aJ

140

COMMITTEE T1 - TELECOMMUNICATIONS
STANDARDS CONTRIBUTION

DOCUMENT NUMBER: T1A1.5/94-xxx

DATE: July 1994

STANDARDS PROJECT: T1A1.5 - Performance

SUBJECT: Modification of video reference impairment system (VIRIS) to improve mosquito noise simulation

SOURCE: Bellcore

CONTACT(S): Anthony M. Lessman
Bellcore, Room NVC 1E-314
331 Newman's Springs Road, Red Bank, NJ 07701
Tel. (908) 758-2513, Fax: (908) 758-4545
E-mail: tony1@cc.bellcore.com

ABSTRACT: This is a proposed contribution to ITU-T Study Group 12 which addresses that portion of Question 22 dealing with a video reference impairment system. We are seeking approval from T1A1.5 to submit it as a Bellcore contribution. It reports on further work completed on the preliminary video reference impairment system (VIRIS) reported last year in Contribution T1A1.5/93-103 and in Contribution ITU-T, Study Group 12, COM 12-21-E. The contribution reports on changes made in the program to improve the mosquito noise simulation based on observations of impairment artifacts in video images produced by an experimental MPEG 1 coding algorithm.

DISTRIBUTION: T1A1.5

INTERNATIONAL TELECOMMUNICATION UNION

DOCUMENT NUMBER
Date
Original: English

TELECOMMUNICATION
STANDARDIZATION SECTOR

STUDY PERIOD 1993 - 1996

Geneva, 28 November - 9 December 1994

Text available only in English

Question:22/12

STUDY GROUP 12 - CONTRIBUTION

Source*: BELLCORE

Title: MODIFICATION OF VIDEO REFERENCE IMPAIRMENT SYSTEM (VIRIS) TO IMPROVE MOSQUITO NOISE SIMULATION

1. INTRODUCTION

A video reference impairment system, VIRIS[1][2], was created in 1992 to simulate the digital video coder impairment artifacts of blocking, blurring, mosquitoes and noise. This contribution reports on changes made in the program to improve the mosquito noise simulation based on observations of impairment artifacts in video images produced by two versions of an experimental MPEG1 coding algorithm developed at Bellcore. These changes were made because of the desire to use VIRIS as an aid in evaluating the picture quality of MPEG1 codecs. The modified program, VIRIS1, is identical to the original VIRIS program except for the mosquito noise simulation changes.

A long term goal for the VIRIS work is to contribute towards the creation of a new draft recommendation in Working Party 12/2, Rec. P.VRU, a video reference impairment unit. This work supports Question 22/12[3] which addresses audio/video quality in multimedia services.

The next section discusses and defines the method used to simulate the mosquito impairment in VIRIS1. Section 3 describes the test plan for subjectively characterizing the simulated mosquito impairment along with two software implementations of the experimental MPEG1 algorithm. Subjective test results are discussed in Section 3 and conclusions are given in Section 4.

2. MPEG1 IMPAIRMENT SIMULATION

The general approach used in simulating MPEG1 impairments in VIRIS1 was to first observe actual impairment artifacts introduced by an MPEG1 algorithm and then try to produce simulated impairments resembling, as close as possible, the actual ones. The next subsection discusses this approach in simulating MPEG1 impairments. Following subsections detail the simulation method and the changes in the VIRIS program.

* CONTACT PERSON Anthony M. Lessman Tel.:1 908-758-2513

2.1 MPEG1 Impairment Artifacts

A software implementation of MPEG1 was used to produce MPEG1 signals (a hardware codec was not available). The MPEG1 algorithm, called Bellcore '91, was developed at Bellcore in support of MPEG standards activities. In the present study, various television picture sequences were subjected to the Bellcore '91 algorithm and then viewed on a television monitor to investigate the type of impairment artifacts produced. In order to better discern the type of impairment artifacts produced by MPEG1 encoding, the algorithm was run at 1.0 Mbit/s in addition to the normal rate of 1.3 Mbit/s to increase the impairment magnitude. Tandem encodings (processing the signal through the algorithm more than once) were also performed in an attempt to increase the impairment magnitude but this resulted in much less effect compared to lowering the bit rate.

Viewing of the processed MPEG1 pictures indicated that the predominant impairment artifacts were mosquito noise and blurring. These impairments always occurred in combination and their perceptibility was directly correlated with the amount of motion and detail in the picture - the more motion and detail, the higher the perceptibility. Blocking artifacts could also be perceived but only when viewing still frames of the picture sequence. Even then, the blocking artifacts could be seen only on certain pictures with high motion content and the magnitude of the blocking impairment was low. The blocking artifacts could not be perceived on any of the pictures when viewed at the normal rate of 30 frames per second as they became indistinguishable from the mosquito noise and blurring artifacts. Random noise artifacts could not be perceived on any of the test pictures.

Mosquitoes were simulated in the original VIRIS program[2] by adding close-in leading and lagging positive echoes (design displacement from main signal of 0.5 micro-seconds) to the SIF (Source Input Format, 352 pels x 240 lines) image in the horizontal direction to create a halo effect around objects. A shimmering of the halo was created by turning the echoes off every fifth frame. However, the mosquito artifacts (tiny spark-like points or heat waves) in MPEG1 pictures occur not only on the sides of objects but also on the top and bottom of objects.

The scheme used in VIRIS to simulate mosquito noise was modified in VIRIS1 to better simulate the actual mosquito noise impairment produced by the MPEG1 algorithm. In VIRIS1, mosquitoes were simulated by adding three different close-in echoes, each on for five frames in turn, to create a shimmering effect. The leading and lagging negative echoes were applied in both the horizontal and vertical directions and had design displacements of 0.375, 0.5 and 0.75 micro-seconds. It was determined that negative echoes (white echoes to black and black to white) resulted in slightly better simulation than positive echoes. It was also determined that it was not necessary to introduce additional blurring to the picture to simulate MPEG1 blurring because the echoes used to simulate the mosquito impairment also tended to blur the edges in the picture.

An imperical approach was used to arrive at the above simulation method. It involved a trial and error process in dealing with different ways of adding echoes and then comparing results with MPEG1 pictures. Changes were made in the simulation method in attempts to improve the simulation, but these changes could not be done in real time since it was necessary to run the VIRIS program to produce the changes. Nevertheless, it is believed that a reasonable simulation of MPEG1 impairments resulted from the above simulation method. Of course, the perceptibility of actual MPEG1 impairments is correlated with motion in the picture, whereas in VIRIS1 the simulated impairments are not. Therefore, the degree to which the simulated impairments resemble the real impairments depends to a great extent on the particular picture sequence being used. In any case, one of the primary concerns in the initial work with VIRIS was that it be simple to implement and be independent of any coding scheme. This precludes, at least for now, correlation of motion in the picture with the simulated impairment. The next subsection details the VIRIS1 mosquito simulation method.

2.2 VIRIS1 Mosquito Simulation Method

Mosquitoes were simulated in VIRIS1 by adding close-in, leading and lagging negative echoes to the SIF image in both the horizontal and vertical directions to create a halo around objects. The echoes had design displacements of 0.375, 0.5 and 0.75 micro-seconds. A shimmering of the halo was created by changing the echo displacement every five frames.

Echoes in the SIF image are the result of ripples in the passband amplitude response of the image array. The ripple frequency determines the echo displacement; the ripple amplitude determines the echo amplitude and the ripple phase determines the echo polarity (positive or negative). The ripples in the frequency response were produced by applying 13-tap digital, multiband filters[4] to the SIF image array in both the horizontal and vertical directions. Table 1 shows the coefficient values for the three sets of multiband filters which produce the three echo displacements. The particular echo displacement used on any frame is determined by user input to the VIRIS1 program.

The coefficient value of tap 0 of each of the filters was scaled to a value of 175. With the exception of one set of positive/negative taps for each of the three echo displacement filters, all other filter tap values were 0. The particular set of filter taps having a non-zero coefficient value determined the echo displacement while the value of the coefficient determined the echo amplitude. The actual coefficient value was input by the user to the VIRIS1 program and could range from -1 to -30, in whole number steps. For each specific impairment level, the same coefficient value was applied to each of the three filters.

The filtering was implemented in VIRIS1 by applying the digital multiband filters to the SIF luminance image array, first along each of the 240 rows and then down each of the 352 columns. The filtered output signal, z[n], for each case is given in Equation 1.

$$z[n] = \frac{1}{S} \times \sum_{i=-6}^{i=6} h[i] \times x[n-i], \qquad \text{[EQ1]}$$

where $x[n]$ = input SIF image array,

$h[i]$ = filter coefficients (See Table 1),

$$S = \sum_{i=-6}^{i=6} h[i],$$

$0 \le n < 352$ in horizontal direction,

$0 \le n < 240$ in vertical direction.

In Equation 1, the input signal, x[n], is extended by 6 samples at the beginning and ending of each line or column.

Table 1: Coefficient Values For 13-Tap, Multiband Filters

Filter Tap	Filter Tap Coefficients		
	Echo Displacement, μ-secs		
	0.375	0.5	0.75
[-6],[6]	0	0	-1 to -30
[-5],[5]	0	0	0
[-4],[4]	0	-1 to -30	0
[-3],[3]	-1 to -30	0	0
[-2],[2]	0	0	0
[-1],[1]	0	0	0
[0]	175	175	175

3. SUBJECTIVE TEST PLAN

Subjective tests were performed to determine the relationship between impairment level and subjective opinion for the mosquito impairment simulated by VIRIS1. Other conditions were also tested and these included a SIF reference condition, an NTSC reference condition and two MPEG1 conditions, one at 1.3 Mbit/s and the other at 1.0 Mbit/s.

In general, the test procedures and recommendations given in ITU-R (formerly CCIR) Recommendation 500-4[5] were used as guidelines in designing the test plan. This recommendation gives guidelines on test methods, viewing conditions and rating scales for use in evaluating TV picture quality. The basic approach used in the test plan was to have nonexpert observers rate the quality of 10-second TV picture sequences representing various levels of the VIRIS1 mosquito impairment, the SIF and NTSC reference conditions and the MPEG1 conditions. They used the discrete 5-point quality rating scale that ranges from EXCELLENT to UNSATISFACTORY. Details of the test plan are given in the following subsections.

3.1 Picture Sequences

Three picture sequences were selected for the tests. The particular number selected was a compromise between the desire to have a large sample to represent different types of pictures (with the added benefit of providing variety for the test subjects) and the processing time required to prepare the test material. Each of the sequences was 5 seconds in length and, after processing, was repeated during the editing process to create a 10-second sequence.

The sequences were selected to obtain varying degrees of motion and detail. Their titles with brief descriptions are:

1. "Race Cars"-This sequence was derived from a 20-second sequence that shows moving race cars on a stationary racetrack background. The sequence includes four scene cuts to different cars and includes some detail (mostly due to numbers on race cars) and moderate motion. It originated from a Bellcore demonstration clip of HDTV which was downloaded to NTSC format.
2. "Table Tennis"-This sequence was derived from a 20-second sequence that shows two men playing a table tennis game and includes moderate detail and motion. There are 3 scene cuts with the

background in two of the scenes stationary and the background in the other moving slowly. The sequence originated from D1 digital tape and was one of the "standard" sequences used by the MPEG community for evaluating MPEG algorithm performance.

3. "Football"-This sequence was used in the previous VIRIS study[2]. It is a sequence from a football game that has considerable detail and moderate motion. It includes no scene cuts and the background moves slowly. The sequence originated from a Bellcore demonstration clip of HDTV which was downloaded to NTSC format.

3.2 Test Conditions and Test Material Preparation

Eight mosquito impairment levels were selected as test conditions based on informal laboratory viewing tests. Table 2 shows the PSNRs calculated by VIRIS1 for each of the mosquito impairment levels, first as an average across each picture sequence and then as an average across the three pictures

Table 2: PSNRs For VIRIS1 Mosquito Impairment Levels

Mosquito Level, M	Average PSNR, dB			
	Race Cars	Football	Table Tennis	3-Pic Average
-1	42.9	43.5	41.1	42.5
-3	41.8	40.7	39.5	40.7
-5	40.3	37.8	37.3	38.5
-10	36.5	32.1	32.6	33.7
-15	33.2	28.2	29.1	30.2
-20	30.3	25.1	26.4	27.3
-25	27.7	22.7	24.1	24.8
-30	25.3	20.5	22.1	22.6

Note: Mosquito Level, M, corresponds to filter tap coefficient values for the variable range of coefficients given for the three multiband filters of Table 1.

The PSNR's of the Football and Table Tennis sequences were fairly close, differing by amounts from 0.5 to 2.4 dB across the mosquito level range. The PSNR's of the Race Cars sequence differed from those of the other two sequences by amounts from 0.6 to 5.2 dB across the mosquito level range. Figure 1 shows a plot of the 3-picture average PSNRs versus the mosquito input level to VIRIS1. A linear function fitted to the data is also shown on the plot to provide a pathway from PSNR to VIRIS1 mosquito level input.

In addition to the 8 VIRIS1 mosquito impairment conditions, the test included conditions of NTSC quality, SIF quality and two conditions of MPEG1, one at 1.3 Mbit/s and the other at 1.0 Mbit/s, for a total of 12 test conditions. Each of the three test sequence pictures was subjected to each of the 12 test conditions for a total of 36 test sequences. During the processing and editing, the sequences were stored on D1 digital tape and then later transferred to a Sony Laser Video Disk Recorder (Model LVR-5000/LVS-5000). The laser disk recorder allowed for high quality NTSC recording and random access playback of the recorded material. It was controlled by a computer program to play back the test material in different random orders as required by the test design.

3.3 . Test Method

A single stimulus test method using a discrete 5-point quality scale was used in the tests. A total of 27 nonexpert observers, selected from an out-of-house subject pool, participated in the tests. They included 3 males and 24 females with ages ranging from 19 to 71 with an average age of 47.3. Observers rated each of the test conditions by placing a check mark next to the appropriate judgement category of the comment scale with terms of EXCELLENT, GOOD, FAIR, POOR and UNSATISFACTORY. There was a voting sheet for each test condition.

3.4 Test Procedures

Viewing conditions were the same as reported previously[2]. Three observers were tested at a time, except for one test where only two were tested (one called subject failed to appear). After the test administrator read the test instructions and answered any questions, a practice session of 12 conditions was conducted to familiarize the subjects with the test procedures and with the test material. Each test condition took 20 seconds and included a 2-second message screen on the receiver identifying the condition number, the 10-second picture sequence and an 8-second voting period with a message screen requesting the observer to vote for the picture quality. Each of the 36 test conditions was tested twice for a total of 72 conditions.

The test was administered in two parts with a 10-minute break between the parts. Each part included the 36 different test sequences and took 12 minutes to complete. The conditions in each part were presented in a pseudo random order with two constraints:

1. the first, second and third group of 12 picture sequences each contained all of the 12 basic test conditions and
2. each of the three groups of 12 picture sequences contained an equal number of the three picture sequences.

These constraints were designed to counterbalance any learning effects on the results such as a sudden awareness of a particular aspect of an impairment only after the test was well underway, or perhaps, a gradual change in a subject's criteria for judging quality as the test progressed. The random orders were different in each half of the test and were changed for each group of subjects.

4. SUBJECTIVE TEST RESULTS

The mean opinion score (MOS) was computed for each of the 12 test conditions for each of the 3 picture sequences. MOS's for the NTSC reference conditions were 4.7, 4.6 and 4.6 for the Race Cars, Football and Table Tennis sequences, respectively. MOS's for the SIF reference conditions for the corresponding sequences were 3.2, 3.6 and 3.8, respectively, indicating a substantial drop in quality from NTSC to SIF. This loss of quality is primarily due to the drop in horizontal and vertical resolution that occurs in the conversion from CCIR 601 to SIF image format. Since MPEG1 algorithms use SIF as the input image format, the quality associated with the SIF also represents the limit on quality that can be obtained with MPEG1 algorithms.

The SIF reference condition with the Football sequence had an MOS of 3.6 which is 0.6 lower than was obtained in the previous VIRIS study[2]. However, the previous study did not use NTSC reference conditions as was done in the present study and as a result, the stimulus material did not cover the full range of the comment scale. Since subjects probably attempt to use all of the comments on the scale to judge the stimulus material, they may have rated the SIF reference conditions in the previous study higher than they would have if a full range of stimulus material had been presented. The purpose of including the NTSC conditions in the present study was to obtain full coverage of the comment scale to reduce or eliminate this type of anchoring problem. The SIF quality obtained in the present study is, thus, probably more indicative of actual SIF quality than the results obtained previously.

Figure 2 shows plots of MOS as a function of the VIRIS1 mosquito impairment PSNR for the 3 picture sequences. A logistic function fitted to the 3-picture data is also shown on the figure. The logistic function is of the same form as used and described in the previous study. Referring to Figure 2, subjective reaction to the mosquito impairment was somewhat different for the three test picture sequences. The results for the Football and Table Tennis sequences were similar, but distinctly different from the results for the Race Cars sequence, especially at the higher PSNRs. However, despite the differences in results for the 3 picture sequences, the logistic function provides a good fit to the 3-picture data with an R squared = 0.90. R squared is a goodness of fit statistic that ranges from 0 to 1 with 1 representing an optimal fit. It also indicates the proportion of variance in the data that is accounted for by the model.

The MOSs for the MPEG1 algorithm at 1.3 Mbit/s were 2.7, 3.0 and 3.6 for the Race Cars, Football and Table Tennis sequences, respectively. Corresponding MOSs for the MPEG1 algorithm at 1.0 Mbit/s were 2.4, 2.8 and 3.5. The average MOSs across the three sequences for the 1.3 and 1.0 Mbit/s MPEG1 algorithms were 3.1 and 2.9, respectively. The picture quality of the MPEG1 algorithms can be expressed in terms of the PSNR for the mosquito impairment of VIRIS1. Figure 3 shows the VIRIS1 logistic model that relates mosquito PSNR to MOS along with the PSNR intersections on the model for the 3-picture average MOSs for the two MPEG1 conditions. The picture quality of the 1.3 and 1.0 Mbit/s MPEG1 algorithms are equivalent to the VIRIS1 mosquito PSNR levels of 31.9 and 30.6 dB, respectively.

5. CONCLUSIONS

It was determined in this study that mosquito noise was the predominant type of impairment artifact produced by MPEG1 video coding algorithms. The mosquito impairment simulation in the original VIRIS program was modified to make it more closely resemble the actual impairment artifacts produced by MPEG1 encoding algorithms. Figure 2 shows the relationship between the PSNR of the new VIRIS1 mosquito impairment and MOS for each of the three test picture sequences and for the 3-picture average. Figure 3 demonstrates that VIRIS1 can be used as an aid in evaluating MPEG1 picture quality. Results of subjective tests indicated that the picture quality of the Bellcore '91 MPEG1 algorithm at 1.0 and 1.3 Mbit/s was equivalent to that produced by the VIRIS1 mosquito impairment having PSNRs of 30.6 and 31.9 dB, respectively.

VIRIS1 results in a reasonable simulation of the mosquito noise that occurs with MPEG1 encoding algorithms and can be used now as an aid in evaluating MPEG1 picture performance. It provides a constant quality reference that can be used to compare the quality performance of different MPEG1 codecs. However, it is believed that there is room for improvement in the simulation. Changing the echo displacement to one of three different values every five frames may have resulted in an echo "shimmer" that was too regular and that resembled, on certain test scenes, not mosquito noise but "ghost" flutter due to an airplane. One possible adjustment in the simulation would be to change the echo displacement in a random manner which would still result in a shimmering of the echo, but not in a regular pattern. This possible simulation improvement may be investigated in future work

As discussed earlier, the perceptibility of the impairment artifacts produced by MPEG1 coders is correlated with the amount of detail and motion in the picture, the faster the motion and the greater the detail, the higher the perceptibility. However, the perceptibility of the simulated impairments in VIRIS1 is not correlated with motion or detail content. Test sequences with moderate motion and detail were, therefore, used to increase the similarity between real and simulated impairments. These types of sequences also exercise the MPEG1 system under test to a greater extent than sequences with slower motion and little detail. Of course, one option for future work could be to include some degree of detail and motion correlation in the impairment simulation algorithms to make the simulation less independent of the type of picture sequences used in the subjective tests. However, the simulation algorithms would necessarily become much more complex than those in VIRIS1.

Future work in this project may involve working with image formats other than SIF. For example, the forthcoming MPEG2 coding standard for bit rates of 3 Mbit/s and higher uses the CCIR 601 image format (720 pels x 480 lines). Coders commonly used for video teleconferencing applications follow the ITU-T H.261 coding standard[6] which uses the CIF (Common Intermediate Format, 352 pels x 288 lines) format and, sometimes, the QCIF (Quarter CIF, 176 pels x 144 lines) format. The simulation of impairment artifacts for any of these other types of codecs would require use of the appropriate and corresponding input image format.

REFERENCES

1. Contribution T1A1.5/93-103, "Report on an experimental video reference system," July 22, 1993.
2. Contribution ITU-T, COM 12-20-E, "VIRIS, AN EXPERIMENTAL VIDEO REFERENCE SYSTEM," Question 22/12, December, 1993.
3. Contribution T1A1.5/93-141, "A report of the May 1993 meetings of ITU-TS SG 12 (formerly CCITT SG XII) and Speech Quality Experts Group in Geneva," Bellcore & AT&T Communications, July 1, 1993.
4. "Digital Filter Design Package," Version 2.12, Atlanta Signal Processors Inc., December, 1986.
5. CCIR Recommendation 500-4, "Method for the Subjective Assessment of the Quality of Television Pictures," 1990.
6. CCITT Recommendation H.261, COM XV-R 37-E, "Video Codec For Audiovisual Services at px64 kbit/s," August 1990.

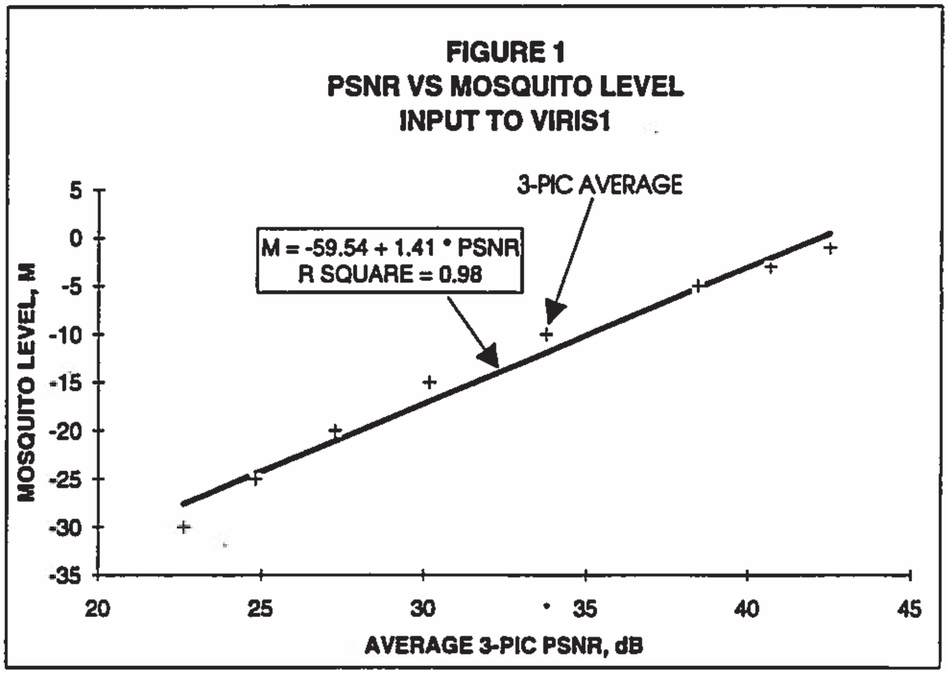
FIGURE 1
PSNR VS MOSQUITO LEVEL
INPUT TO VIRIS1
3-PIC AVERAGE
M = -59.54 + 1.41 * PSNR
R SQUARE = 0.98
MOSQUITO LEVEL, M
5
0
-5
-10
-15
-20
-25
-30
-35
20
25
30
35
40
45
AVERAGE 3-PIC PSNR, dB

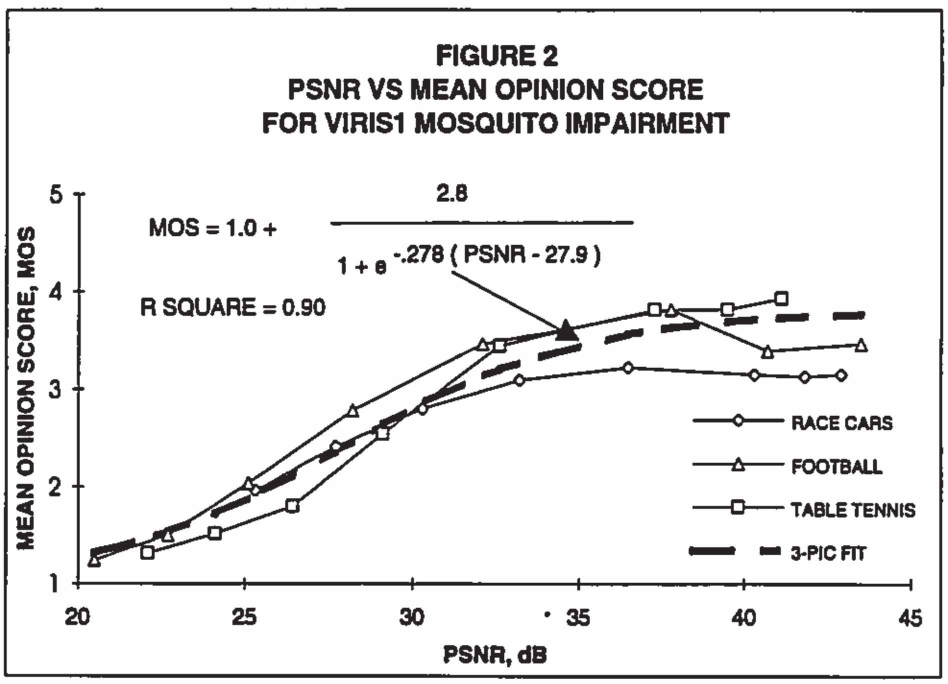
FIGURE 2
PSNR VS MEAN OPINION SCORE
FOR VIRIS1 MOSQUITO IMPAIRMENT
MOS = 1.0 + 2.8 / (1 + e -.278 (PSNR - 27.9))
R SQUARE = 0.90
MEAN OPINION SCORE, MOS
5
4
3
2
1
20
25
30
35
40
45
PSNR, dB
RACE CARS
FOOTBALL
TABLE TENNIS
3-PIC FIT

FIGURE 3
USE OF VIRIS1 TO EVALUATE
PERFORMANCE OF MPEG1 ALGORITHMS