INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.930

(08/96)

SERIES P: TELEPHONE TRANSMISSION QUALITY
Audiovisual quality in multimedia services

Principles of a reference impairment system for video

ITU-T Recommendation P.930

(Previously CCITT Recommendation)

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	Series	P.30 P.300
Transmission standards	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of quality	Series	P.80 P.800
Audiovisual quality in multimedia services	**Series**	**P.900**

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION P.930

PRINCIPLES OF A REFERENCE IMPAIRMENT SYSTEM FOR VIDEO

Summary

This Recommendation describes the principles of an adjustable video reference system that can be used to generate the reference conditions necessary to characterize the subjective picture quality of video produced by compressed digital video systems. A Reference Impairment System for Video (RISV) can be utilized to simulate the impairments resulting from the compression of video sequences, independent of compression scheme. The subjective evaluation methods are described in Recommendation P.910.

Appendix I describes VIRIS (a Video Reference Impairment System developed by Bellcore), which is a specific implementation of an adjustable reference impairment system for video. Although the studies done at Bellcore were with MPEG-1, VIRIS can also be used with other compression schemes, such as H.261.

It is recommended that a RISV be capable of producing the following categories of distortions, either singly or in combinations, with independent adjustment of each impairment level:

a) Artifacts due to conversions between analog and digital formats (e. g. noise and blurring).

b) Artifacts due to coding and compression (e. g. jerkiness, edge busyness, and block distortion).

c) Artifacts due to transmission channel errors (e. g. errored blocks).

In this Recommendation, five types of impairments (block distortion, blurring, edge busyness, noise, and jerkiness) are defined and general methods for implementing these impairments are provided. Appendix I describes a specific implementation of these impairments in VIRIS. Other impairments are the subject for future study.

From the viewer's point of view, the impairments produced by the RISV should be a good approximation of impairments generated by digital video coding and transmission systems.

Three possible applications for the RISV are:

1) creating reference conditions in subjective tests of digital video systems to ensure that the quality of the scenes presented to viewers covers the entire range of picture quality;

2) defining standard video impairment levels that can be used to compare subjective test results; and

3) quantifying the user-perceived quality of a video system with respect to a known reference.

Although this Recommendation describes the principles of an RISV, before an implementation can be recommended, validation tests are required.

Source

ITU-T Recommendation P.930 was prepared by ITU-T Study Group 12 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 30th of August 1996.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

CONTENTS

Introduction

Reference video conditions are often used in subjective video quality evaluations because category rating judgements can vary from one test to the next or from one laboratory to another depending on the experimental setting (i.e. range and frequency effects, subject population, etc.). The reference conditions allow for relative comparisons between the reference video conditions and the video conditions of interest and allow for subjective quality to be stated in terms of an objective measure of the reference. Including reference video conditions in a test allows for a wide range of test conditions to be presented and provides for a measure of video quality that can be more reliably replicated in another test or laboratory.

In considering a system capable of producing such a set of adjustable video reference conditions, several issues need to be addressed:

1) the type of video impairments or digital compression artifacts that should be simulated must be determined;
2) the appearance of the simulated video impairments should be similar to the actual impairments;
3) the means of generating the impairments should be precisely defined so that they can be reproduced in any laboratory; and
4) the impairment level range should provide a wide range of picture quality performance.

Video compression algorithms may introduce numerous impairment artifacts into the picture. Five of the more common types are block distortion, blurring, edge busyness, jerkiness, and noise. These impairment artifacts are created in the compression stage and are generally due to loss of significant information by quantization, coding and other compression techniques. They are usually correlated with motion in the picture.

An adjustable video reference system should be rather simple to implement and should simulate the above five impairments similar to those produced by actual digital video encoders and independent of the compression scheme. The simulated impairment range should cover a wide range of picture quality for evaluating the picture quality of digital video coders. This requires that subjective tests be performed to determine the picture quality range provided by the simulated impairments.

This Recommendation describes some of the impairments that should be produced by an adjustable video reference impairment system and the general methods used to produce them. Appendix I contains a description of VIRIS, which is a specific implementation of reference impairment system for video for the evaluation of MPEG-1 digital video coders. Note that before VIRIS or any implementation of a RISV can be accepted by all Administrations, validation tests are required.

This Recommendation reflects the current early status of research on reference impairment system for video. As progress on this work continues, understanding of these methods no doubt improve. As new knowledge is attained, this Recommendation will be revised.

Recommendation P.930

PRINCIPLES OF A REFERENCE IMPAIRMENT SYSTEM FOR VIDEO

(Geneva, 1996)

1 Scope

This Recommendation describes an adjustable video reference system that can be used to generate the reference conditions necessary to characterize the subjective picture quality of video produced by compressed digital video systems. A Reference Impairment System for Video (RISV) can be utilized to simulate the impairments resulting from the compression of video sequences, independent of compression scheme. Subjective evaluation methods are not covered in this Recommendation but are instead described in Recommendation P.910.

2 References

The following Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

- ITU-R Recommendation BT.500-6 (1994), *Methods for the subjective assessment of the quality of television pictures*.
- ITU-R Recommendation BT.601-4 (1994), *Encoding parameters of digital television for studios*.
- ITU-R Recommendation BT.802 (1992), *Test pictures and sequences for subjective assessments of digital codecs conveying signals produced according to Recommendation 601*.
- ITU-T Recommendation P.920 (1996), *Interactive test methods for audiovisual communications*.
- ITU-T Recommendation P.910 (1996), *Subjective video quality assessment methods for multimedia applications*.
- ITU-T Recommendation H.261 (1993), *Video codec for audiovisual services at p × 64 kbit/s*.

3 Abbreviations and definitions

For the purposes of this Recommendation, the following abbreviations and definitions are used.

3.1 **pel (or pixel)**: A picture element that describes the brightness or colour of a discrete point in an image.

3.2 **block**: Group of pels. For example, a block of 8 × 8 pels is the smallest coding block used in MPEG-1 algorithms. There are 1320 such blocks in a SIF image, 44 in the horizontal direction (352 pels/8) and 30 in the vertical direction (240 lines/8).

3.3 **block distortion (or tiling)**: Distortion of the image characterized by the appearance of an underlying block encoding structure.

3.4 **blurring**: Global distortion of the entire image, characterized by reduced sharpness of edges and spatial detail.

3.5 **colour errors**: Distortion of all or a portion of the final image characterized by the appearance of unnatural or unexpected hues or saturation levels. These hues or saturation levels were not present in the original image.

3.6 **edge busyness**: Distortion concentrated at the edges of objects, and further characterized by its temporal and spatial characteristics.

3.7 **error blocks**: A form of block distortion where one or more blocks in the image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks.

3.8 **jerkiness (or jerky motion)**: Motion that was originally smooth and continuous perceived as a series of distinct "snapshots".

3.9 **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).

3.10 **quantization noise**: A "snow" or "salt and pepper" effect similar to a random noise process but not uniform over the image.

3.11 **CIF**: Common Intermediate Format used by H.261 coders, 352 luminance pels × 288 lines.

3.12 **QCIF**: Quarter CIF, 176 luminance pels × 144 lines.

3.13 **SIF**: Source Input Format used by MPEG coders, a progressive, non-interlaced format of 352 luminance pels × 240 lines × 29.97 Hz or 352 luminance pels × 288 lines × 25 Hz.

3.14 **CCIR format**: ITU-R (formerly CCIR) digital video standard using interlaced formats of 720 luminance pels × 480 lines × 30 Hz and 720 luminance pels × 576 lines × 25 Hz.

3.15 **MPEG**: Moving Pictures Experts Group, a Working Group organized by ISO and charged with the responsibility to develop multimedia/systems standards.

4 Attributes of a Reference Impairment System for Video (RISV)

It is recommended that a RISV be capable of producing artifacts (either singly or in combinations) due to:

a) conversions between analogue and digital formats (e.g. noise and blurring);

b) coding and compression (e.g. jerkiness, edge busyness, and block distortion);

c) transmission channel errors (e.g. errored blocks).

From the viewer’s point of view, the impairments produced by the RISV should be a good approximation to the impairments generated by digital video coding and transmission systems. Next, five of these impairments are described in more detail, and Appendix I describes algorithms for producing these types of impairments.

4.1 An adjustable range of video impairment levels

The amount of distortion for each impairment should be independently adjustable. The ranges of adjustment should be sufficient to cover the full range of quality for digital video coding and transmission systems. In addition, there should also be a means to combine two or more impairments and adjust the combined level of these impairments over a range that is useful for testing these systems. Both these topics are subject of continuing study, and subclause 5.6 provides additional guidance on these topics.

4.2 Potential applications for a RISV

Three potential applications for the RISV are:

1) creating reference conditions in subjective tests of digital video systems to ensure that the quality of the scenes presented to viewers covers the entire range of picture quality;
2) defining standard video impairment levels that can be used to compare subjective test results; and
3) quantifying the user-perceived quality of a video system with respect to a known reference.

Standardization of the RISV is sufficient for the first application. In a subjective test, a reference condition is the video scene that is processed through the RISV to add impairments of a specified level. Usually several reference conditions are added to the list of subjective test scenes and the impairment levels in the reference conditions are chosen to cover complete range of quality – from EXCELLENT to BAD. This allows for the subjective test participants to make a more reliable judgement on quality.

For the second and third applications, standardization of the RISV alone is not sufficient. For these two applications, the video scenes used with the RISV must also be standardized. This is because the perceived level of digital video impairments is highly dependent on the video scene content (i.e. the spatial and temporal information in the scene).

With the second application, use of a RISV becomes necessary if several different laboratories are performing subjective tests independently on the same digital video system. The laboratories may want to compare the test results, determine the inter-laboratory differences, and combine the results for a global analysis. If different laboratories want to compare their independent test results for a global analysis and calculate the inter-laboratory differences, the laboratories must utilize the same test scenes and reference conditions. This is possible by using a RISV and producing the associated impairment levels for a specific set of test scenes. In general, reporting only the impairments settings of the RISV is not sufficient. The video scene used to determine the RISV impairments levels settings must also be reported (and this scene should be one of a widely available standard test scenes such as those found in ITU-R Recommendation BT.802).

Figure 1 illustrates the third potential application of a RISV. A RISV can be used to quantify user quality of a video sequence passed through a compression system. By comparing the quality of video as produced by a compression system with that produced by a RISV, an objective measure of quality in terms of the RISV impairment levels is possible. Using Figure 1 the operation of the RISV is as follows: A video scene is injected into the video system under test and into the RISV. Next, the video out from the video system under test and the RISV are alternately displayed on a monitor. The RISV impairment levels are adjusted by a set of adjustment "knobs" (effectively software parameters) until a viewing test subject judges that the video qualities are equal.

The use of the RISV in this manner allows for specifying the quality of video objectively in terms of the RISV impairment levels. As implementations of RISVs are developed, the algorithm specified to assess video quality, as described by the third application may be refined. Therefore, the operation of an RISV in this manner is quite preliminary and is a subject for future study.

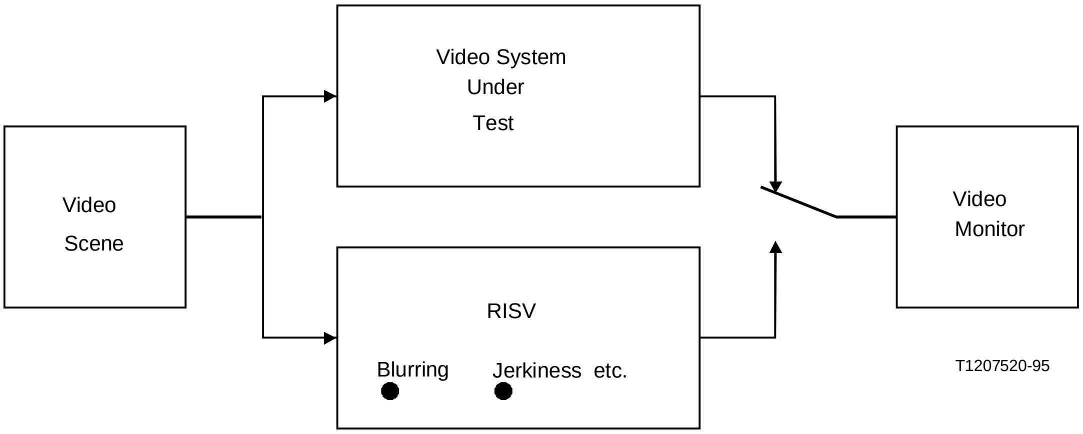

FIGURE 1/P.930

Example of RISV used for quantifying performance of a digital video system

NOTE – Depending on the implementation, the RISV may not operate in real time as it requires intensive signal processing power. Therefore, the generated scenes may need to be stored for later playback during the matching process.

5 Impairments produced by the RISV

There are numerous impairments which the RISV should simulate. At present the five impairments that a RISV can be designed to simulate are: block distortion, blurring, edge busyness, noise and jerkiness. Other impairments that should be simulated, such as error blocks and colour distortion, are the subject for future study.

5.1 Block distortion

Block distortion (also known as blocking or blockiness) is often caused by coarse quantization of the spatial frequency components during the encoding process. For a given quantization level, block distortion is usually more visible in smoother areas of the picture. One way to simulate block distortion is by changing the luminance values in a block of pels in such a way as to make that block of pels distinguishable from its neighbouring pels.

For instance, in many coders which use the Discrete Cosine Transform (DCT), the pels are separated into blocks of 8 × 8 pels. In the MPEG-1 coder for example, there are 1320 blocks in a SIF image derived from the CCIR 601 525-line format, 44 in the horizontal direction (352 pels/8) and 30 in the vertical direction (240 lines/8). As block distortion is commonly seen in smooth high motion areas, a simulation may involve impairing a subset of the blocks found in these areas by changing the luminance value of the selected blocks according to a numerical algorithm. The position and number of blocks impaired along with the intensity of the change in luminance determines the perceptibility and level of the impairment.

Other methods for simulating block distortion are subject for future study.

5.2 Blurring

Blurring is the reduction in sharpness of edges and spatial detail and is often caused in compression algorithms by trading off bits to code resolution and motion. Image formats such as SIF, CIF and QCIF are lower in resolution than the CCIR 601 image format and compared to it are effectively blurred even before compression.

The blurring impairment can be simulated by applying a symmetric, two-dimensional FIR (Finite duration Impulse Response) low-pass filter to the digital image array. Several filters with varying cut-off frequencies can be provided to allow control over the degree of blurring. Other methods of simulating the blurring impairment are being considered.

5.3 Edge busyness

Edge busyness is often caused by the use of coarse quantization levels during the encoding process within a block containing both edges delimiting smooth areas and pels with a significantly different average level. The result is to produce distortion concentrated at the edges of objects, characterized by temporally varying sharpness or spatially varying noise.

Edge busyness can be simulated by adding close-in, leading and lagging positive or negative echoes to the digital image in the vertical and/or horizontal directions to create a halo around objects edges. The displacements of the echoes can be varied to control the perceived degree of edge busyness. A shimmering of the halo can be created by changing the echo displacement every several frames.

Echoes in the digital 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). Thus, echoes can be produced by filtering the image with a multiband filter containing passband ripples.

5.4 Noise

There are many types of noise impairments commonly produced by compression algorithms. Two common impairments are mosquito noise and quantization noise. Mosquito noise is a form of edge busyness characterized by moving artifacts and/or blotchy noise patterns superimposed over objects. The algorithm to simulate mosquito noise is a subject for future study. Quantization noise may be created in the quantization step during the compression coding process. This noise looks like a random noise process (snow) and can be a grey or coloured noise, but is not uniform over the image.

For the case of block-based encoding schemes, the quantization noise impairment can be simulated by replacing the luminance value of pels at random locations with a constrained random value. The random location to impair can be determined by drawing two random numbers, the first which is a position for the row location of the pel and the second which is a position for the column location of the pel. After a pel location is determined, the unimpaired luminance value for the selected pel is replaced by a random value chosen out of the range specified by the lowest and highest luminance level (or expanded range) found within the block which encloses the pel location to impair.

Additional pel locations are selected until the desired number of pels is impaired. The number of pels impaired is an indication of the level of the quantization noise impairment.

5.5 Jerkiness

Jerkiness or jerky motion is motion that was originally smooth and continuous perceived as a series of distinct "snapshots". It is commonly observed on video telephones and other low-bit rate video systems. It is the result of skipping video frames to reduce the amount of video information that the system is required to transmit or process per unit of time. The resulting frame rate is the level of this impairment.

Jerky motion can be simulated by repeating video frames. For instance, when a 30 frame-per-second system is used, to create a 10 frames-per-second rate every third frame is used to replace the next two frames in a video sequence. Other methods of simulating jerkiness that correlates the amount of motion in a sequence to the severity of the jerkiness impairment are being studied.

5.6 Combination of impairments

A RISV should be capable of producing simulations of at least the following five types of impairments: block distortion, blurring, edge busyness, noise and jerkiness. The RISV may also simulate other impairments, such as colour errors, mosquito noise and error blocks, but these are subject for further study. For each of the five defined impairments, the RISV should be capable of adjusting the level of the impairment over a range from imperceptibility of the visual impairment to a very annoying level of the visual impairment. The selection of maximum and minimum impairment levels is usually determined by experimentation and relies on visual observation and subjective judgement. The impairments to use and the number of impairment level steps to cover the desired range of quality is left to the experimenter. Usually the process consists of accounting for the scene content and coder being tested and appropriately selecting the proper level and range of each impairment experimentally.

A RISV should be capable of introducing more than one of the five impairments into a video scene. With multiple impairment injection, order of introduction becomes important. When adding two or more impairments to a scene, the RISV may operate in a sequential fashion. The scene frames are processed for the first chosen impairment, then the frames are reprocessed again for each chosen impairment. The following issues must be addressed when deciding the sequence of impairments:

1) It is most efficient, from a processing time perspective, to first drop frames to simulate jerkiness before other impairment processes. It is also important to prepare the same frame sequence that the simulated compression encoder would deal with when (selecting a quantizer and) producing block distortion. In most cases it is desirable that the motion area detection use this sequence with the decreased number of frames, while the edge detection processing time also benefits from operating on fewer frames.
2) All detection operations should be conducted on this reduced frame sequence and before applying the other impairments (consistent with item 1). For example, accurate edge detection must be conducted before adding blurring and motion areas must be identified after adding jerkiness.
3) Blurring must be added before other localized impairments to avoid modifying their levels.
4) Again, processing efficiency suggests that frame repetition (to complete the simulation of jerkiness) should take place last. This is also needed to correctly portray other impairments, since they should not change in repeated frames.

Further considerations are sought as part of continuing study.

Figure 2 illustrates the same order of impairments used by VIRIS, a specific implementation of an RISV (see Appendix I). Each of the impairment levels can be independently adjusted over the appropriate range or the RISV can have a master control that can adjust the levels of all the impairments simultaneously. However, use of a master control is not yet well-understood and is thus the subject for future study.

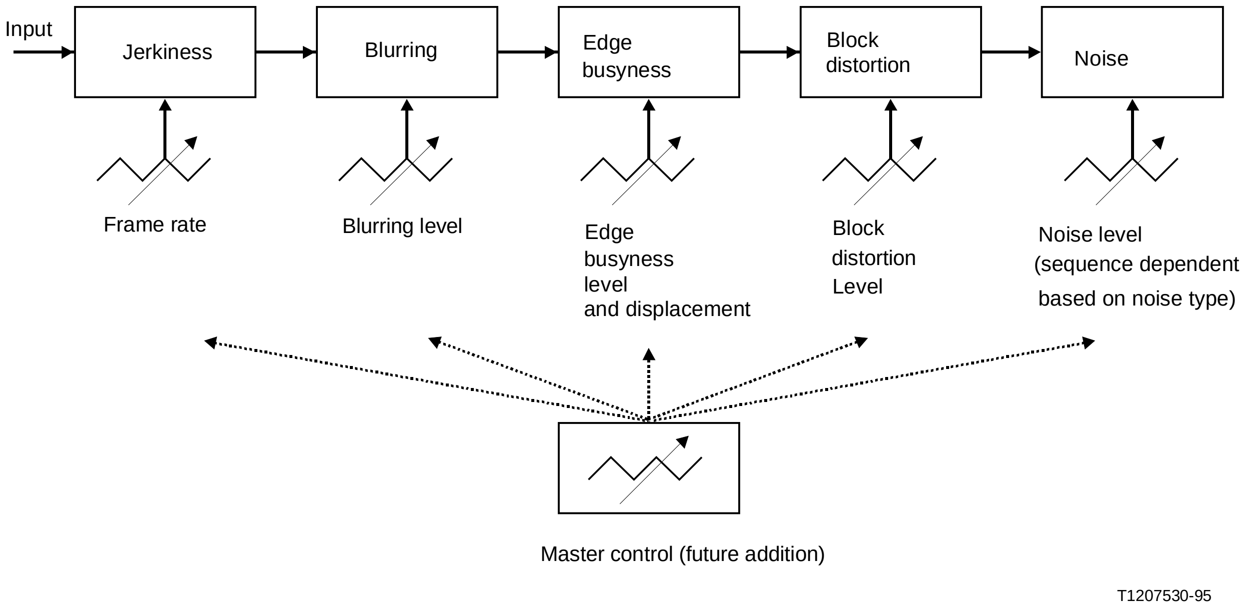

FIGURE 2/P.930

Block diagram of RISV showing order of impairments and level controls

6 Conclusion

This Recommendation describes the principles of an adjustable video reference system that can be used to generate the reference conditions necessary to characterize the subjective picture quality of video produced by compressed digital video systems. Use of reference conditions during a subjective test allows for relative comparisons between reference video conditions and video conditions of interest, and thus a more reliable measurement of video quality. This Recommendation describes three applications for an adjustable reference system and the impairments that such a system should produce. Although the characteristics of a RISV are described, before an implementation can be recommended, validation tests are required.

Appendix I

VIRIS, a specific implementation of a RISV

I.1 Introduction

The purpose of this appendix is to report efforts to design a Video Reference Impairment System (VIRIS) which simulates the impairments of block distortion, blurring, edge busyness, jerkiness and noise. This system is intended as a general video laboratory tool for evaluating the performance of digital video coders. VIRIS (sometimes referred to as VIRIS1, a more recent version of the Video Reference Impairment System) is a software system implemented in the C programming language to manipulate digital video files to introduce simulated coding impairments into a video image. It is designed to operate on SIF images but the method can be applied to other image formats such as CIF, QCIF and CCIR 601 format. VIRIS is in a preliminary stage and is useful only as a general purpose laboratory tool.

The next subclause discusses and defines the methods used to simulate the impairments in VIRIS. Subclause I.3 describes the calculation of Peak Signal-to-Noise ratio (PSNR). Subclause I.4 describes the characteristics of the latest implementation of VIRIS. Subclauses I.5 and I.6 describe:

1) the subjective test plan for determining the impairment ranges to produce appropriately wide ranges of picture quality performance; and
2) the results of those subjective tests performed to characterize the relationship between impairment level and objective quality. Conclusions are given in I.7.

I.2 Impairment simulation

The input image format for MPEG-1 compression algorithms is SIF, which is a 30 Hz, progressive, non-interlaced format of 240 × 352 pels of Y (luminance signal) and 120 × 176 pels of U and V (chrominance signals). SIF is converted from the 30 Hz, interlaced, CCIR 601 format of 480 × 720 pels of Y and 480 × 360 pels of U and V. The resolution of SIF is half that of CCIR 601 in the vertical and horizontal directions. It was chosen as the input format for VIRIS because of VIRIS's potential use for evaluating 1.5 Mbit/s video codecs utilized within Video Dial Tone services. However, the methods described in this subclause for simulating impairments can be applied to other image formats with only minor changes to account for the differences in image formats.

The C programme, VIRIS, operates on one frame of SIF video at a time. Only the luminance pels are manipulated in creating the simulated impairments – the colour pels are not changed.

The next five subclauses detail the methods used in VIRIS to simulate the block distortion, blurring, edge busyness, noise and jerkiness impairments in SIF video images.

I.2.1 Block distortion

Block distortion is often caused by a too coarse quantization during the compression process which results in a distortion or loss of high frequency components. The result of such a distortion is the visibility of the underlying block encoding structure. For a given quantization level, block distortion is usually more visible in smooth areas of a sequence.

A block of 8 × 8 pels is the smallest coding block used in MPEG-1 algorithms. There are 1320 such blocks in a SIF image, 44 in the horizontal direction (352 luminance pels/8) and 30 in the vertical direction (240 lines/8).

Block locations to impair are selected by identifying the regions of the image where one would expect block distortion to appear. These are typically smooth areas with associated motion. The following is the process which VIRIS uses to simulate the block distortion artifact [1]:

a) Edges in the current frame are identified by filtering the zero-padded frame[1] using the Sobel [8, 10] filtering technique. As block distortion is commonly seen in smooth areas of an image, by identifying the edges of an image, the smooth areas are inherently located where compression-generated block distortion is commonly visible.
b) Block distortion is placed within those smooth areas of the image which indicate movement.

The manner in which movement or change in Temporal perceptual Information (*TI* for short) is determined is to perform a subtraction of pixel luminance intensities of location (x,y) across successive frames.

---

[1] "Zero-padding" is the addition of a band of zeros around the sides of the frame. This band of zeros will be of a width greater than or equal to (N-1)/2 pixels, where N is the number of filter taps used in the largest VIRIS filter.

$$TI(x,y,t+1) = Y(x,y,t+1) - Y(x,y,t) \quad \text{(I.2-1)}$$

where $Y(x,y,t+1)$ is the zero-padded luminance image at time $t+1$, $Y(x,y,t)$ is the zero-padded luminance image at time t and $TI(x,y,t+1)$ is the temporal perceptual information at time $t+1$ and pixel (x,y). As block distortion is commonly seen in areas of motion, to determine the block location to impair, the *TI* is accumulated for the current and past frame for each block of the present scene:

$$\sum_{i=-7}^{0} \sum_{j=-7}^{0} abs(TI(8r+i, 8s+j, t+1)) \quad \text{(I.2-2)}$$

[If $(8r+i,8s+j)=(x,y)$ is an edge at t or $t+1$, $TI(.)=0$, and $r=1$ to 44, $s=1$ to 30, the subsampled block position of the SIF image.]

As the larger this sum, the greater the block motion, a search is performed across the entire frame to locate the block (r,s) with the greatest indication of motion (or equivalently block with largest *TI*). If this block also contains no more than 5 pels within it that have been classified as edges[2], the block is impaired and the position is eliminated in further determinations of block distortion locations for that frame. Based on the desired number of blocks to impair, the process is repeated to locate the next block with largest degree of motion located within a smooth region of the frame.

After the proper blocks to impair have been selected, the block distortion can be simulated by changing the luminance values in the block of pels in such a way as to make that block of pels distinguishable from its neighbouring pels. The impairment process is as follows: firstly, the average luminance value of the 64 pels in the block, P_{ave}, is determined. Each pel value in the block is then set to the sum of the two quantities:

1) the average between the current pel luminance value and the block luminance, P_{ave};
2) a randomly selected number between –2 to 2. When adding blocking to an image sequence, the location of the blocks is changed every 15 frames to hold the distorted block position constant such that the eye perceives the impairment.

I.2.2 Blurring

Blurring is the reduction in sharpness of edges and spatial detail in a picture. This impairment is mainly caused because of a compression algorithm trading off bits to code resolution and motion. Image formats such as SIF, CIF and QCIF are inherently lower in resolution than the CCIR 601 image format and compared to it appear blurred even before compression. The blurring impairment is created in VIRIS by applying a symmetric, one-dimensional, fifteen-tap FIR (Finite duration Impulse Response) low-pass filter to the SIF digital image array in the horizontal direction. Six filters with varying cut-off frequencies were designed using a digital filter design software package [6]. The selection of specific design cut-off frequencies was based on a trial and error procedure. A few frames of a SIF video image were processed with different low-pass filters and informal comparison tests of the filtered and unfiltered images were performed. Specific filters were then selected to cover a wide range of picture quality. These filters have design cut-off frequencies of 1.5, 1.0, 0.75, 0.5, 0.375 and 0.25 MHz and correspond to blurring impairment levels of 1 to 6, respectively. The coefficient values for the 15 taps of each filter are given in Table I.1.

---

[2] Through experimentation it has been determined that if the quantity returned through applying the Sobel filter to 3 × 3 blocks of pels is greater than 500, the pel is classified as an edge for the purposes of the block distortion impairment.

TABLE I.1/P.930

Coefficient values for low-pass filters

Filter tap	Filter tap coefficient					
	Blurring impairment level (LPF cut-off frequency, MHz)					
	1 (1.5)	2 (1.0)	3 (0.75)	4 (0.5)	5 (0.375)	6 (0.25)
[–7],[7]	–1	–2	–3	0	2	–2
[–6],[6]	1	–1	–3	–3	1	2
[–5],[5]	3	1	–1	–5	–4	4
[–4],[4]	6	5	3	–3	–6	–3
[–3],[3]	10	9	8	5	–1	–9
[–2],[2]	13	14	15	15	13	3
[–1],[1]	15	17	20	24	28	31
[0]	16	19	22	28	34	47

The filtering is carried out in VIRIS by applying the low-pass filter to the SIF luminance image array along each of the 240 rows. The filtered output signal, y_n, is given by:

$$y_n = \frac{1}{S} \sum_{i=-7}^{7} h_i \, x_{n-i} \qquad \text{(I.2-3)}$$

where:

x_n input SIF image array

h_i filter coefficients (Table I.1)

$S = \sum_{i=-7}^{7} h_i$

$0 \le n < 352$

In Equation (I.2-3), the input signal outlining the vertical frame boundaries, x_n, is extended by seven samples at the beginning and ending of each line to permit calculation of the blurred output signal near the frame boundaries. The value assigned to these extended samples equals the value of the luminance pixel which borders the edge.

I.2.3 Edge busyness

There are two different algorithms that were developed to implement the edge busyness impairments:

1) the edge busyness impairment is applied to only the vertical edges of objects;

2) the edge busyness impairment is applied to both the vertical and horizontal edges.

The algorithm for single directional edge busyness impairment simulation is no longer recommended. The following subclauses describe the latest VIRIS edge busyness impairment simulation.

I.2.3.1 VIRIS implementation of edge busyness

Edge busyness distortion is caused by too high a quantization level in a block containing both a smooth area and some pels with a significantly different average level. Edge busyness is simulated in VIRIS by adding close-in, leading and lagging negative echoes to the SIF image in both the vertical and the horizontal directions to create a halo around objects' edges. The echoes have design displacements of 0.375, 0.5 and 0.75 micro-seconds and the shimmering of the halo is created by changing the echo displacement every five frames.

As VIRIS was originally designed to evaluate the performance of MPEG-1 encoders, the edge busyness impairment simulation was designed to best simulate the edge busyness impairment as seen in MPEG-1 video. It was determined by experimental observation that the edge busyness artifact in MPEG-1 video occurs both on the sides of objects (i.e. vertical edges) and on top and bottom of objects (i.e. horizontal edges); hence the simulation algorithm employs a two-dimensional filtering algorithm. In addition, although MPEG-1 does introduce a blurring artifact, by filtering the image to simulate the halo effect, it was not necessary to introduce additional blurring to the picture to simulate the MPEG-1 blurring artifact. The echoes used to simulate the edge busyness impairment also tended to blur the edges in the picture.

I.2.3.2 VIRIS edge busyness simulation method

As alluded to above, the edge busyness impairment is simulated in VIRIS by using a filter with ripples in the passband amplitude response to create an echo impairment. 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 are produced by applying 13-tap digital, multiband filters [6] to the SIF image array first in the horizontal direction and second in the vertical direction. Table I.2 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 VIRIS programme.

The coefficient value of tap 0 of each of the filters is scaled to a value of 175. With the exception of one pair of tap weights for each of the three echo displacement filters, all other filter tap weights are 0. The placement of filter tap having a non-zero coefficient value relative to the centre tap determines the echo displacement while the value of the coefficient determines the echo amplitude. The actual coefficient value is input by the user to the VIRIS programme and can range from -1 to -30, in integer steps. The same coefficient value is applied to each of the three filters.

The filtering is implemented in VIRIS 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 by:

$$z_n = \frac{1}{S} \sum_{i=-6}^{6} h_i \, x_{n-i} \qquad \text{(I.2-4)}$$

where:

x_n input SIF image array

h_i filter coefficients (see Table I.2)

$S = \sum_{i=-6}^{6} h_i$

$0 \le n < 352$ (horizontal)

$0 \le n < 240$ (vertical)

In Equation (I.2-4), the luminance value of the input signal outlining the vertical and horizontal frame boundaries, x_n for n=0, n=352 (horizontal) and n=0, n=240 (vertical), is extended by six samples beyond the left, right, upper, and lower frame boundaries to permit calculation of output signal z_n.

TABLE I.2/P.930

Coefficient values for 13-tap, multiband filters

Filter Tap	Filter tap coefficients		
	Echo displacement, (µs)		
	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

The approach used to arrive at the above simulation method was experimental. It involved a trial and error process in dealing with different ways of adding echoes and then comparing results with MPEG-1 pictures. Changes were made in the simulation method to attempt to improve the simulation, but these changes could not be done in real time since it was necessary to recompile the VIRIS programme to produce the changes. Nevertheless, it is believed that a reasonable simulation of MPEG-1 impairments resulted from the above simulation method. Of course, the perceptibility of real MPEG-1 impairments is correlated with motion in the picture, whereas in VIRIS most of 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 design of an adjustable video reference impairment system is that it be simple to implement and be independent of any coding scheme. This precludes, at present, correlation of motion in the picture with the simulated impairment.

I.2.4 Noise

There are two types of noise produced by VIRIS, quantization noise and signal correlated noise.

I.2.4.1 Quantization noise

Quantization noise is a noise sometimes created in the quantization step of the compression process. This noise looks like a random noise process but is not uniform over the image. It is simulated in VIRIS by replacing the luminance value of pels at random locations with a random value between 16 (black) and 255 (peak white). The random location is determined by drawing two random numbers, the first with a value between 1 and 240 for the row location of the pel and the second with a value between 1 and 352 for the column location of the pel. After a pel location is determined, the pel luminance value is replaced with a random value between 16 and 255. That pel location is then eliminated in further determinations of random noise pel locations for that frame. The number of noise pel values that can be changed in a particular frame ranges from 1 to 84 480 (352 × 240 pels) and is determined from user input to VIRIS. With each successive frame, the random number generator seed is changed as part of the input data to VIRIS so that the noise pel locations change from frame to frame.

I.2.4.2 Signal correlated noise

Signal correlated noise is a term utilized to describe the appearance resulting from the combination of edge busyness and mosquito noise impairments as seen in compressed video. Signal correlated noise is a distortion typically seen around the edges of moving objects and is characterized by a fine grained, "halo" noise pattern superimposed near and on edges of objects in motion. A simulation of the combination of these impairments requires extracting edges within a scene, determining whether they are in motion, and finally injecting a noise impairment [1].

The edges are located in a scene using the Sobel filtering technique. If the value returned by applying the Sobel filter over the nine adjacent pixels is greater than 50, the pixel location is classified as an edge. If the pixel location also has an associated degree of motion (determined by differencing the luminance value of the pixel location across two successive frames and verifying that the resulting *TI* is greater than 2) then that pixel location is flagged as one to impair.

To impair the pixel location, a random number between $-\beta$ and β is added to the luminance value of the unimpaired pixel.

$$Y(x,y,t) = Y(x,y,t) + \text{random}\ (-\beta, +\beta) \qquad \text{(I.2-5)}$$

where $Y(x,y,t)$ is the luminance value of the pixel location (x,y) at time t identified as an edge in motion.

For each pixel location satisfying the conditions of being an edge in motion, the luminance value of the unimpaired image is altered by an amount between $-\beta$ and β. The level of the impairment, specified by β, controls the intensity of the impairment. Increasing β, increases the visibility of the impairment. To correlate the impairment more effectively to the temporal variation, the pixels to impair are actually the union of pixels identified as edges in the previous and present frame.

I.2.5 Jerkiness

Jerkiness is defined as motion, originally smooth and continuous, perceived as a series of distinct "snapshots". It is commonly observed on picture telephones and on some low-bit rate video teleconferencing systems. It is the result of repeating video frames to reduce the amount of video information that a system is required to transmit or process. VIRIS utilizes a Frame Repetition Factor, FRF, to control the level of jerkiness present. An FRF of 3, for example, results in every third frame used to replace the next two frames in the video sequence. The Effective Frame Rate, EFR, is calculated as 30/FRF. The EFR for the example FRF of 3 would, thus, be 10 frames per second.

I.3 Calculation of the Peak Signal-to-Noise Ratio (PSNR)

Currently each of the simulated impairments in VIRIS is objectively characterized by calculating the Peak Signal-To-Noise Ratio (PSNR) over each processed frame and the average PSNR over all the frames of a processed picture sequence. The unweighted PSNR is one measure for assessing the distortion of the processed sequence. For each processed frame, k, the RMS noise, N_{rms_k}, is computed as:

$$N_{rms_k} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left[U_{ij} - I_{ij}\right]^2}{N \cdot M}} \qquad \text{(I.3-1)}$$

where:

U_{ij} luminance value of unimpaired pel at row i and column j location

I_{ij} luminance value of impaired pel at row i and column j location

N 240 for SIF image

M 352 for SIF image

k kth frame

To calculate the PSNR over a sequence of K frames, the per-frame average noise, N_{rms_k}, is first determined in Equation (I.3-1). Next, the average noise, N_{rms}, across the sequence of K frames is calculated in Equation (I.3-2) as follows:

$$N_{rms} = \frac{1}{K} \sum_{k=1}^{K} N_{rms_k} \qquad \text{(I.3-2)}$$

Finally, the PSNR is calculated using Equation (I.3-3) as follows:

$$PSNR = 20 * \log_{10} \frac{S_p}{N_{rms}} \qquad \text{(I.3-3)}$$

where S_p equals the number of levels to which the luminance intensity is quantized. The system on which VIRIS operates quantizes the luminance pels to 8 bits, or $S_p = 255$.

Because the unweighted PSNR as described above is loosely correlated to the human visual system, a more accurate objective measure may be required. This is a subject for future study.

I.4 VIRIS programme

Shown in Figure I.1 is the process of how VIRIS impairs one frame of SIF video at a time. VIRIS injects impairments into a SIF image converted from a CCIR 601 image (using the *preproc* programme) stored on a video disk. With the VIRIS levels stored in data files, frames are processed and the impaired sequence is converted back to CCIR 601 (using the *postproc* programme) format and transferred to a digital video disk for viewing.

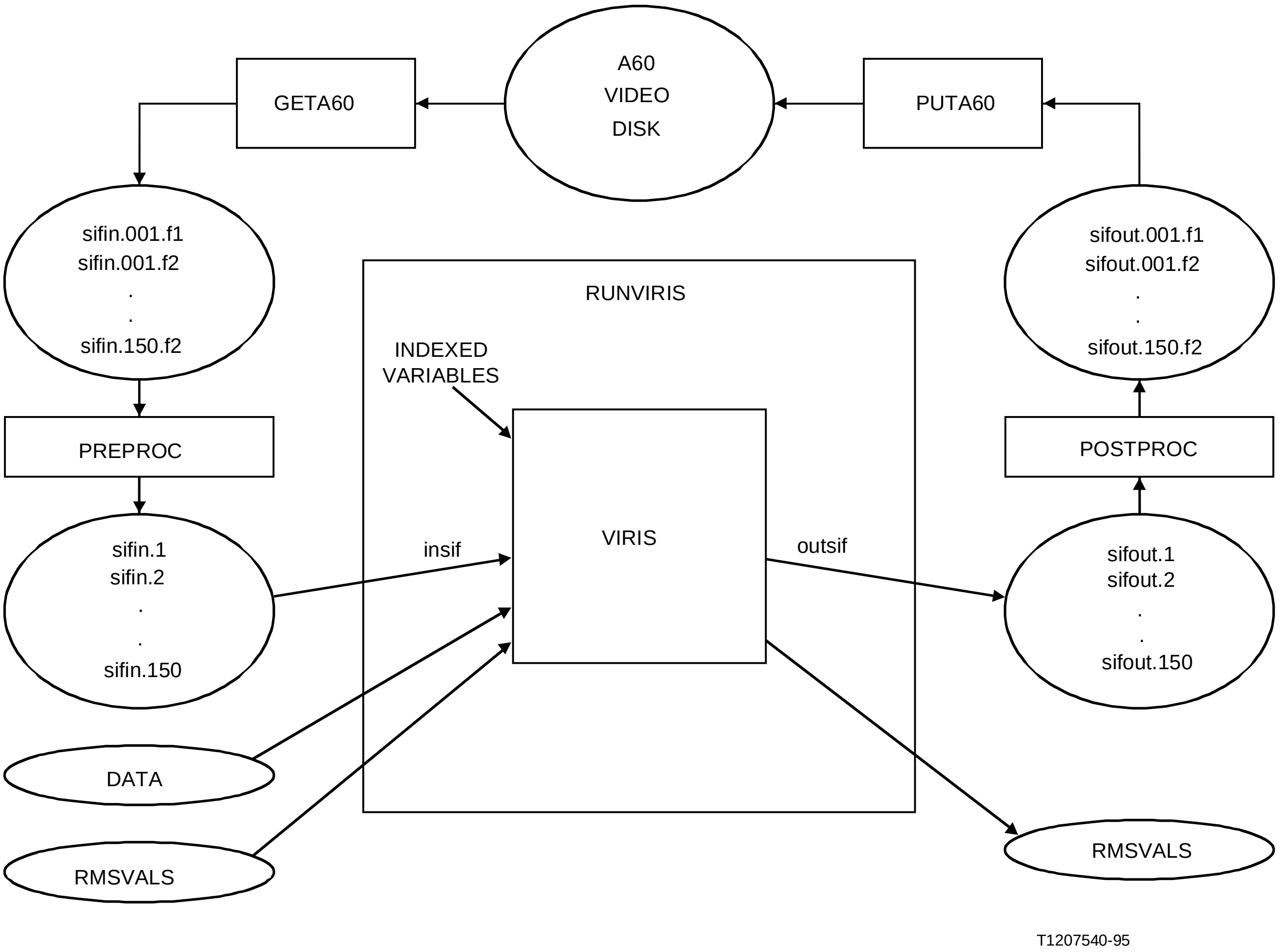

FIGURE I.1/P.930

Operation of VIRIS

The data file controls the level of impairments added to a video sequence. There are essentially six impairment level control parameters to operate VIRIS:

a) The block distortion level, input as a whole number and represents 0.1 per cent of the 1320 total blocks to be changed (result is rounded to nearest whole number). For example, a block distortion level of 10 results in 13 blocks that are impaired (10 × 0.001 × 1320 = 13.2 which rounds to 13).

b) The quantization noise level, input as a whole number and representing 0.001 per cent of the 84 480 total luminance pel values (rounded to nearest whole number) to be changed. For example, a noise level of 10 results in a change in the luminance value of 8 pels (10 × 0.00001 × 84 480 = 8.4 which rounds to 8).

c) The signal correlated noise level, input as a whole number, which represents the range of luminance values by which a pel can be altered. For instance, a level of 10 indicates that the luminance value of pels classified as edges can be randomly altered by a range from –10 to 10 luminance levels from its original value.

d) The blurring level, input as whole numbers between 0 and 6 with 0 indicating no blurring. The numbers 1 to 6 select low-pass filters with cut-off frequencies of 1.5, 1.0, 0.75, 0.5, 0.375 and 0.25 MHz.

e) The edge busyness echo displacement which consists of a whole number between 0 and 3, with 0 selecting no edge busyness simulation and 1, 2 or 3 selecting 0.5, 0.75 and 0.375 msecs displacements.

f) The edge busyness echo amplitude level. The input data item is an integer number between –30 and –1 and represents the filter tap coefficient value for the particular echo displacement selected.

With the VIRIS implementation, the jerkiness level is specified by the Frame Repetition Factor, FRF, input as a command line variable. Using the data file, VIRIS injects impairments into the SIF image file based on the specified levels. After each frame is impaired, the PSNR for the current frame and average PSNR across all processed frames are calculated.

I.5 VIRIS subjective test plan

Subjective tests were performed to determine the relationship between impairment level and subjective opinion for impairments simulated by VIRIS. In general, the test procedures and recommendations given in ITU-R Recommendation BT.500-6 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 non-expert observers rate the quality of 10-second TV picture sequences containing various levels of the VIRIS impairments. They used the discrete 5-point quality rating scale that ranges from EXCELLENT to UNSATISFACTORY. The test plan and results provided in the following two subclauses describe the relationships between impairment level and subjective opinion for a subset of the impairments discussed in I.2. Specifically, subjective tests have not been carried out at this time to develop relationships between impairment level and opinion for the block distortion, signal correlated noise, and jerkiness impairment based on the algorithms described within this appendix. The information presented below was extracted from the relevant clauses of [2].

I.5.1 VIRIS picture sequences

Three picture sequences were selected for the tests to characterize the noise and blurring impairments. 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 were 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:

a) "Bond" – This sequence was obtained from a laser disk. It is a sequence of two people, Bond and a woman, riding double on a horse in a desert scenario. The sequence has little detail and a slow "horse-walking" motion.

b) "Chase" – This sequence was obtained from a laser disk movie. It is a sequence of a young man on a skateboard being chased by several other young men. The sequence has considerable detail and motion with three scene cuts.

c) "Football" – This sequence was derived from a 20-second sequence that had been used in previous tests to evaluate MPEG-1 algorithms. It is a sequence from a football game that has considerable detail and moderate motion.

I.5.2 VIRIS test conditions

For the noise and blurring impairments, each picture sequence was processed by VIRIS 12 times (2 impairments × 6 impairment levels). The impairment levels were selected based on the results of informal viewing tests and on one limited pilot test with non-expert observers. Table I.3 shows the PSNRs calculated by VIRIS for each of the three pictures and the average PSNR across the three pictures for all impairment levels. The table also shows the data input to VIRIS (read from a data file) for each of the 6 levels of the 2 impairments.

TABLE I.3/P.930

PSNRs for VIRIS impairment levels

Impairment level	VIRIS input	Average PSNR, dB			
		Bond	Chase	Football	3-Pic Average
QN1	1	60.8	60.4	59.6	60.3
QN2	3	55.3	54.9	54.2	54.8
QN3	7	52.2	51.9	51.1	51.7
QN4	15	48.7	48.4	47.8	48.3
QN5	62	42.6	42.3	41.6	42.2
QN6	125	39.4	39.1	38.5	39.0
BLR1	1	47.2	41.8	42.7	43.9
BLR2	2	43.4	38.1	38.1	39.9
BLR3	3	40.0	35.4	34.2	36.5
BLR4	4	38.6	33.7	32.6	35.0
BLR5	5	36.3	31.8	30.1	32.7
BLR6	6	34.2	30.2	28.0	30.8

QN Quantization Noise
BLR Blurring

The PSNR varies considerably across the three picture sequences for the blurring and noise impairments. The variation of blurring impairment is on the order of 1 to 7 dB depending on the level. The PSNR variation across the sequence for the noise impairment is considerably less, on the order of 1 dB.

Plots of the PSNR, averaged across the three pictures, versus the input to VIRIS for the two impairment levels are shown in Figures I.2 and I.3. An exponential function fitted to the data is also shown on each of the plots to provide a pathway from PSNR to VIRIS input for each of the impairments.

As discussed previously, the 150 frames of each 5-second picture sequence processed by VIRIS were converted from the SIF to CCIR 601 image format and then transferred to the disk. Each sequence was then transferred to a digital tape twice in succession to make a 10-second sequence. Because this subjective test for characterizing the noise and blurring artifacts also characterized the older block distortion and edge busyness impairments, there was actually 72 picture sequences processed by VIRIS (3 sequences × 4 impairments × 6 impairment levels) along with the three unprocessed reference pictures of SIF quality. These sequences were stored on a tape. After all of the test material had been stored on digital tape, it was recorded on a laser video disk recorder. This system allows for high quality NTSC recording and random access playback of the recorded material. It was controlled by a computer to play back the test material in different random orders as required by the test design.

I.5.3 VIRIS test method

A single stimulus test method using a discrete 5-point quality scale was used in the tests. 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.

A total of 23 non-expert observers, selected from an out-of-house subject pool, participated in the tests. They included three males and twenty females with ages ranging from 25 to 67 with an average age of 52.9. Each had normal visual acuity (equal to or better than 20/25) and normal colour vision.

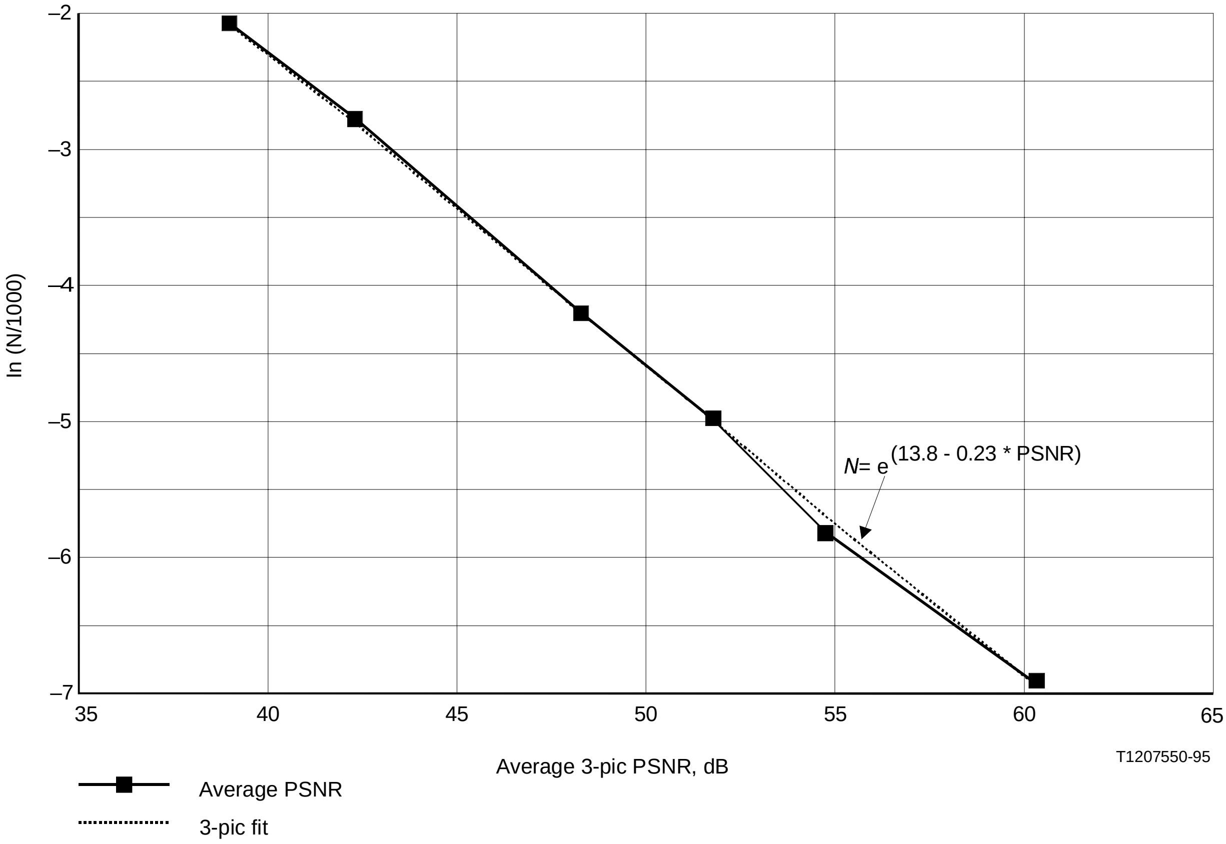

FIGURE I.2/P.930

Quantization noise PSNR vs QN data input to VIRIS

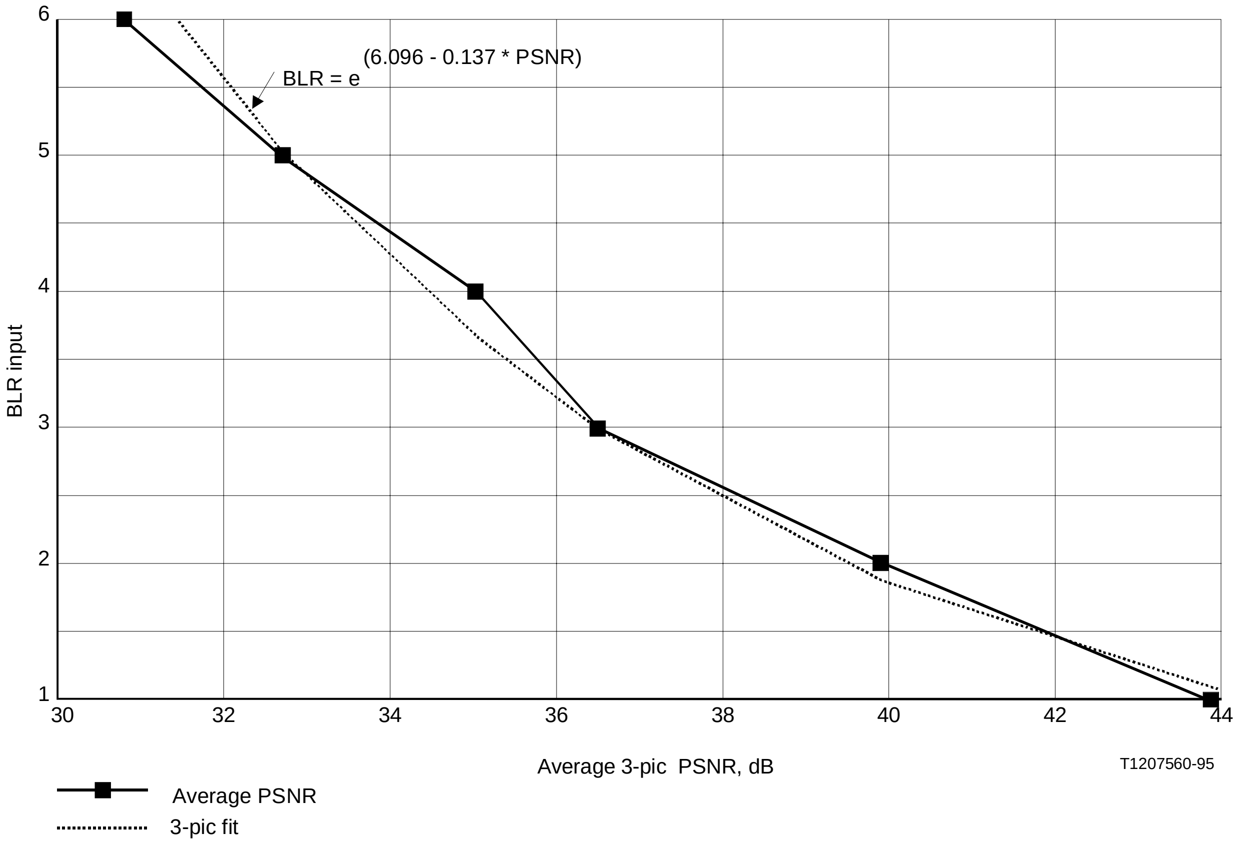

FIGURE I.3/P.930

Blurring PSNR vs BLR data input to VIRIS

I.5.4 VIRIS viewing conditions

The observers who took the test viewed a television receiver. They sat at a viewing distance of 7.5 times the height (7.5*H*) of the picture (about 7.5 feet from the picture screen) which is midway between the two viewing distances that had been used in previous tests of compressed bit rate algorithms. A viewing distance of 5*H* is in the middle of the range recommended by ITU-R and a 10*H* distance is about the average viewing distance in the United States. The 7.5*H* distance used in the study was a compromise between a distance (5*H*) which the author believed was too close for comfortable viewing in the present tests and the desire for more critical judgements than could be obtained at the 10*H* distance.

The test room was approximately 12 × 20 feet with light grey, sound absorbing walls and controlled, incandescent ceiling lighting. Ambient lighting and screen luminance were measured with a digital photometre. The ambient lighting in the test room was dimmed to approximately 6.5 lux at the subject seating positions by adjusting the dimmer controls for the ceiling lamps. The resulting ambient lighting at the receiver was approximately 1.1 lux. The receiver controls were adjusted to provide a peak luminance of approximately 24.8 millilambert and a minimum luminance of approximately 0.16 millilambert for a contrast ratio of about 150. Black burst and 100 IRE test signals from the digital generator were used in making these adjustments.

I.5.5 VIRIS test procedures

A total of 81 basic test conditions were utilized during the subjective test. The conditions consisted of 72 SIF sequences (3 sequences × 4 impairments × 6 impairment levels) processed by VIRIS (the older block distortion, edge busyness algorithms and current blurring, quantization noise algorithms) and 3 repeats of each of the three unprocessed SIF video sequences. Each 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 81 conditions was tested twice for a total of 162 conditions.

Three observers were tested at a time, except for one test where only two were tested (one called subject failed to appear). They were seated in front of the receiver at a viewing distance of 7.5H. They filled out their name and seating position on the first page of a notebook which contained a voting page for each practice and test condition. The test administrator then read the test instructions. After answering any questions, a practice session of 11 conditions was then conducted to familiarize the subjects with the test procedures and with the test material.

The test was administered in two parts with a 20-minute refreshment break between the parts and a short 2- to 3-minute stretch break within each part after 40 sequences were shown. Each part included the 81 basic test conditions and took about 30 minutes to complete. The conditions in each part were presented in a pseudo-random order with two constraints:

a) the first, second and third group of 27 conditions each contained all of the impairment levels for the four impairment types including the three SIF reference pictures; and

b) each of the three groups of 27 conditions 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 under way, 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.

I.5.6 VIRIS subjective test results

I.5.6.1 Data analysis

Data analysis consisted of computing the Mean Opinion Score (MOS) and 95% confidence intervals for each of the impairment levels and SIF reference, for each of the three picture sequences. The subject voting opinions of EXCELLENT, GOOD, FAIR, POOR and UNSATISFACTORY were first assigned the numbers 5, 4, 3, 2 and 1, respectively. The MOS for each condition was calculated as:

$$MOS = \frac{1}{N}\sum_{i=1}^{N} V_i \qquad \text{(I.5-1)}$$

where:

V_i Vote

N number of votes

The total number of votes for the impairment conditions was 46 (23 subjects × 2 repeats) and for the SIF reference conditions was 138 (23 subjects × 6 repeats). The 95 percent confidence intervals (CI) about the MOSs were computed from the 23 subject means for each impairment level and SIF reference from Equation (I.5-2) [4]:

$$95\%\ CI = \mu \pm \frac{t \cdot \sigma}{\sqrt{N}} \qquad \text{(I.5-2)}$$

where:

95% CI	95% Confidence Interval
μ	MOS
t	two – tailed t – statistic for 0.05 significance level = 2.069
σ	standard deviation of subject means
N	number of subject means (23)

I.5.6.2 VIRIS results

The results, including PSNR, MOS and 95% CI are shown in Table I.4 for all impairment levels and SIF reference for each of the three picture sequences.

Only the results acquired for the impairment simulation algorithms as described in I.2 are presented. Although the actual subjective test characterized a set of four impairments, block distortion, edge busyness, blurring and quantization noise, the algorithms to simulate the first two impairments have been modified, and therefore the results obtained during this subjective test do not apply to the current simulation algorithms. Subclause I.6 will describe a second subjective test to characterize the edge busyness impairment based on the simulation technique outlined in I.2.

MOSs for the three reference pictures, Bond, Chase and Football were 4.0, 3.8, and 4.2, respectively. These opinion ratings are somewhat low for reference conditions but not unexpected since the SIF image format limits picture quality because of its limited horizontal and vertical resolution. In a previous test using continuous quality scale, the average rating of six SIF picture sequences was about 60, where 60 to 80 covered the GOOD quality opinion region. The present results with SIF picture sequences are not inconsistent with previous results.

The width of the 95% CIs varied from 0.3 to 0.7. In general, CI widths became narrower as picture quality decreased, indicating that the subjects were more in agreement about their judgements of poor picture quality than about their judgements of good picture quality. Normally, when the impairment range covers the full range of the comment scale, the CI widths will be narrowest at both ends of the comment scale and wider for qualities in between.

TABLE I.4/P.930

Subjective test results for VIRIS impairments

Impairment	Picture Sequence								
	Bond			Chase			Football		
Type/Level	PSNR (dB)	MOS	CI Width *X*	PSNR (dB)	MOS	CI Width *X*	PSNR (dB)	MOS	CI Width *X*
Ref-SIF	-	4.0	0.7	-	3.8	0.6	-	4.2	0.7
QN1	60.8	3.9	0.6	60.4	3.8	0.6	59.6	3.9	0.6
QN2	55.3	3.4	0.6	54.9	3.6	0.6	54.2	3.3	0.6
QN3	52.2	2.7	0.5	51.9	3.4	0.6	51.1	3.3	0.6
QN4	48.7	2.5	0.4	48.4	3.0	0.5	47.8	2.7	0.5
QN5	42.6	1.8	0.4	42.3	1.8	0.4	41.6	2.2	0.4

Impairment	Picture Sequence								
	Bond			Chase			Football		
Type/Level	PSNR (dB)	MOS	CI Width X	PSNR (dB)	MOS	CI Width X	PSNR (dB)	MOS	CI Width X
QN6	39.4	1.8	0.4	39.1	1.7	0.3	38.5	2.0	0.4
BLR1	47.2	3.8	0.7	41.8	3.3	0.6	42.7	3.7	0.6
BLR2	43.4	3.4	0.6	38.1	2.7	0.5	38.1	3.0	0.5
BLR3	40.0	3.0	0.5	35.4	2.0	0.4	34.2	2.6	0.5
BLR4	38.6	2.3	0.5	33.7	1.7	0.4	32.6	2.1	0.4
BLR5	36.3	1.9	0.4	31.8	1.5	0.3	30.1	1.3	0.3
BLR6	34.2	1.6	0.3	30.2	1.3	0.3	28.0	1.3	0.3

PSNR	Peak Signal-to-Noise Ratio
MOS	Mean Opinion Score
95% Confidence Interval	MOS X
Ref-SIF	Reference SIF
QN	Quantization Noise
BLR	Blurring

Plots of MOS as a function of PSNR are shown for the three picture sequences in Figures I.4 and I.5 for the noise and blurring impairments, respectively. A logistic function fitted to the three picture data of each of the impairments is also shown on each of the figures. The logistic function is of the form given in Equation (I.5-3).

$$MOS = K_1 + \frac{K_2}{1 + e^{-K_3(PSNR - K_4)}} \tag{I.5-3}$$

where:

K_1	lower MOS asymptote = 1.0
K_2	upper MOS asymptote - K_1 = 3.2
K_3	slope parameter
K_4	PSNR value where MOS = $K_1 + \frac{K_2}{2}$

In fitting the logistic function, the lower asymptote, K_1, was set at 1.0 and the upper asymptote, $K_1 + K_2$, was set at 4.2 in all cases. The lower asymptote value of 1.0 is the lowest MOS that can result with the 5-point comment scale. The upper asymptote value of 4.2 represents the highest average MOS obtained in the present study for one of the SIF reference conditions (Football sequence). The two fitting parameters were, thus, K_3, a slope parameter, and K_4, the value of PSNR where the MOS is midway between the low and high asymptote values.

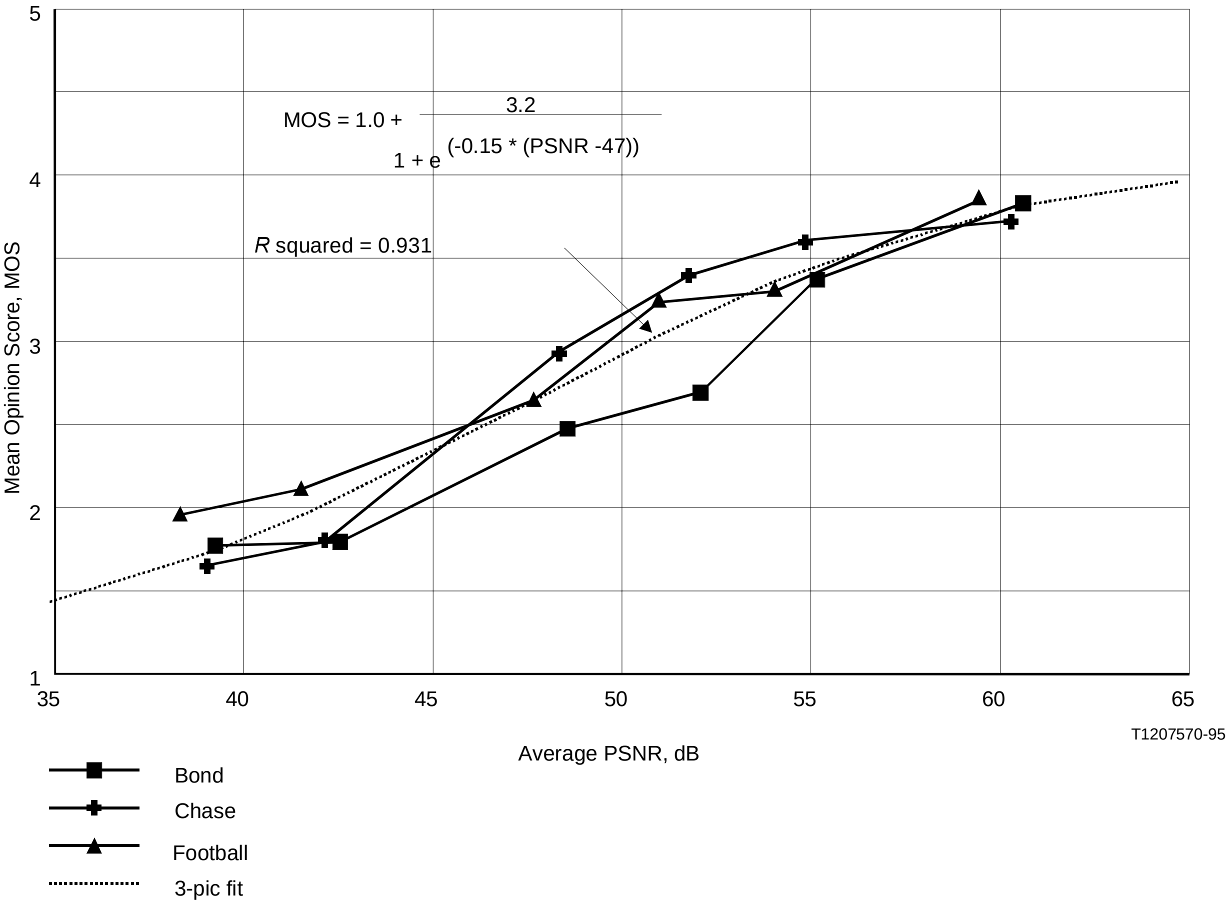

FIGURE I.4/P.930

Quantization noise PSNR vs MOS

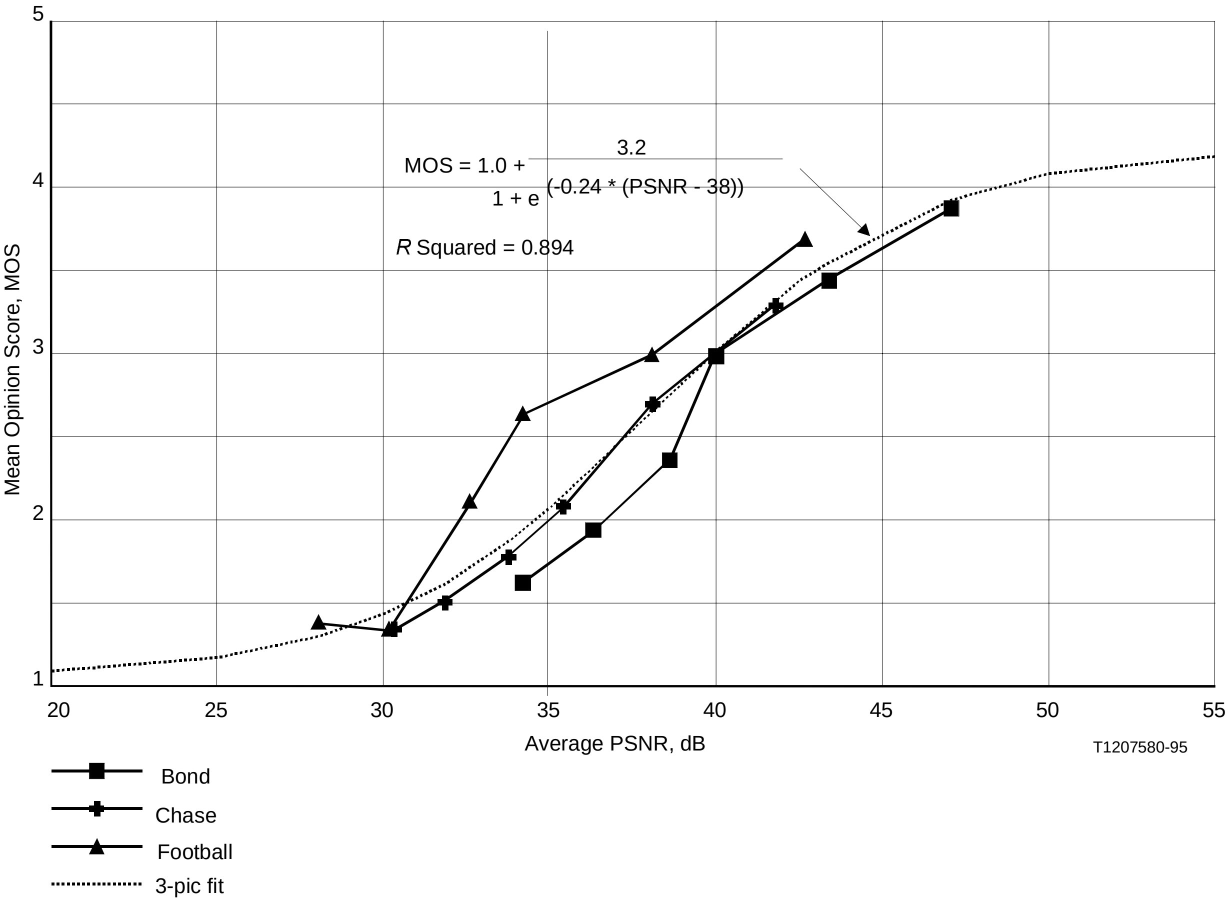

FIGURE I.5/P.930

Blurring PSNR vs MOS

Referring to the quantization noise results in Figure I.4, subjective reaction to the noise PSNRs was somewhat different depending on the particular picture being viewed. However, the range of MOSs for the three pictures is fairly compact and the logistic function appears to be a good fit to the three picture data with *R* squared = 0.9309. (*R* squared is a goodness of fit statistic that ranges from 0 to 1 with 1 representing an optimal fit).

Figure I.5 indicates that subjective reaction to blurring PSNRs was more sensitive to picture than was the case for noise and block distortion PSNRs, with somewhat more spread between the three picture MOSs. The logistic function is still a good fit to the three picture data with *R* squared = 0.8941.

I.6 Edge busyness subjective test plan

A second subjective test was performed to determine the relationship between impairment level and subjective opinion for the edge busyness impairment as described in I.2. Other conditions were also tested and these included a SIF reference condition, an NTSC reference condition and two MPEG-1 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 Recommendation BT.500-6 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 non-expert observers rate the quality of 10-second TV picture sequences representing various levels of the VIRIS impairment, the SIF and NTSC reference conditions and the MPEG-1 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 subclauses.

I.6.1 Edge busyness 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:

"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 demonstration clip of HDTV which was downloaded to NTSC format.

"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 three 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 for evaluating MPEG algorithm performance.

"Football" – This sequence was used in the previous VIRIS test. 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 demonstration clip of HDTV which was downloaded to NTSC format.

I.6.2 VIRIS test conditions and test material preparation

Eight edge busyness impairment levels were selected as test conditions based on informal laboratory viewing tests. Table I.5 shows the PSNRs calculated by VIRIS for each of the edge busyness impairment levels, first as an average across each picture sequence and then as an average across the three pictures.

TABLE I.5/P.930

PSNRs for VIRIS edge busyness impairment levels

Edge busyness level, EB	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 – Edge Busyness level, EB, corresponds to filter tap coefficient values for the variable range of coefficients given for the three multiband filters of Table I.2.				

The PSNRs of the football and table tennis sequences were fairly close, differing by amounts from 0.5 to 2.4 dB across the edge busyness level range. The PSNR of the race cars sequence differed from those of the other two sequences by amounts from 0.6 to 5.2 dB across the edge busyness level range. Figure I.6 shows a plot of the three picture average PSNR versus the edge busyness input level to VIRIS. A linear function fitted to the data is also shown on the plot to provide a pathway from PSNR to VIRIS edge busyness level input.

In addition to the eight edge busyness impairment conditions, the test included conditions of NTSC quality, SIF quality and two conditions of MPEG-1, 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 laser video disk recorder. 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.

I.6.3 Edge busyness subjective test method

A single stimulus test method using a discrete 5-point quality scale was used in the tests. A total of 27 non-expert 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.

I.6.4 Edge busyness subjective test procedures

Viewing conditions were the same as used in the first VIRIS test (see I.5.4). 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.

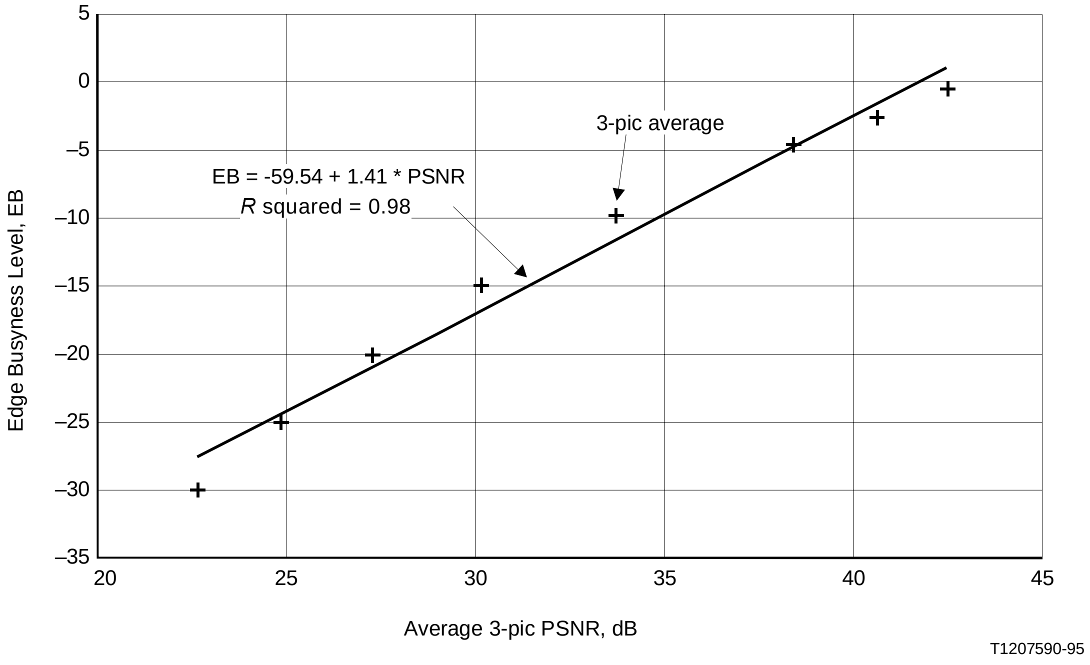

FIGURE I.6/P.930

Edge busyness PSNR vs EB data input to VIRIS

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:

a) the first, second and third group of 12 picture sequences each contained all of the 12 basic test conditions; and

b) 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 under way, 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.

I.6.5 Edge busyness subjective test results

The Mean Opinion Score (MOS) was computed for each of the 12 test conditions for the three picture sequences. The MOS of the NTSC reference conditions were 4.7, 4.6 and 4.6 for the race cars, football and table tennis sequences, respectively. MOSs 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 MPEG-1 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 MPEG-1 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 VIRIS test. However, the earlier test did not use NTSC reference conditions as was done in this test 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 past test higher than they would have if a full range of stimulus material had been presented. The purpose of including the NTSC conditions in the second subjective test was to obtain full coverage of the comment scale to reduce or eliminate this type of anchoring problem. The SIF quality obtained in the second test is, thus, probably more indicative of actual SIF quality than the results obtained in the first subjective test.

Figure I.7 shows plots of MOS as a function of the VIRIS edge busyness impairment PSNR for the three picture sequences. A logistic function fitted to the three 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 I.7, subjective reaction to the edge busyness 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 three picture sequences, the logistic function provides a good fit to the three 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.

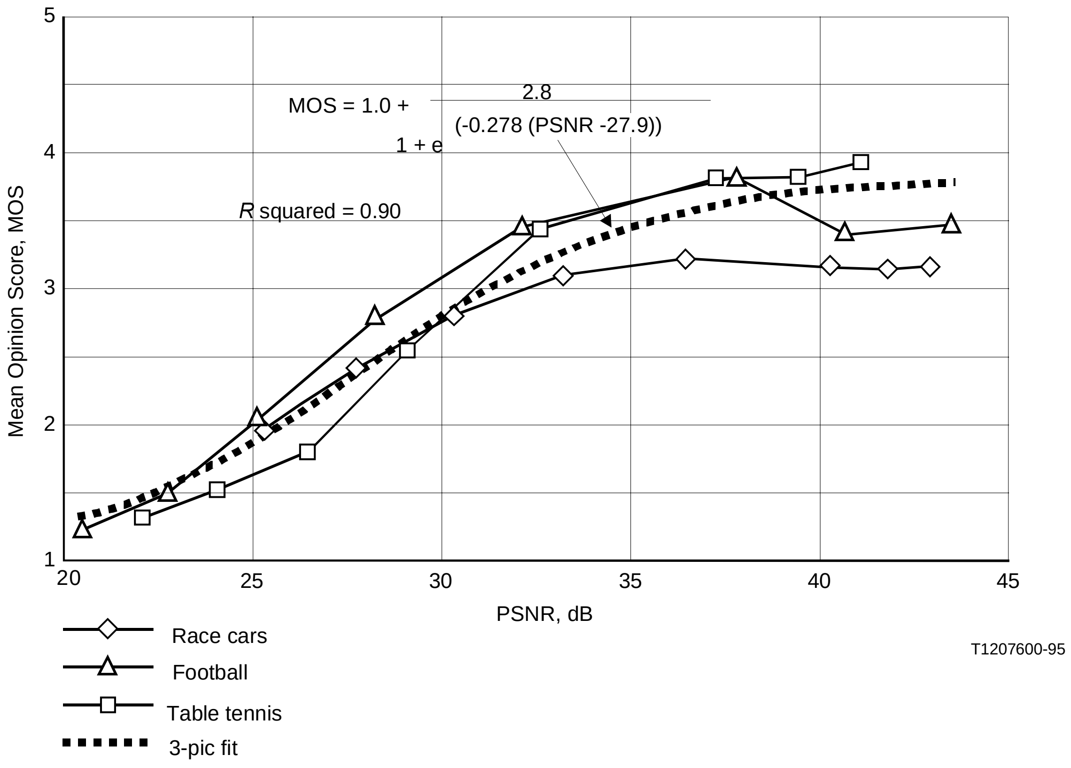

FIGURE I.7/P.930

Edge busyness PSNR vs MOS

I.7 Conclusions

An adjustable video reference impairment system is needed for the picture quality performance evaluation of compression digital video systems. Such a system needs to produce a wide range of picture quality by introducing simulated picture impairments similar to those produced by the digital coders. The Video Reference Impairment System (VIRIS) described in this appendix is a beginning towards that end.

The VIRIS RISV can introduce simulations of block distortion, blurring, edge busyness, signal correlated noise, quantization noise and jerky motion impairments to SIF picture sequences. These impairments can be added singly or in combination. All of these impairments can be objectively characterized in terms of the impairment level and average Peak Signal-To-Noise Ratio (PSNR) over all of the frames in the picture sequence. Subjective test results for three of these impairments (block distortion, signal correlated noise and jerkiness has not been characterized subjectively) have indicated that subjective reaction to the simulated impairments covers a wide range of picture quality. These relationships are shown in Figures I.4, I.5, and I.7 for quantization noise, blurring and edge busyness, respectively, for each of the three picture sequences tested. A logistic function fitted to the three picture data relates MOS and PSNR for each of the three impairments.

Figures I.2, I.3 and I.6 show plots of the PSNRs averaged across the three pictures and the input level to VIRIS for each of the three impairments. An exponential function fitted to the data is also shown on each of the plots to relate PSNR to VIRIS data input for each of the four impairments and a linear function was used for VIRIS. Use of these functions along with the logistic functions shown in Figures I.4, I.5 and I.7 provides a pathway between desired MOS and VIRIS data input for each of the four impairments for those scenes used in the test results. Using these functions to predict the results for other scenes is not recommended at this time and is a topic for further study.

Further work needs to be done with VIRIS. This includes the simulation and subjective testing of various combinations of the impairments. The aim is to relate desired subjective opinion, MOS, to VIRIS data input for various combinations of impairments. The result may be in the form of a single relationship for a combination of all impairments or, perhaps, multiple relationships (which would include the five developed in the current study) for specific combinations of impairments. Future work involves learning more about the actual impairments that are created by compression algorithms in digital video coders, how to simulate these impairments, and designing additional methods to correlate objective measures of quality to a subjective opinion.

Bibliography

[1] ITU-T COM 12-53 (1995), *Addition of Motion Correlated Artifacts to VIRIS*, Bellcore.

[2] ITU-T COM 12-21 (1993), *VIRIS, An Experimental Video Reference Impairment System*, Bellcore.

[3] ITU-T COM 12-35 (1994), *Modification of Video Reference Impairment System (VIRIS) to improve Edge Busyness Impairment Simulation*, Bellcore.

[4] HAYS (William L.): Statistics For Psychologists, *Holt, Rinehart and Winston*, 1963.

[5] ANIL (Jain): Fundamentals of Digital Image Processing, *Prentice Hall*, 1989.

[6] *Digital Filter Design Package*, Version 2.12, Atlanta Signal Processors Incorporated, December 1986.

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Telephone network and ISDN
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media
Series H	Transmission of non-telephone signals
Series I	Integrated services digital network
Series J	Transmission of sound-programme and television signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound-programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminal equipment and protocols for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communication
Series Z	Programming languages