hockey. These videos are commonly used for video experiments and are publicly available at the Video Quality Group website (<http://www.vqeg.org>). The second step was to generate videos in which one type of artifact dominated and produced a relatively high level of annoyance. We also created a set of videos with equal proportions of blurring and ringing. Therefore, for each original video, 3 new videos were created:  $X_{blurry}$ , with only blurring,  $X_{ring}$ , with only ringing, and  $X_{comb}$ , with a combination of blurring and ringing.

The usual approach to subjective quality testing is to degrade a video by a variable amount and ask the test subjects for a quality rating [6]. Since both the type and the strength of artifacts vary from frame to frame and region to region, this method cannot be used to measure the visibility and annoyance produced by specific artifacts at specific strengths. To do this we use an experimental paradigm in which impairments are restricted to an isolated region (defect zone) of the video clip for a short time interval. The rest of the video is left in its original

state [7].

Finally, the test sequences,  $Y$ , were generated by linearly combining the original video with the video containing the impairment ( $X_{blurry}$ ,  $X_{ring}$ , or  $X_{comb}$ ) in different proportions, as given by the following equation:

$$Y = X + r \cdot (X_I - X), \quad (3)$$

where  $X$  is the original video,  $X_I$  is the sequence with the impairment, and  $r$  is the strength parameter of the test sequence ( $r \geq 0$ ). Before adding them, the videos were transformed to the linear light domain using a gamma approximation. A total of 95 test videos were used in this experiment (5 originals  $\times$  6 strengths  $\times$  3 types of impairments).

To create artifacts that were identical except for strength, we first created a single artifact using the algorithm. This was combined with the original video using eq. 3 with different values of  $r$ . The values of  $r$  were chosen to cover the ranges of both the psychometric and the annoyance functions in so far as possible, using results of a pilot study. However, we found that the highest strengths of ringing that we could produce without saturation had lower values of TSE and produced less annoyance than our highest strength blurring artifacts. 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 estimate of its parameters.

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. They were asked to wear any vision correcting devices (glasses or contacts) that they normally wear to watch television. There were five stages to the experimental session: instructions, training, practice, experiment, and interview. In the first stage, the subject was verbally given instructions. In the training stage, we showed sample sequences to the subject to establish the range for the strength and annoyance scales. In the practice stage, the subject carried out 8 practice trials to allow the responses to stabilize. At the interview stage, we asked the test subjects for qualitative descriptions of the defects that were seen.

The main experiment was performed with the complete set of test sequences presented in random order. The test subjects were instructed to search each video for defective regions. After each video was played the 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 instructed to enter a positive numerical value indicating how annoying the defect was. Any defect half as annoying should be given 50, as annoying 100, twice as annoying 200 and so forth.

## 4. DATA ANALYSIS

We used the standard methods [6] for analyzing the visibility and annoyance judgments provided by the test subjects. We first computed two measures for each test sequence: Probability of Detection ( $PD$ ), and Mean Annoyance Value ( $MAV$ ).  $PD$  was estimated by dividing the number of subjects who detected the artifact by the total number of subjects.  $MAV$  is calculated by averaging the annoyance levels over all observers for each video.

The visibility threshold is defined as the logarithm of the TSE such that the impairment is seen by 50% of the subjects. To

estimate this threshold, the probability of detection data for each impairment were fitted using the Weibull function [6], which is defined as:

$$P(x) = 1 - 2^{-(Sx)^k}, \quad (4)$$

where  $P(x)$  is the probability of detection,  $x$  is the logarithm of the Total Squared Error (TSE),  $S$  is the sensitivity, and  $k$  is a constant that determines the slope of the transition. The 50% detection threshold is simply  $x_T = 1/S$ . Columns 2, 3, and 4 of Table 1 show the psychometric function fitting parameters ( $x_T$  and  $k$ ) and the sum of squared fitting residuals. The empty spaces correspond to cases where a fit was not possible because more than 50% of subjects detected the weakest impairment.

The mean annoyance values for each test sequence were fitted with the standard logistic function [6]:

$$y = y_{\min} + (y_{\max} - y_{\min}) / (1 + \exp(-(x - \bar{x}) / \beta)) \quad (5)$$

where  $y$  is the predicted annoyance and  $x$  is the logarithmic error energy. The parameters  $y_{\min}$  and  $y_{\max}$  were set to '0' and '100'. The parameter  $\bar{x}$  translates the curve in the  $x$ -direction and the parameter  $\beta$  controls the steepness of the curve. Columns 5, 6, and 7 of Table 1 show the annoyance function fitting parameters ( $\bar{x}$  and  $\beta$ ), and the sum of squared fitting residuals. The fits of both functions were generally quite good.

In Table 1, it can be seen that some of the original videos had higher visibility thresholds for ringing, while others had higher thresholds for blurring. The same is true for the annoyance parameter  $\bar{x}$ . This is shown in Figures 2-5 which depict the psychometric and annoyance functions for the test videos 'Cheerleader' and 'Calendar'.

From Table 1 it is clear that the values of  $x_T$  and  $\bar{x}$  vary quite a bit over the different impairments (2 log units for  $x_T$  and 1.4 log units for  $\bar{x}$ ). However, these differences are due primarily to differences in the original videos and not due to the type of impairment inserted in them. To show this, we performed an ANOVA analysis to test for the main effects ('original' and 'impairment type') on fitting parameters. Table 2 shows the  $P$  values obtained from this analysis. As expected, 'impairment type' did not have a significant effect on any of the fitting parameters. On the other hand, 'original video' had a significant effect on three out of the four parameters. Further, an interaction between 'original video' and 'impairment type' seems apparent in the data, although our two-way ANOVA does not test for this. In bus, calendar, and flower ringing is more visible and annoying; in cheerleader and hockey blurring is more visible.

It is interesting to observe that the same videos which had lower visibility thresholds for ringing, also had lower  $\bar{x}$  for ringing. This indicates a positive correlation between these two factors. We calculated the Pearson correlation [8] between  $x_T$  and  $\bar{x}$ . As expected, these parameters are highly correlated ( $R^2 = 0.977$ ) and are related by the following expression  $\bar{x} = 1.29 x_T - 2.08$ . This implies that if we know the visibility thresholds of these impairments we can estimate their annoyance. Moore [7] also found a high correlation (0.971) between the  $x_T$  and  $\bar{x}$  values for MPEG blocky/blurry and fuzzy artifacts and the relation  $\bar{x} = 1.15 x_T$ .

We performed a linear regression analysis on our annoyance data to test if we could predict the  $MAVs$  of the combined artifacts from the  $MAVs$  of the ringing and blurring. The results show a very significant correlation ( $R^2 = 0.96$ ; t-test,  $P = 1.8 \cdot 10^{-19}$ )

**Table 1:** Annoyance and visibility fitting parameters.

Group	$x_T$	$k$	$\Sigma(r_i)^2$	$\bar{x}$	$\beta$	$\Sigma(r_i)^2$
BlurBus	4.17	11.25	0.04	4.77	0.23	9.88
BlurCalendar	4.17	15.63	0.01	4.86	0.20	6.16
BlurCheer	3.45	7.13	0.06	4.49	0.37	22.2
BlurFlower	4.00	8.82	0.08	4.76	0.23	11.39
BlurHockey	2.38	3.41	0.07	3.48	0.28	10.34
MixBus	4.17	15.94	0.06	4.82	0.23	5.67
MixCalendar	4.17	16.52	0.03	4.84	0.21	4.61
MixCheer	3.57	6.73	0.10	4.55	0.31	7.58
MixFlower	4.00	14.09	0.02	4.71	0.21	7.87
MixHockey				3.56	0.37	10.38
RingBus	3.70	8.44	0.13	4.40	0.32	1.69
RingCalendar	3.45	9.11	0.06	4.25	0.34	4.57
RingCheer	4.00	9.15	0.18	4.65	0.29	6.85
RingFlower	3.45	14.81	0.06	4.11	0.23	8.41
RingHockey	2.70	3.63	0.17	3.78	0.50	0.00

**Table 2:**  $P$  values obtained from the ANOVA analysis on the annoyance and visibility fitting parameters (Table 1).

Annoyance	$\bar{x}$		$\beta$	
	original	impairment	original	impairment
$P$	0.003	0.2348	0.609	0.1495
Visibility	$S$		$k$	
	original	impairment	original	impairment
$P$	0.057	0.3584	0.0152	0.7084

among these values. We found coefficients of 0.631 and 0.4058 for blurring and ringing, respectively. Figure 6 depicts the plot of the observed  $MAV$  (combined artifacts) versus the predicted  $MAV$ . The line in the graph corresponds to  $y=x$ .

We also fitted a Minkowski metric to the data. We obtained a Minkowski factor of 1.017 and coefficients 0.633 and 0.411 for blurring and ringing, respectively. These results were very close to the linear fit, indicating that the linear model had a better fit with our data.

## 5. CONCLUDING REMARKS

A new algorithm for creating synthetic ringing was proposed. The resulting artifacts are 'relatively pure' and provide a valuable tool for studying the characteristics of this type of artifact. A psychophysical experiment was performed to compare the annoyance and visibility of ringing and blurring.

We used standard methods to find the psychometric and annoyance curves. We found that both of these curves have the same form as has been used for MPEG artifacts. Although ringing occurs only near edges and blurring can occur over wide areas of the images, there is no consistent difference between either the thresholds or mid-annoyance strengths. There are, on the other hand, interactions between the specific video and artifact type in the determination of these values. What this means is that, although high strengths of ringing rarely occur in practice, annoyance depends on ringing strength in essentially the same way that it depends on the blurring strength.

Mid-annoyance strength is highly correlated with the threshold, and the relation can be described by a linear function. We also found that a simple linear model with no interactions predicted how ringing and blurring combine to determine overall

annoyance. The linear fit presented a high correlation and the weights for blurring and ringing are approximately 0.6 and 0.4, respectively.

### 6. REFERENCES

[1] M. Yuen, and H.R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, 70 (1998), pp. 247-278.

[2] K. T. Tan and M. Ghanbari, "A multi-metric objective picture-quality measurement model for MPEG video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 1208-13, 2000.

[3] Recommendation ITU-R BT.500-930, "Principals of a reference impairment system for video," ITU-T 1996.

[4] S.K. Mitra, *Digital Signal Processing: A Computer Based Approach*, 2nd Edition, 2000, McGraw-Hill.

[5] J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol.8, No. 6, 1986.

[6] ITU Recommendation BT.500-8, "Methodology for Subjective Assessment of the Quality of Television Pictures," 1998.

[7] M.S. Moore, *Psychophysical Measurement and Prediction of Digital Video Quality*, Ph.D. thesis, University of California Santa Barbara, June 2002.

[8] W. Hays, *Statistics*, 2nd Edition, 1981, CBS College Publishing.

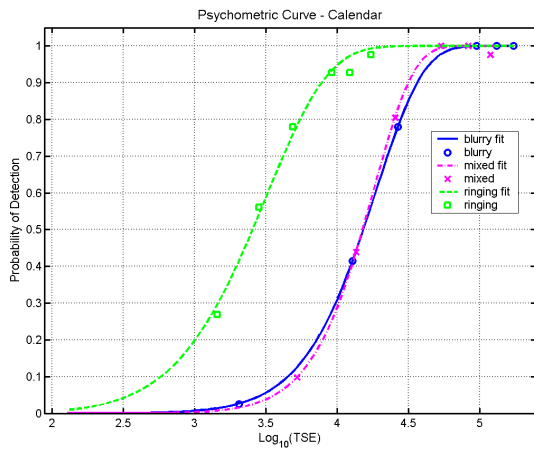


Figure 2. Psychometric curve for the sequence Calendar.

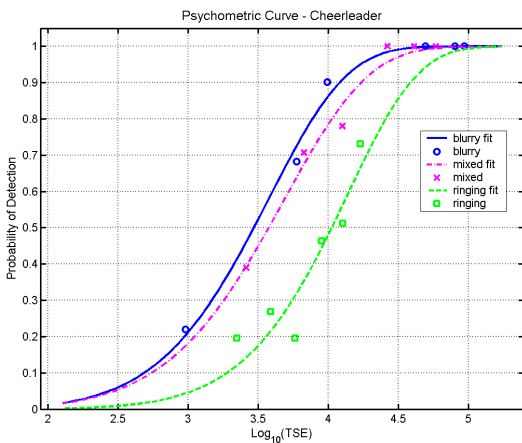


Figure 3. Psychometric curve for the sequence Cheerleader.

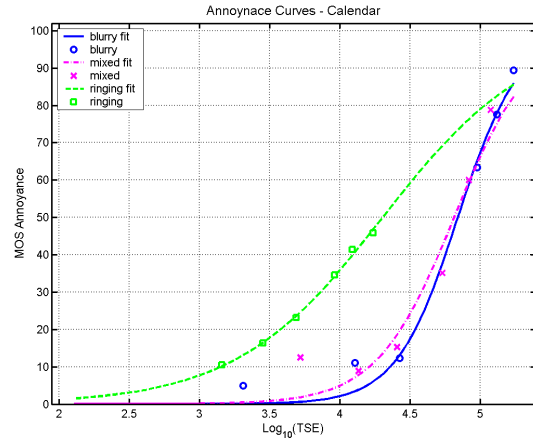


Figure 4. Annoyance curve for the sequence Calendar

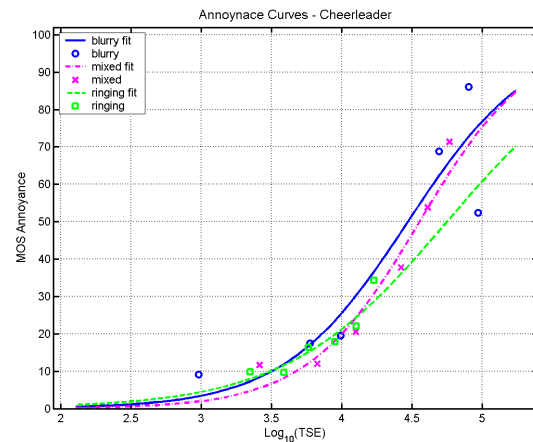


Figure 5. Annoyance curve for the sequence Cheerleader.

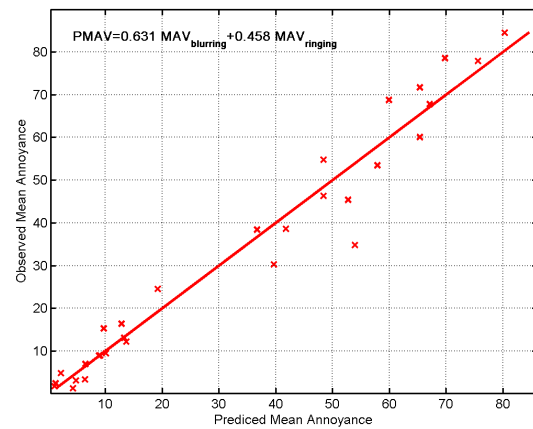


Figure 6. Predicted MAV of combined artifacts.