Mosquito noise in MPEG-compressed video: test patterns and metrics

Charles Fenimore, John Libert, and Peter Roitman
National Institute of Standards and Technology[1]
Gaithersburg, Maryland 20899-8114

ABSTRACT

Mosquito noise is a time dependent video compression impairment in which the high frequency spatial detail in video images having crisp edges is aliased intermittently. A new synthetic test pattern of moving spirals or circles is described which generates mosquito noise (MN) under Motion Pictures Expert Group (MPEG) compression. The spiral pattern is one of several NIST-developed patterns designed to stress specific features of compression based on motion estimation and quantization. The "Spirals" pattern has several spirals or circles superimposed on a uniform background. The frames are filtered to avoid interline flicker which may be confounded with MN. Motion of the spirals and changing luminance of the background can be included to reduce the correlation between successive frames. Unexpectedly, even a static pattern of spirals can induce mosquito noise due to the stochastic character of the encoder.

We consider metrics which are specific to the impairment being measured. For mosquito noise, we examine two separable detectors: each consists of a temporal (frame-to-frame) computation applied to the output of a spatial impairment detector which is applied to each frame. The two spatial detectors are: FLATS, which detects level 8x8-pixel image blocks; and the root-mean-square (RMS) applied to the image differences between original and compressed frames. The test patterns are encoded at low bit rates. We examine the measured mosquito noise as a function of the Group-of-Pictures (GOP) pattern in the MPEG-2 encoding and find the GOP structure defines the periodicities of the MN.

Keywords: digital video compression, quality metrics, test patterns,mosquito noise, flats, time dependent, stochastic process.

1. INTRODUCTION

At low bit rates, MPEG-2 (Motion Pictures Experts Group) video compression induces a variety of impairments which are characteristic of block transform-based coders, such as image blocking and blurring. These specific impairments are found in both MPEG- and JPEG- (Joint Photographic Experts Group) compressed moving imagery. The measurement of video impairments (or artifacts in the parlance) follows two distinct approaches.

The first approach is to quantify specific impairments such as blocking, blurring, and ringing. For single frames extracted from MPEG-compressed video sequences, Fenimore, van de Grift, and Field [1] described a blocking detector (FLATS defined in Section 5) which is effective in measuring blocking in I-frames. In a subsequent investigation, Libert and Fenimore [2] found that a modified FLATS detector and a discrete cosine transform (DCT) error detector are equally effective in finding the threshold for subjective perception of blocking in I- B- and P-frames. In the case of JPEG compression, Meesters and Martens [3] have measured and correlated the appearance of all three impairments. They report that a single parameter quantifies the subjective appearance of all three impairments in JPEG compressed images. In a recent paper de Ridder [4] extracts independent measures of blocking, blurring, and ringing in JPEG-compressed images and is able to measure the relative contributions of each. It is interesting to note that the first two blocking detectors mentioned above are "single-ended" in requiring input of only the processed video.

In the second approach to quality measurement, an overall score is determined which incorporates all effects contributing to the impairment. Such measurements are described in work of Tong, Heeger, and van den Lambrecht [6], Lubin [5], Watson [7], Winkler [8], and others. Typically, these global quality metrics are double-ended in that they require input of the original, unprocessed video as well as the compressed video.

[1] Electricity Division, National Institute of Standards (NIST), Technology Administration, U. S Department of Commerce.

A definition of mosquito noise [9] appears in work reported to the International Telecommunications Union (ITU): *"Distortion concentrated at the edges of objects, and further characterized by its temporal and spatial characteristics. Sometimes associated with movement, characterized by moving artifacts and/or blotchy noise patterns..*" We adopt a different definition emphasizing the intermittency of mosquito noise (Section 3).

The generation of MN and other impairments of a known magnitude is addressed in P930, but not their measurement. Indeed, there appears to be little work on the measurement of specific dynamic MPEG impairments. For global quality metrics, Winkler [8] and Watson [10] have explicitly addressed the measurement of temporal effects. Referring to various models for the temporal mechanism in human visual perception, these authors implement infinite impulse response filter(s) which approximate such models.

In the present study, we describe a technique for generating patterns of spirals and circles. The patterns are loosely modeled on high contrast patterns seen in such video test clips as "Mobile and Calendar" (sample frame in Fig.1). The NIST spiral patterns are mathematically defined and then rendered on an image grid. In order to avoid aliasing associated with the sharp edges of the image a finite impulse response (FIR) filter is applied to the image (Section 4).

Although MN is a temporal phenomenon, we find that it is generated in static (unmoving) spirals under MPEG compression (Section 5). The intermittency in noise is associated with the frame-to-frame non-uniformity of MPEG compression. MPEG organizes a video sequence into Groups of Pictures (GOP) having I-, P-, and B-frames. I-frames are coded independently; P-frames depend on I-frames; and B-frames depend on I- and P-frames. The intermittency is quantified by either of two new metrics: one is based on the root-mean-squared (RMS) error of the compressed frames and the other on the FLATS measure of blocking. In each case, a simple temporal FIR filter is applied: F(z) = 1-z (Section 5.) The FLATS-based metric has higher sensitivity to MN than does the RMS-based metric. Both metrics exhibit the footprint of the GOP in the mosquito noise amplitude plots.

2. MPEG IMPAIRMENTS

The quantization and motion estimation stages of an MPEG encoder are the two main contributors to bit-rate reduction and so to impairment generation. At low bit rates, image blocking is a dominant impairment. Blocking arises from quantizing too coarsely the coefficients of the discrete cosine transform (DCT) and from failure of motion search to find good motion estimates.

2.1 I-frames and DCT compression blocks in MPEG2

The DCT and quantization stages of MPEG2 compression can introduce blocking impairments into video frames on the scale of the 8 x 8-pixel blocks into which each frame is decomposed. MPEG2 groups four DCT blocks into a single macroblock. Each macroblock is handled in one of two modes: intraframe (I-frame) compression mode in which the four DCT blocks making up the macroblock are encoded without reference to other frames in the video sequence and interframe mode (discussed below) in which motion estimation is used. For I-frames all of the macroblocks are DCT-encoded without motion estimation. Picture information is lost in quantizing the transform coefficients.

For an image with pixel values $s(p, q)$ on an *N* x *N* block, the 2-dimensional DCT, $S(j, k)$, is defined by:

$$S(j, k) = \frac{2}{N} \cdot C(j) \cdot C(k) \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} s(p, q) \cos\left(\frac{\pi(2p+1)j}{2N}\right) \cos\left(\frac{\pi(2q+1)k}{2N}\right), \text{ where}$$

and both *j* and *k* = 0 ... *N*-1. In MPEG-2, *N* is taken to be 8 [11].

$$C(k) = \begin{cases} 1/\sqrt{2} & \text{for } k = 0 \\ 1 & \text{otherwise} \end{cases}$$

The quantization of the coefficients, $S(j, k)$, occurs through integer division by the factors, $M_{QUANT} \cdot Q(j, k)$. The matrix *Q* is fixed while the parameter M_{QUANT} is set in a feedback loop to provide control of the bit rate. The quantization is coarser, that is $Q(j, k)$ is larger, for higher values of j and k (i.e. for higher frequencies). Similarly, as M_{QUANT} increases, the coefficients are represented with less resolution. This loss of resolution can produce visible image blocks. The blocking impairment detector which we have developed attempts to exploit the appearance of a large number of zero coefficients in quantized video frames.

2.2 P- and B-frames, motion estimation, and noise

In interframe mode the DCT is not applied to the original frame but to the residual image formed as the difference between the original frame and a motion-estimated frame. Doing so introduces a new class of blocking impairments. Although there are two types of interframe macroblocks, predicted (P-frames) and bidirectional (B-frames), in each case motion estimation is used to find an estimate of each 16 x 16 pixel macroblock. A macroblock in the encoded (or target) frame is compared with linear translates of equal-sized blocks in encoded frames which precede and/or follow it. The block which most closely approximates the target macroblock is used as an initial estimate of the target block. The associated translation gives the value of a motion vector.

DCT encoding is applied to the residual macroblock. Even in the absence of motion in the video, the encoder will quantize the motion-estimation residual of the original frame. Because the DCT is applied to a residual dominated by high frequency components, the structure of the blocking for inter-coded frames may differ from that of the I-frames.

3. WHAT IS MOSQUITO NOISE?

The VIRIS project (a Video Reference Impairment System developed by Bellcore [9]) has defined edge busyness and mosquito noise as follows:
edge busyness: *Distortion concentrated at the edges of objects, and further characterized by its temporal and spatial characteristics.*
mosquito noise: *Form of edge busyness distortion sometimes associated with movement, characterized by moving artifacts and/or blotchy noise patterns superimposed over the objects (resembling a mosquito flying around a person's head and shoulders).*

We take the point of view that mosquito noise is introduced into a video sequence by compression processes operating on a time scale corresponding to the length of a Group of Pictures. Thus, intermittency in the noise is akin to a periodicity in the impairment. It will be seen that for our test pattern, the amplitude of the metrics for image blocking and image error has a component at the scale of the GOP. That is our intermittency.

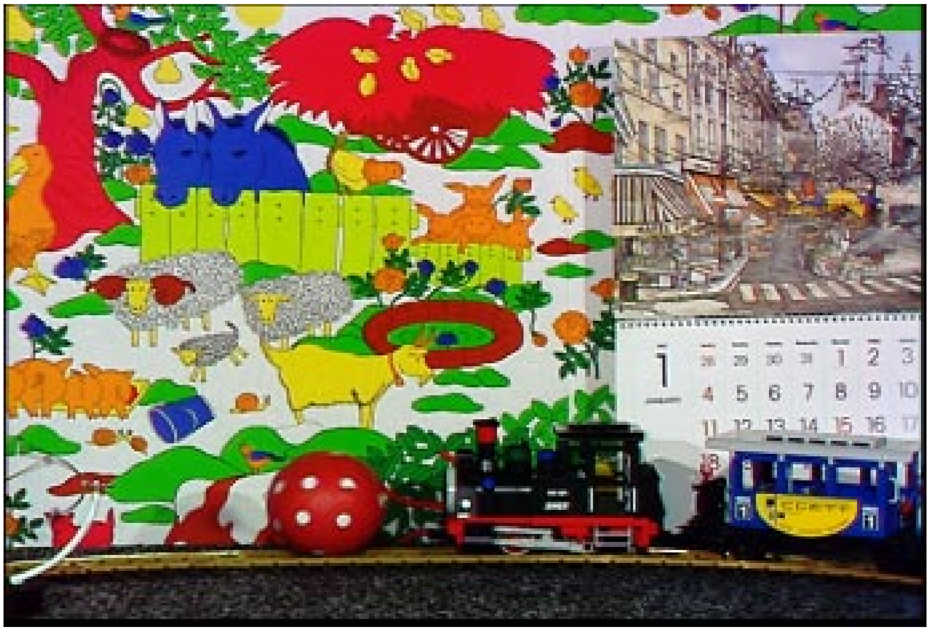

Figure 1: Mobile and Calendar is challenging to MPEG compression. Mosquito noise is produced in image regions with sharp edges, such as in the lettering.

4. TEST PATTERNS FOR GENERATING MOSQUITO NOISE

The classic Rec. 601 test clip, 'Mobile & Calendar' [14], has the edges associated with mosquito noise in the lettering of the calendar and the "wool" of the sheep, among other portions of the images. A sample frame is presented in Figure 1, as an illustration of materials challenging to MPEG encoders.

4.1 Synthetic video test pattern: Spirals

The synthetic pattern 'Spirals' (Figure 2, and a similar pattern 'Circles') are designed to emulate those features of natural, camera-captured video which stimulate the production of mosquito noise under MPEG compression. The spirals are defined mathematically by their center, outer radius, number of windings, and the width of the "brush". Representing the spirals on the image raster requires the use of filtering to avoid aliasing. For Spirals we apply two spatial filters: one filter applies sub-pixel sampling which reduces rastering and Moire' effects; the second filter is a low pass FIR filter to reduce flicker.

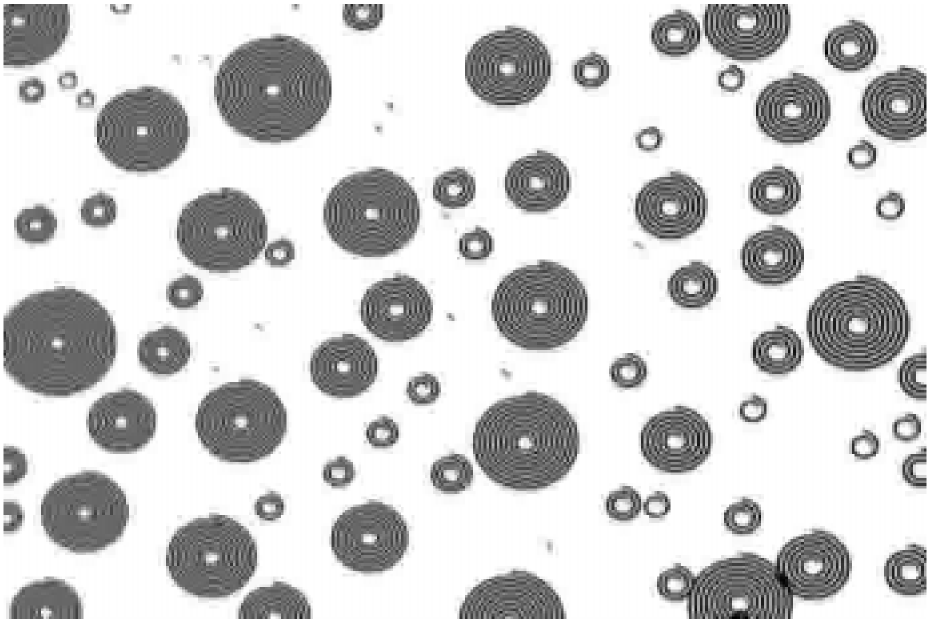

Figure 2: One frame from the Spirals pattern after MPEG-2 compression at 1.7 Mbits/second.

4.1.1 Math model for pattern: rendering to raster with sub-pixel sampling

Sub-pixel sampling emulates the capture of an image on a camera raster by generating intermediate luminance levels at pixels which lie on the transition between two regions.

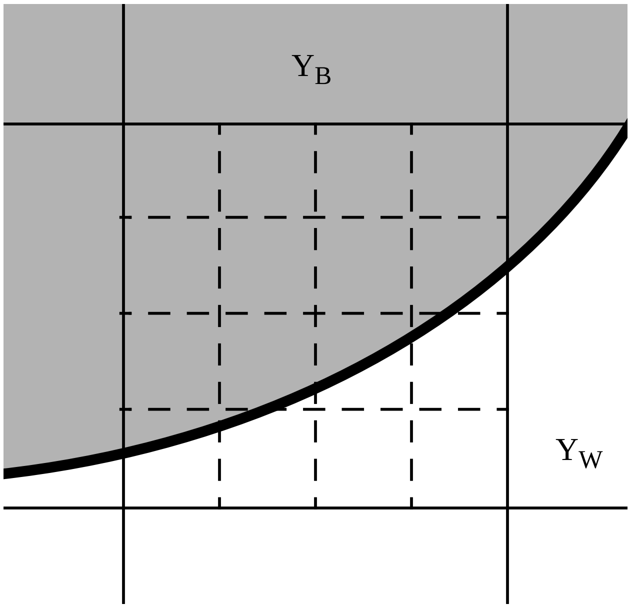

Figure3: In the “Spirals” image which is modeled as a bi-level image, pixels on the boundary between regions of constant luminance are rendered by averaging the luminances at the sub-pixel center points. This filtering avoids Moire’ aliasing and rastering.

In Figure 3, a pixel straddles the edge of the dark region in which the luminance is a constant, $Y_{B,}$ and the light region with luminance Y_W. The luminance, Y, of any pixel is the average of the two luminances with weight given by the fraction, f, of the sub-pixels centered in the Y_B region:
$Y = f * Y_B + (1-f) * Y_W$.

4.1.2 Filtering to meet requirements of the sampling theorem.

In the presence of sharp edges in our test images, the patterns may not meet the requirements of the sampling reconstruction theorem, that there be two samples per wavelength at the highest frequency. For this reason, the test patterns have been low-passed filtered. We examined a variety of filters and concluded that the simple FIR filter (.5, .5) applied both horizontally and vertically to each image, reduced interline flicker and other aliasing. In the vertical direction, this filter is the customary "line pairing" which is used to reduce interlacing flicker. The Spirals are motion blurred to reduce temporal aliasing and judder [9]. In general, the velocity of objects in video may be too high to satisfy the sampling theorem without unacceptable blurring[15].

5. GENERATING AND MEASURING MOSQUITO NOISE

The Test Model 5 MPEG-2 compression software was used for this study. The encoding and decoding parameters are discussed in the documentation for the package [12] and the choices made for this study are described in [1]. The focus of the present study is the correlation between these encoding parameters and the character of the induced mosquito noise.

We find that mosquito noise depends on the Group of Pictures (GOP) structure. GOP is specified by two indices (m, n), where m in the number of frames between successive I-frames in the GOP and n is the number of frames between successive I or P frames. For example, the following GOP indices have the indicated frame-types (IBP) shown below in at least two GOPs:

GOP indexing	GOP frames types sequence	
(3, 3)	IBB IBB	IBB IBB
(6, 3)	IBBPBB	IBBPBB
(6, 2)	IBPBPB	IBPBPB
(1, 1)	I I I I I I	I I I I I I

FLATS with local luminance adaptation threshold

Libert and Fenimore [2] have defined luminance-adapted FLATS as 8 x 8 blocks of pixels having constant luminance, Y_0, inside the block in either row, column, or both directions and differing from the 4 nearest neighboring blocks by a threshold amount (4). Formally, consider those 8 x 8 block cornered at image coordinate (J,K) , having pixels indexed by (j,k), j-J and k-K = 0 … 7. Select those blocks for which the luminance is either:

(a) constant on the entire block, $Y(j,k) = Y_0$, (2a)

(b) constant in the vertical direction, $Y(j,k) = Y_0(j)$, (2b)

or (c) constant in a horizontal direction, $Y(j,k) = Y_0(k)$. (2c)

In addition, calculate a luminance-adapted contrast using the mean luminance value of the 8 x 8 block under examination and the means of its nearest 4 neighboring blocks which share a boundary according to the expression (3). If $Y_{J,K}$ designates the average luminance on the 8 x 8 image block cornered at pixel (J,K), we consider the local contrast, $C_{Y_{JK}}$, based on four directional differences, $D_N = |Y_{JK} - Y_{(J-8)K}|$; $D_S = |Y_{JK} - Y_{(J+8)K}|$; $D_E = |Y_{JK} - Y_{J(K+8)}|$; and $D_W = |Y_{JK} - Y_{J(K-8)}|$ of the block averages and the average luminance on the surrounding 24 x 24 pixel block, Y_{24x24}.

$$C_{Y_{JK}} = \frac{\min(D_N, D_S, D_E, D_W)}{Y_{24x24}} \qquad (3)$$

Given that the block satisfies the level conditions (2) it is accepted as a flat only if the contrast value exceeds a visibility threshold determined empirically by subjective measurement. As in [2], we use a contrast threshold value of 0.03. Thus, a block, Y, satisfying (2) is a FLAT if only if

$$C_Y > 0.03. \qquad (4)$$

Two detectors for mosquito noise: FLATS-based and Root Mean Square-based

The FLATS detector produces a count, F_n, of flats in frame n. F_n, cannot exceed the maximum number of blocks in the image. For our Recommendation 601 video [3], the image width = M = 720 pixels, and the image height = N = 486 pixels, yielding a peak value
$F_{peak} = M * N / 64 = 5400$.

We also consider an RMS frame impairment metric. For an original video sequence, O_n, and compressed video sequence, C_n, This second metric is based on the frame-by-frame root mean square (RMS) of the difference of the two sequences:

$$R_n = \| O_n - C_n \|_2 .$$

The peak value of the RMS is the same as the peak luminance value. For 8-bit luminance values, Recommendation 601 implies $R_{peak} = Y_{peak} = 235$.

Each of these two metrics computes an estimate of the impairment level in each frame of a video sequence. To convert a frame-based metric into a metric on a video clip which captures the intermittent character of the mosquito noise impairment, we use the time-averaged magnitude of the frame-to-frame change in the impairment. For any frame impairment metric, I, (such as F_n or R_n) the temporal metric, M_I is

$$M_I = mean\{| I_n - I_{n-1} |\}.$$

In addition to its simplicity, this metric is peaked at 30 Hz. It is a simple, if not very precise, approximation to the continuous perceptual filters described in Watson [10] and Winkler [8].

We use peak signal-to-noise ratio (PSNR) measured in dB to provide a common scale for these metrics.
For RMS one has :

$$\mathrm{PSNR}_R = - 20 \log_{10}\{ M_R / R_{peak} \}.$$

For FLATS one has :

$$\mathrm{PSNR}_F = - 20 \log_{10}\{ M_F / F_{peak} \}.$$

6. RESULTS AND CONCLUSIONS

The FLATS-based and RMS-based metrics were applied to the *Spirals* test pattern, using four GOP patterns. These GOP patterns are those identified earlier: (1,1), (3,3), (6,2), and (6,3). Although the target bit rate was set at 1.7 Mb/s for all four encodings, the actual rates were:

GOP indices	Actual bitrate (Mb/s)
(6,3)	1.80
(6,2)	1.85
(3,3)	2.00
(1,1)	3.72

Figure 4 displays the RMS and FLATS data. The most striking feature is that in spite of the high level of blocking in the all I-frame encoding, there is an absence of mosquito noise signal. Except for an initial settling period of three frames, the flat portions of both the (1,1) curves indicates there is little variation in either F or R from frame to frame. In the case of F, the number of blocks is strictly constant. The measures of blocking are highest (and the blocking is readily apparent) for the (1,1) coding. However, informal viewing of the compressed test clip confirms that the noise is static (and the mosquito noise is imperceptible) in the asymptotic region following settling and the noise is dynamic (and the mosquito noise is visible) if the viewing includes the first three frames.

Indeed, there is a settling period for each of the GOP patterns which can be observed in viewing the video. We compare the metrics with and without these transients. Settling may be regarded as a design flaw in the MPEG-2 implementation. For the other 3 GOP patterns, there is a constantly cycling of I- ,B- , and P-frames The magnitude of the mosquito noise is affected by the encoding bit rate and the relative proportion of I-, B-, and P-frames.

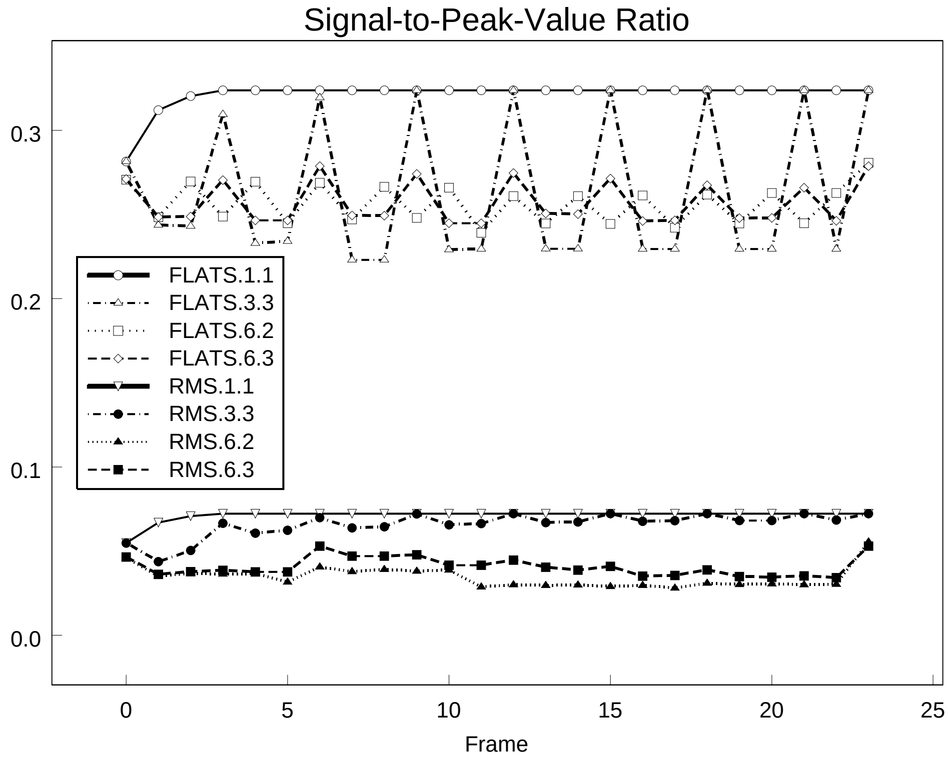

Figure 4: Comparison of the FLATS and RMS image impairment measures. The upper graphs show the frame-by-frame FLATS measure, $\boldsymbol{F_n}$, and the lower graphs that for the RMS metric, $\boldsymbol{R_n}$. The mosquito noise metrics are based on the average variation in $\boldsymbol{F_n}$ and in $\boldsymbol{R_n}$.

Figure 4 suggests (and the data in Table 1 quantifies) the high sensitivity of the FLATS-based metric, $\boldsymbol{M_F}$, compared to the RMS-based metric, $\boldsymbol{M_R}$. The PSNR for $\boldsymbol{M_R}$ was no less than 48 dB while $\boldsymbol{M_F}$ had a PSNR ranging from 23 to 35 dB, except in the absence of Mosquito Noise, GOP=(1,1).

Table 1: Peak signal-to-noise ratio (in dB) for two Mosquito Noise metrics applied to a ‘Spirals’ video clip compressed using four different GOPs. Data is presented for full clips and for ‘settled’ regime.

	RMS (dB) Full series	RMS (dB) Asymptotic	FLATS(dB Full series	FLATS (dB) Asymptotic
GOP=3,3	48.373317	50.114289	24.263875	23.927344
GOP=6,3	49.368541	53.111225	34.084941	34.167673
GOP=6,2	66.504697	66.942597	35.286341	35.502284
GOP=1,1	62.383305	Undefined	54.723735	Undefined

The sensitivity of FLATS can be attributed to its selectivity for DCT blocks. As noted in [4] the FLATS detector is very effective in finding 8 x 8 pixel blocks but may fail to identify other blocking, such as that seen in B- and P-frames. Although this selectivity might be regarded as a weakness in a pure blocking metric, in detecting mosquito noise it emphasizes the measured difference between inter- and intra-encoded frames and appears to improve performance. This suggests that the motion estimation of blocks and the addition of high frequency in coding the residuals of B- and P- frames is a significant component of the mosquito noise.

The most surprising result of this study is the finding that Mosquito Noise occurs in static scenes. This helps in understanding the source of MPEG impairments. MN is strongly associated with the GOP structure. The two frame impairment detectors both exhibit the pattern of the GOP in the trace of frame-by-frame error. The second surprise is the sensitivity of the FLATS-based metric. Our results suggest that a subjective study of mosquito noise would be useful in determining a threshold value for the perception of mosquito noise and in assessing our two metrics. The threshold for MN is likely to be significantly higher than that for static blocking, due to the dynamic character of MN. In particular, the blocking threshold of about 30 dB found in [2] will be higher for Mosquito Noise.

REFERENCES

1. C. Fenimore, B.F. Field, and C. VandeGrift, "Test Patterns and Quality Metrics for Digital Video Compression", *HVEI II*, SPIE vol 3016, San Jose, CA [1997]

2. J. Libert and C. Fenimore, "Visibility Thresholds for compression-induced image blocking", *HVEI IV*, SPIE vol 3644, San Jose, CA [1999]

3. L. Meesters and J-B. Martens, "Blockiness in JPEG-coded images",*HVEI IV*, SPIE vol 3644, San Jose, CA [1999]

4. Huib de Ritter, "Percentage scaling: a new method for evaluating multiply impaired images," *HVEI V*, SPIE vol 3959, San Jose, CA [2000].

5. J. Lubin, *A Visual Discrimination Model for Imaging System Design and Evaluation*, David Sarnoff Research and Development Report, Princeton NJ, [1995].

6. X. Tong, D.J. Heeger, C.J.v.d.B. Lambrecht, "Video quality evaluationusing ST-CIELAB," *HVEI IV*, SPIE vol 3644, San Jose, CA [1999]

7. A.B. Watson, J. Hu, J.K. McGowan, J.B. Mulligan, "Design and performance of a digital video quality metric"*HVEI IV*, SPIE vol 3644, San Jose, CA [1999].

8. S. Winkler, "A perceptual distortion metric for digital color video," *HVEI IV*, SPIE vol 3644, San Jose, CA [1999]

9. ITU-T Recommendation P.930, *Principles of a reference impairment system for video*, 8/96.

10. A.B. Watson, "Toward a perceptual video quality metric"*HVEI III*, SPIE vol 3299, San Jose, CA [1998].

11. ISO/IEC 13818-2, *Generic Coding of Moving Pictures and Associated Audio - Part 1: Video*, International Organization for Standardization, [1995].

12. ISO/IEC JTC1/SC29/WG11/N0400, *Test Model 5 (draft)*, MPEG93/457,[1993] Software is available by FTP on the World Wide Web at ftp://ftp.crs4.it/mpeg/programs/.

13. ITU-R Recommendation 601-5, *Studio Encoding Parameters of Digital Television.*

14. ITU-R Recommendation 802-1, *Test Pictures and Seauences for Subjective assessments of Digital Codecs*.

15. W.F. Schreiber, *Fundamentals of Electronic Imaging Systems*, 3rd Ed., Springer, Berlin [1993].