Université d'Ottawa · University of Ottawa

Operational (In-Field) Calibration Methodology of Video Quality Evaluation of MPEG2 Compressed Data for Broadcast

By

Husam R. Hassan

A thesis submitted to the

Faculty of Graduate And Postdoctoral Studies

In partial fulfillment of the requirements for the degree of

Master of Applied Science

In

Electrical and Computer Engineering

Supervised by:

Prof. Emil Petriu

Ottawa-Carleton Institute for Electrical and Computer Engineering

School of Information Technology and Engineering

Faculty of Engineering

University of Ottawa

September 2001

Contents

i- List of Figures

ii- List of Tables:

iii- Abstract:

This Thesis discusses the different techniques used in subjective and objective measurement for MPEG-2 compressed video for broadcast according to the ITU-R BT-500 recommendation for both world standards ATSC and DVB.

The technique for subjective/objective quality correlation described in ITU recommendations and VQEG work [VQEG00] for Double Stimulus Continuous Quality Scale (DSCQS) & Single Stimulus Continuous Quality Evaluation (SSCQE) is used to generalize an application for In-Field operational calibration standard of subjective / objective digital video evaluation for TV network and broadcast stations.

A practical experiment for validating the correlation between the subjective evaluations of some MPEG2 coded streams vs. the objective measurement obtained using the IFN / Rohde & Schwarz single stimulus blockiness measurement algorithm is realized, demonstrating the correlation to other measurements and the ease and applicability to perform this in the field.

iv- Acknowledgement

I would like to express my deep thanks and gratitude to Prof. Dr. Emil M. Petriu for his guidance and support throughout the full course of my Masters studies and Thesis.

I also would like to express my deep thanks and appreciation to the team of the Digital TV and Video research group at the Communication research Center – Industry Canada, Ottawa, Canada: Philip Corriveau (currently at Intel Corp.), Ron Reneaud & Andre Vincent for their help and support during my work, and allowing me to use their facility for the practical part of my study.

My special thanks to the whole team of division 7 of Rohde & Schwarz in Germany, USA and Canada for their continuous help and support during my work, especially:

1. Thomas Bichlmaier, Germany
2. Harald Ibl, Germany
3. Peter Foulger, Canada
4. Alexander Woerner, USA

There are a number of people who have helped me with papers, support or encouragement including the VQEG members. I especially want to thank:

5. Scott Daly, Sharp, USA
6. Charles Fenimore, NIST, USA
7. Brian Flowers, VQEG
8. Prof. Mohamed Ghanbary, Univ. of Essex, UK
9. Margaret Pinson, NTIA, USA
10. Stefan Winkler, EPFL, Switzerland
11. Stefan Wolf, NTIA, USA

My biggest thanks goes to my parents and family for their continuous help, support and encouragement throughout my whole life, and my daughter Nadine, who have sacrificed many weekends without going to the movies or having fun!

Chapter 1

Introduction

"The basic purpose of television systems is to extend the sense of vision and hearing beyond their natural limits" [Bru99]

The first television systems were mechanical, and then they became electronic. The first color broadcast was an NTSC in the USA in 1953. SECAM followed in France in 1957, then PAL in Germany in 1961. The next revolution was in HDTV, which had some momentum going for it, but didn't pick up worldwide until the introduction of digital television.

The first success of digital TV was in post-production applications, where the high cost of digitizing the video stream was compensated by the capability of limitless layering of effects while preserving the same quality. This is one of the huge drawbacks of analog TV. The use of digital TV could only be extended if the bandwidth of data could be reduced. As the digitized video stream of a standard definition picture is over 200 MBits/sec.(ITU-601 standard), and the new digital HDTV standard (ITU-292) is about 1.5 Gbits/sec., it's extremely difficult and costly to manage this bandwidth in storage and transmission. The cost of transmission is directly proportional to the bit rate transmitted.

Thus compression emerged as a practical solution to this problem. "Compression is the science of reducing the amount of data used to convey information" [Sym98]

A typical bit rate for standard definition signal, which is currently in our digital Satellites and digital cable systems, is around 2-8 Mbits/sec., and around 20 Mbits/sec. for HDTV. This is almost a 100:1 compression ratio, and this puts a big question mark on the quality of signal compared to the original one.

Compression is not new to Television. Interlace is a simple form of compression giving a 2:1 reduction in bandwidth. The use of color difference signals C_B, C_R instead of RGB is another

form of compression. "The human eye is more sensitive to luminance than Chrominance"[Tek98].

When color TV was introduced, the composite color systems like PAL, SECAM and NTSC were a form of compression, because the color was superimposed on the monochrome signal, and used the same bandwidth.

"With the variety of video compression methods in use and being developed, especially in the encoding and decoding, there is a strong requirement for picture quality evaluation, which is independent of the compression algorithm and it's related artifacts" [Tek98].

1.1 The problem of TV Networks

Currently there are a few systems for objective video quality evaluations based on some of the algorithms mentioned in this thesis. Subjective video quality evaluation has been the dominant method of quality measurement at TV stations, TV Networks and distribution companies. The later are sometimes cable companies, satellite companies or even phone utility companies. They would have a group of professional quality evaluators referred to sometimes as "Golden Eyes" to see the programming material and sense any problems in quality. However with the increased number of TV channels to monitor, and the need for TV networks to increase that number on the available transport streams and links they have, it has been a problem to maintain quality to a certain level. In fact there has been no standards to follow in order to decide quality levels acceptable, as is the case of analog TV.

At this time, we see many satellite companies worldwide advertising hundreds of channels in their networks, and it's not difficult to calculate how many programs are put in each transport stream, and deduce that some programs are left with 1 Mbits/sec for example, which is not enough for some programs like sports.

The other issue for broadcasters or other companies involved in the distribution and transmission of digital TV signal is the newly introduced objective measurement systems. There has been a few systems introduced, mainly double-ended systems, which would test the

degradation in transmission. They are used also by manufacturers of encoders and MPEG2 codecs for design purposes. In broadcast, they are used in an out of service mode, before the actual transmission or after, and also in cases of evaluation of MPEG2 CODECs.

The first single stimulus technique system was introduced by IFN/R&S, Germany, in 1998, and a few other systems have been developed since then.

The main issue for broadcasters, is what do the readings they see really mean, and how do they correlate to the subjective viewer evaluation? How do the values correlate as well to other objective systems they may already have. And the big question after that how do they use all that to create a standard of quality for MPEG2 compressed video that is homogeneous all over the network.

1.2 Thesis Organization and Contribution

Chapter 2 gives a general overview of the MPEG2 compression standard, giving a glimpse on the DCT transform, the construction of the stream with I,P and B frames, and goes up to the blocks, macroblocks and slices, and then the whole structure of MPEG transport stream with the different tables, flags and bit structure of the MPEG2 packets. It demonstrates along the way, the factors used in compressions that may affect quality.

Chapter 3 gives a brief history of TV, and the standards adopted in the World today for digital TV broadcast. Added to that, there is a description of some of the main digital tape standards used in TV stations and Networks worldwide, describing the level of compression used in them. This is another cause of quality degradation in the broadcast cycle.

Chapter 4 describes the basics of video quality. This is the core of the whole subject. Describing the human eye concept of vision, luminance and chrominance sensitivity in the human vision system (HVS) and perceived concept of quality. Then we describe the different video artifacts known in video streams. Then describe the ITU-500 recommendation for classifying the methods for subjective quality evaluation, followed by corresponding objective

quality evaluation techniques adopted. Then describing types of video quality metrics for objective systems' performance.

Chapter 5 gives an over view of the World standard algorithms developed for objective quality evaluation including 9 double stimulus systems, and one single stimulus system. These were presented at the VQEG evaluation round done in 1998. There has been some more standards developed afterwards, and may be presented at future work for VQEG or other international bodies.

Chapter 6 discusses the VQEG work for the subjective/objective quality evaluation, and the practical test done for subjective /objective evaluation.

Chapter 7 presents the conclusion of this thesis, and the need and applicability of this new suggested standard in the Broadcast Industry. It also highlights current and future work areas and potential areas.

The contribution of this Thesis is:

Generalization of the ITU recommendations and VQEG work to develop a simple application for In-field calibration procedure for TV networks to validate the video quality subjective/objective correlation for an objective video quality monitoring system based on their individualized set of Network characterization. The correlation is realized using the Pearson linear regression technique.

Chapter 2

Background on Video Compression and MPEG2:

"The quality of presentation that can be derived by decoding the compressed multimedia signal is the most important consideration in the choice of the compression Algorithm. The goal is to provide near transparent coding for the class of multimedia signals that are typically used in a particular service" [Has97]

There are a number of compression techniques and methods used, depending on the application. They include JPEG, MPEG-1, MPEG-2, MPEG-4, MPEG-7, H261, H263, fractals and wavelets among others. It's beyond the scope of this study to get into the details of each one of those.

Since the broadcast industry has adopted MPEG2 as the standard of choice in both systems chosen worldwide (ATSC and DVB), we will focus on MPEG2, discussing the techniques involved, and the potential areas of possible signal degradation affecting quality.

2.1- Entropy Coding:

In any program material, there are 2 types of components; The real information of the signal, which is new and not repeated, and the redundant information, which can be predicted from previous information. The true information is called "Entropy". And ideal encoder would take only the entropy of the signal and encode it, leaving all the redundant information and thus saving bandwidth.

Redundancy can be spatial (Related to space or the values of one picture or frame), like a large area of pixels having almost the same value – a blue sky for example. Redundancy can also be temporal (Related to time), where the successive frames are almost identical. For example a car moving along the same background, or scenery. This stream can be represented by the

information from previous frames, plus the difference signal between frames and the motion vector of the car.

Therefore a talking announcer has much redundancy, whereas the water surface in a lake or tree leaves blowing in the wind has a large amount of information. In fact streams with water surfaces are commonly used to stress test encoders in real life applications, because of the complexity involved in compressing such streams.

Fig. 2.1 shows the entropy versus redundant information in a sample program material. There are many techniques used in Entropy coding. One of the most common ones is "Hoffman Coding". This technique uses the probability of existence of a certain symbol. For example, symbols that occur often, will use small codes, and the code becomes larger for symbols that occur less frequently.

The amount of information transferred in a symbol A that occurs with probability p is:

$$I_A = Log_2\left[\frac{1}{P_A}\right] \qquad [2.1]$$

Where I_A is the number of bits required to express the amount of information conveyed in symbol A, and P_A is the probability of symbol A occurring.

The entropy of code sequence is the average amount of information contained in each symbol of the sequence:

$$H = \sum P(s) * Log_2\left[\frac{1}{P_S}\right] \qquad [2.2]$$

Where H is the entropy of the coding representation, and s ranges through all symbols in the alphabet of symbols. Ref. to [Wes97] as an example, the entropy of the symbols in a sample of 6 symbols shown below is:

H = (1/2) X 1 + (1/4) X 2 + (1/16) X 4 + (1/16) X 4 + (1/16) X 4 + (1/16) X 4 = 2

If we use a fixed word length of 3, we could represent the 6 symbols in the table, but using Hoffman, we get an average code length of 2.

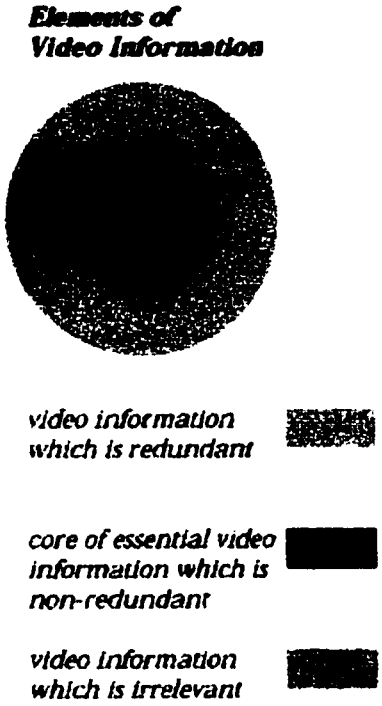


Fig.2.1 Elements of Video Information [Pan00]

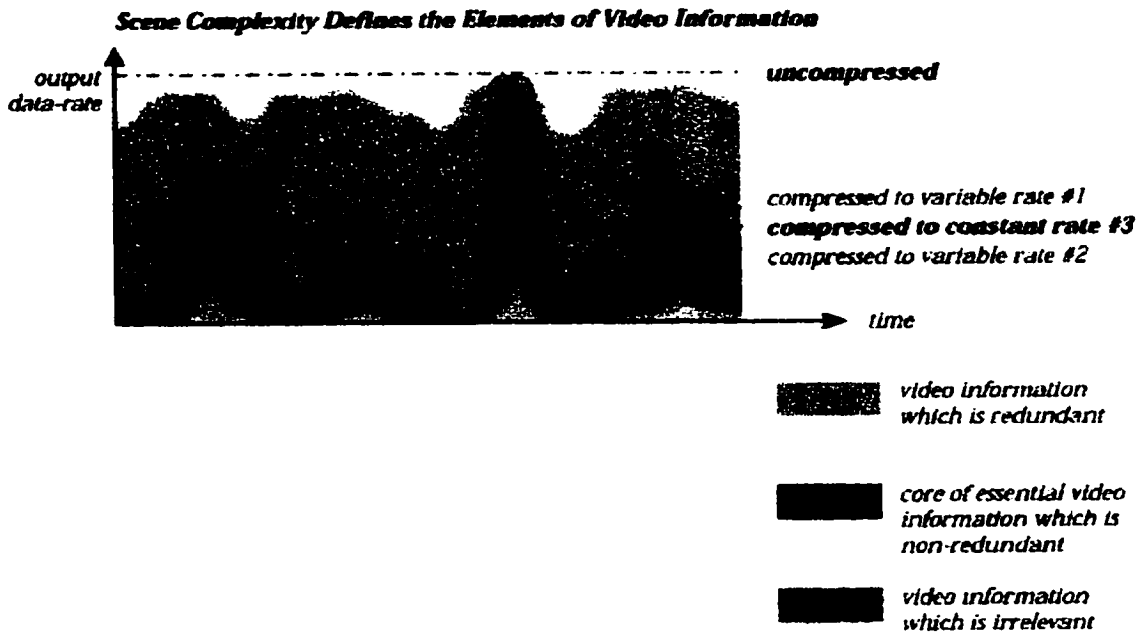


Fig 2.2 Scene Complexity [Pan00]

Intra coding (intra=within) is a technique that exploits spatial redundancy, or redundancies within one picture or frame. Inter coding (inter=between) is a technique exploits temporal redundancy.

Intra coding can be used alone as in JPEG, or along with inter-coding as in MPEG. Intra-coding depends on 2 things in typical images. First, not all spatial frequencies are present, and also, the higher the spatial frequency, the lower the amplitude.

Cutting out high frequency components from images significantly cuts the components, and this is one of the ways of compression used. The coefficients created in transforms like wavelets, FFT or DCT are much less, where most of them will be zero.

Inter-coding deals with similarities between successive images (Temporal). The picture of an object moving on the same background needs sending motion vectors to compensate for the motion between each frame and the next. Then using predicted information from other frames, this frame would be recreated at the decoder side.

The vector transmission requires less data than sending the picture difference data.

2.2- Spatial Coding:

In spatial coding, we transform the waveform into another domain, changing the frequency components in a picture into coefficients.

Fourier transform is one of the most common transforms, where we multiply the input signal by a "Basis Function" sharing different multiples of the frequency and integrating the product. When the input signal contains that frequency, there will be a coefficient from the integral giving the amplitude of that frequency. When the frequency doesn't exist, the integral will be zero at that frequency.

The Fourier transform requires coefficients for both sine and cosine components of each frequency. However, in the cosine transforms, the input waveform is time-mirrored with itself before multiplying it with the basis function. This mirroring cancels the sine coefficients and doubles all of the cosine components.

2.3- Discrete Cosine Transform (DCT)

The discrete cosine transform is the sampled or discrete version of the cosine transform and is used in 2-dimension form in MPEG.

In video display systems, we deal with pixels in a frame, and the next level of entities is called blocks. The block in MPEG case is an 8X8 pixel array. There is another term, called Macro block, which is a 16X16 pixels. A block of 8X8 pixels is transformed into a block of 8X8 coefficients via the DCT transform. The DCT equation is:

$$F(u,v) = 0.25\, C(u)\, C(v) \sum_{x=0}^{x=7} \sum_{y=0}^{y=7} f(x,y) \frac{\cos(2x+1)u\pi}{16} \frac{\cos(2y+1)v\pi}{16} \quad [2.3]$$

Where $C(u)=C(v)=\frac{1}{\sqrt{2}}$ for (u=v=0) ; C(u)=C(v) = 1 (Otherwise)

In some cases, the coefficients give longer values than the pixel length. However, the next steps of weighting or quantization and coding use those values to create a much-compressed value. The inverse DCT on the decoder side will create back the original signal. There is of course some loss induced in this process, because quantizing these coefficients into discrete values, the decoder on the other side cannot determine the original values, but rather deals with the quantized discrete values created in this process.

Fig.2.3 shows the DCT for an 8X8 block. For an 8X8 DCT block, the top left coefficient is the average brightness or DC component of the whole block. Then there are 63 AC coefficients for the

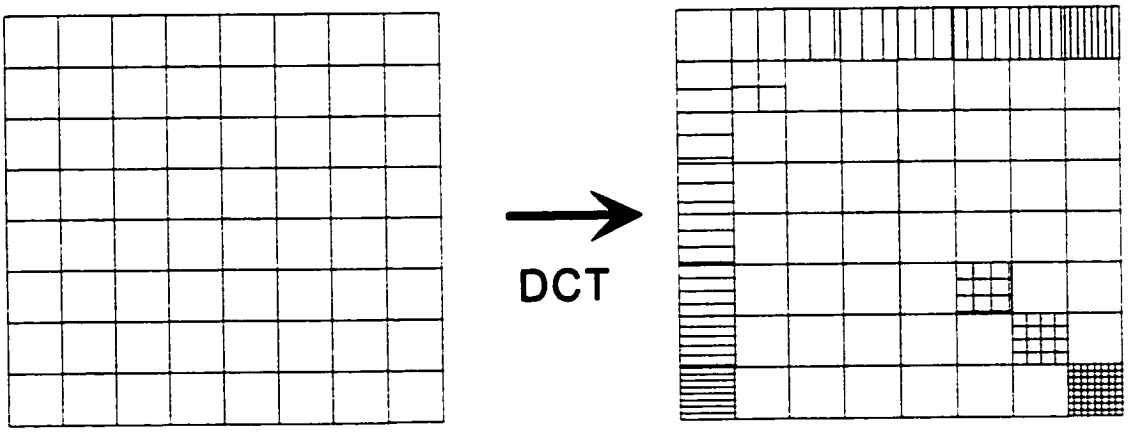


Fig 2.3 The MPEG2 Basis Function and a DCT

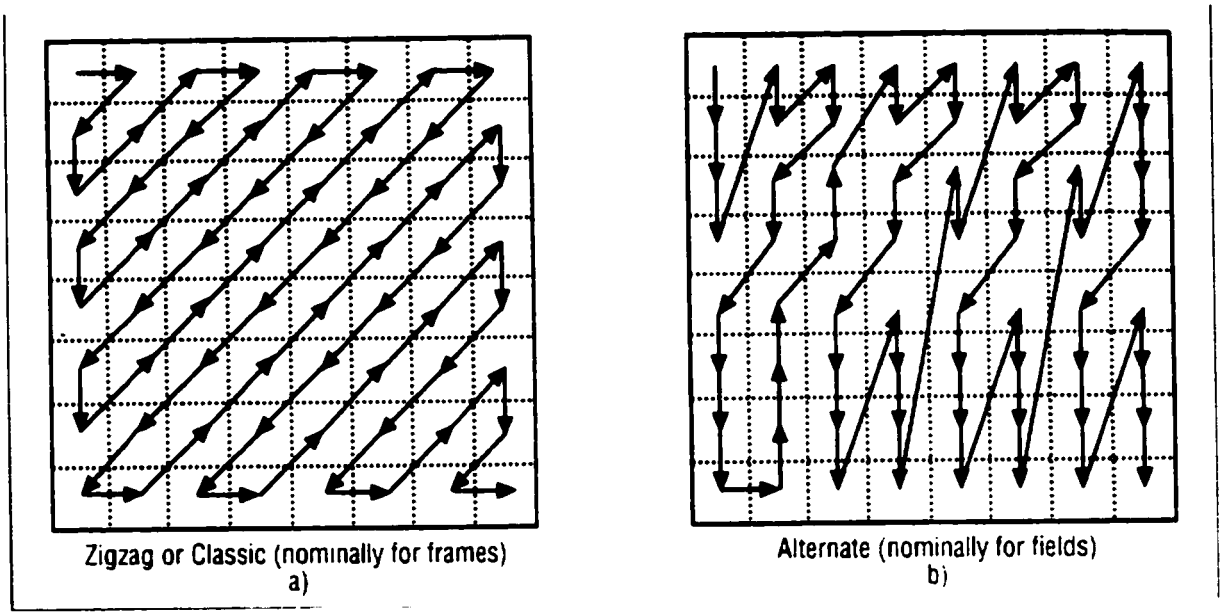


Fig. 2.4 Scanning Techniques [Tek98]

8X8 values. Going from left to right, the spatial function increases, and top to bottom the vertical frequency increases. For color difference signals C_b and C_r, there are separate 8X8 arrays other than the luminance Y array and transformed separately.

2.4- Quantization (Weighting):

"Human perception of noise in pictures is not uniform but is a function of the spatial frequency. More noise can be tolerated at high spatial frequencies" [Tek98]. Thus giving accuracy to low frequency coefficients and less accuracy to high frequency coefficients will not affect the human perception of the content. This is done in the weighting process. In the same token, the human eye is more sensitive to luminance or brightness than Chrominance (Color), and giving less accuracy to the color coefficients will also not affect the human perception as well.

In this case, low frequencies are divided by small numbers to create quantizing levels, whereas high frequencies are divided into high numbers making fewer levels. In the

inverse process of re-quantization on the decoding side, low frequency coefficients will suffer low noise added to the signal, whereas high frequency coefficients will suffer from more noise. High frequency components in a picture are like sharp building edges for example. And those could be seen blurred sometimes in MPEG compressed streams.

2.5 Scanning:

In regular programs, the top left pixels are the ones that have the significant values, and the rest of the block consists mostly of zeros. Sending those values first, then putting a string for the rest of pixels with one value which is zero, will give even more compression of data sent. This technique is called Run-length coding. The scanning methods used in MPEG accomplish this task, by scanning the pixels towards the top left corner of the block first, and moving in sort of a bottom right direction. Fig 2.4 shows 2 of the most common scanning techniques used (Zigzag and Alternate scanning).

2.6 I, P and B frames:

There are 3 types of pictures or frames in an MPEG stream. "I" is the "Intra-coded" picture that is completely encoded and when decoded doesn't need any further information.

The "P" picture is the "Forward Predicted" picture, is predicted from earlier pictures, that could be "I" or "P" frame. "P" picture data comprise of vectors that describing where in the previous picture, the macro blocks should be taken from, and transform coefficients that describe the correction or difference data that must be added to the macro blocks. "B" pictures are "Bi-directionally Predicted" from earlier or later "I" or "P" pictures. B-picture data consist of vectors describing where in earlier or later pictures data should

be taken from, and transform coefficients that provide the correction. Because of the feature of bi-directional in "B" frames, it only requires one quarter the data of an "I" frame, and half the data of a "P" frame.

There is a term called GOP or "Group of pictures" in an MPEG stream. It's an "I" frame followed by "B" and "P" frames, and the GOP ends before the next "I" frame.

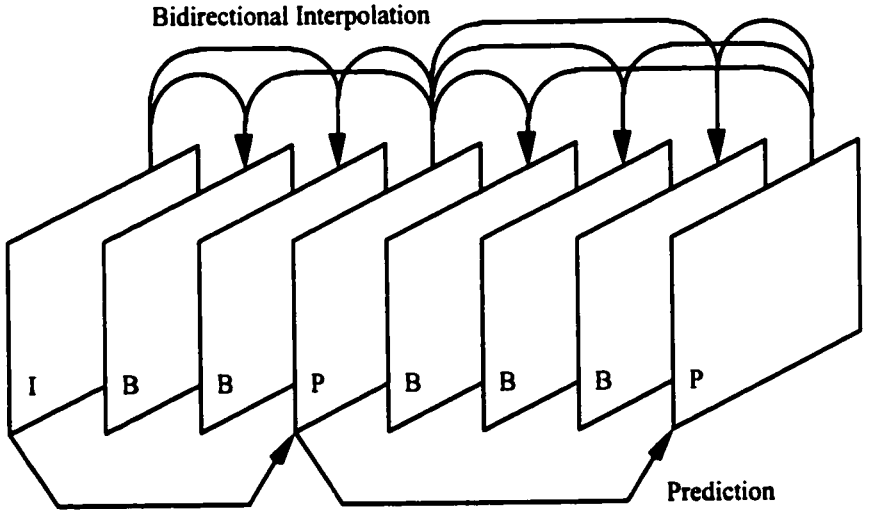


Fig 2.5 I, P and B frames in a Group of Pictures Structure (GOP)

2.7 Profiles and Levels:

MPEG can be used for a number of applications and with different performance. That's why MPEG2 is divided into different profiles, each profile has different levels. For example the profile 4:2:2 indicates 4 samples of luminance to 2 samples of both color difference components C_b and C_r. This profile is used mostly for off Air Broadcast. 4:2:0 uses less one color difference signal sampled, and has more compression though. This application is popular in cable and satellite applications.

Fig. 2.6 shows the different levels and profiles. The simple profile doesn't have B frames, thus needs simpler hardware and has less delay in decoding. The main profile is designed for large portion of applications, especially at main level.

Level	Profile								
					Spatial		High		
	Description	Simple	Main	SNR	Enhancement Layer	Base Layer	Enhancement Layer	Base Layer	4:2:2
	Frame Type	I & P	I, P & B	I, P & B	I, P & B	I, P & B	I, P & B	I, P & B	I, P & B
	Chroma Sampling	4:2:0	4:2:0	4:2:0	4:2:0	4:02:00	4:2:0, 4:2:2	4:2:0, 4:2:2	4:2:0, 4:2:2
High	Samples/line		1920H				1920H	960H	1920H
	Lines/frame		1152V				1152V	576V	1152V
	Frames/sec		60 fr/s				60 fr/s	30 fr/s	60 fr/s
	Max Bit rate		80 Mbps				100 Mbps	100 Mbps	
High 1440	Samples/line		1440H		1440H	720H	1440H	720H	
	Lines/frame		1152V		1152V	576V	1152V	576V	
	Frames/sec		60 fr/s		60 fr/s	30 fr/s	60 fr/s	30 fr/s	
	Max Bit rate		60 Mbps		60 Mbps	60 Mbps	80 Mbps	80 Mbps	
Main	Samples/line	720H	720H	720H			720H	352H	720H
	Lines/frame	576V	576V	576V			576V	288V	512/608V
	Frames/sec	30 fr/s	30 fr/s	30 fr/s			30 fr/s	30 fr/s	30 fr/s
	Max Bit rate	15 Mbps	15 Mbps	15 Mbps			20 Mbps	20 Mbps	50 Mbps
Low	Samples/line		352H	352H					
	Lines/frame		288V	288V					
	Frames/sec		30 fr/s	30 fr/s					
	Max Bit rate		4 Mbps	4 Mbps					

Fig 2.6. MPEG-2 Levels and Profiles [Sym98, Tek98, Rob98]

2.8 Macroblocks

A macroblock is a group of DCT blocks, which correspond to the information content of a window of 16X16 pixels in the original picture [Rob98]. It contains the basic number of luminance and chrominance blocks in the sampling structure used. So, in a 4:2:2 profile

signal, a macroblock contains 4 (8X8) blocks of luminance, 2 blocks of Cb color difference signal, and 2 blocks of Cr color difference signal.

[bha97] states: "When the coded bit stream has no scalable extensions, then the blocks within the macroblock are 8x8 pixels. With scalable extensions, macroblocks may contain scaled blocks with lower resolution like 1x1, 2x2 or 4x4".

The macroblocks are used also for temporal redundancy prediction. When comparing one 16X16 pixel region from one frame to the next in the following time slot, the temporal redundancy processor calculates the difference between the 2 frames. If the 2 frames have a high degree of temporal redundancy, then the difference frame would have a large number of pixels with values near zero.

The 16x16 pixel size is chosen because of the compromise between providing efficient temporal redundancy and requiring moderate computational requirements. The macroblock header contains information about the type (Y or Cb or Cr) and the corresponding motion compensation vectors.

2.9 Slices

The slice is a string of consecutive Macroblocks of arbitrary length running from left to the right of the picture. The slice could be as small as one block or as large as the whole picture. The first and last macroblocks in a slice must be transmitted, but ones in between could be skipped. A slice cannot exceed the right edge of the picture, and slices must not overlap.

The slice header contains information about its' position within the picture, and its' quantizer scaling factor, in case an error occurs and re-transmission is necessary.

2.10 MPEG Transport Stream:

The MPEG2 transport stream normally contains several elementary streams. The elementary stream is the output of the MPEG2 encoder for a single audio or video stream. The content of the transport stream is described in the program specific information table (PSI). Each transport stream contains a Program Associated Table (PAT), as well as one or more Program Map Table (PMT). The PAT is contained in the transport stream packet with the PID 0X000. It refers all the programs contained in the transport stream indicating the program number and the corresponding PID for the Program Map Table (PMT).

The elementary streams whether video, audio or data that belong to the individual programs are described in a PMT.

The elementary stream can be packetized in a packetized elementary stream (PES). The packets have a fixed length of 188 bytes. Each packet is synched with a byte (47Hex). The packets can also have the length of 204 bytes in DVB and 208 in ATSC. The different 16 or 20 bytes are used for Forward error correction before transforming the signal from base-band MPEG into a modulated RF signal. The forward error correction used in broadcast applications is (Reed-Solomon forward error correction).

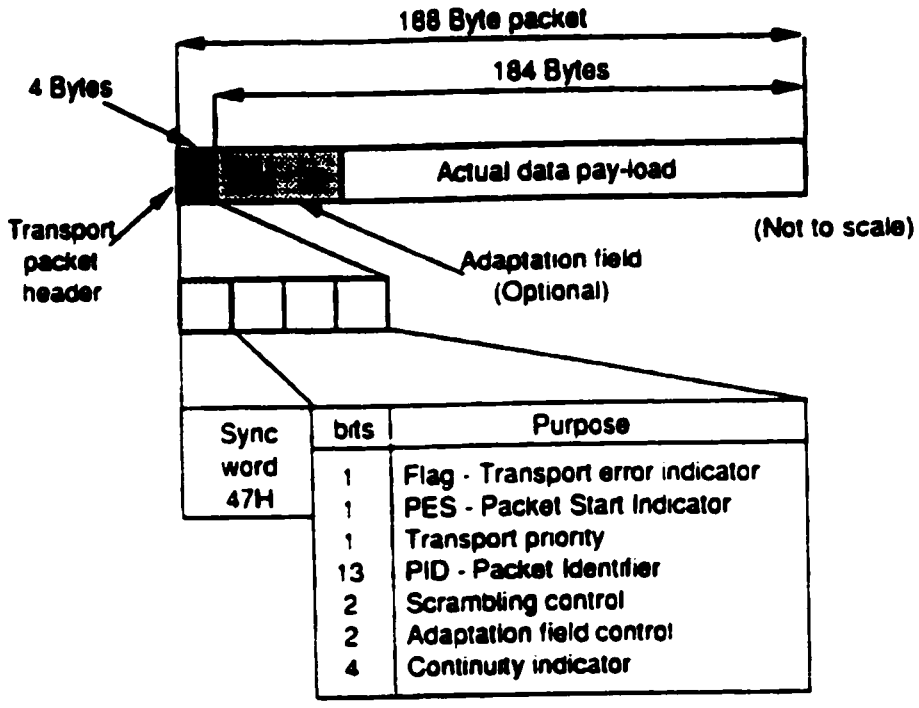


Fig 2.7 The MPEG-2 188 Byte Packet [Rob98]

For program synchronization, there is a common clock reference called "Program Clock reference (PCR)". This signal is carried every 40 ms of time at the most to synchronize the whole program. The individual elementary streams contain time stamps such as Decoding Time Stamps (DTS), and Presentation Time Stamps (PTS).

For service, there are tables like Program Specific Information (PSI), Service Information (SI). PSI contains partly PAT, PMT, CAT (Conditional Access Table) and NIT (Network Information Table). The CAT gives information for encrypted programs, whereas NIT contains data given by the network operator for tuning the receivers (e.g. orbit positions or transponder numbers).

BAT (Bouquet Association Table) contains information about programs from a supplier irrespective of the propagation paths of the programs. SDT (Service Description Table) describes the programs offered.

EIT (Event Information Table) supplies the database for an electronic TV guide with info about the type of program and age classification of the viewer.

RST (Running Status Table) comprises status information about the individual programs and especially serves for controlling video recorders.

TDT (Time and Data Tables) provides information about the date and current time. TOT (Time Offset Table) provides information about the local time offset in addition to date and time.

SI (Stuffing Table) has no relevant content. It's generated when invalid tables are overwritten during transmission.

The MPEG transport stream is an extremely complex structure, with linked and interlinked tables and identifiers to separate elementary streams, vectors, coefficients and quantization tables.

Failures could be in within the encoder, decoder or due to the transmission medium. Modulation errors, RF effects in noise, multipath fading among others could contribute to the reception of a defective stream without the right information to decode it. Fig 2.8 shows a typical transport stream and packets.

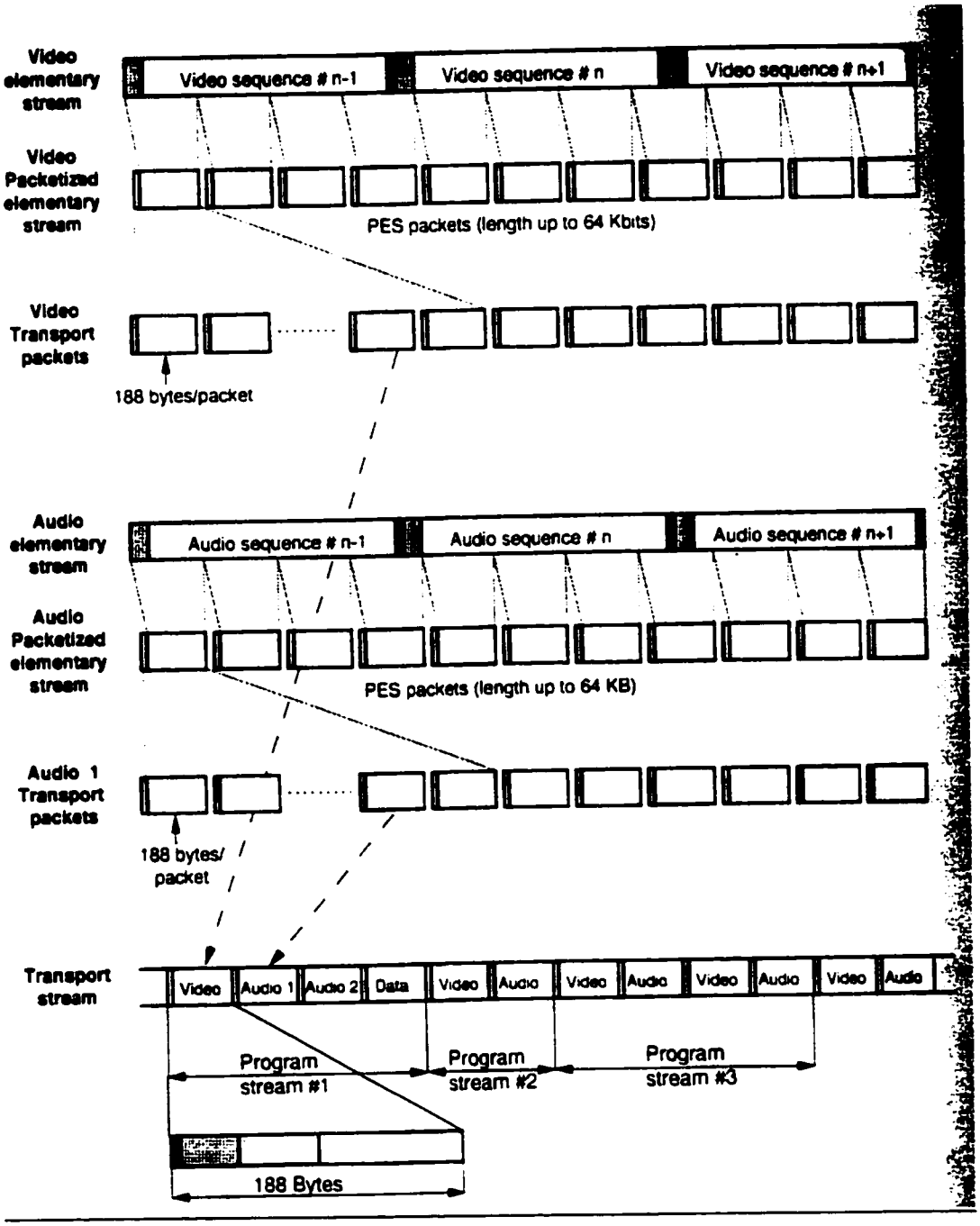


Fig 2.8 The MPEG-2 Transport Stream [Rob98]

Chapter 3

Digital TV and Video Standards

3.1 History of TV

The development of TV started with the German Paul Gottlieb Nipkow invented his mechanical television system in 1884 [Bru99]. The Americans and Europeans worked on that until electronic TV was created. There were different line systems introduced like 405 line system in Great Britain in 1935 developed by EMI under Sir Isaac Shoenberg, and adopted by the BBC in 1836. A 441 interlaced lines, 50Hz introduced in Germany in 1937. Then later in France emerged an 819 line interlaced system. The USA followed in 1941 with the 525 lines 60 Hz system, and then Japan followed in 1953 with a similar system.

As for color television, there was a German patent for such a system dating back to 1904 [Bru99]. Baird demonstrated the first color mechanical TV system in 1928 with a Nipkow disk with 3 spirals, one for each primary color (Red, Green & Blue). The first color TV broadcast started in the USA in 1951 using "Abortive frame sequential system". Then in 1953 the National Television systems Committee (NTSC) developed their system, and it started broadcasting in 1954 in the USA followed by Japan in 1960.

The NTSC system was sensitive to some distortions during transmission causing color hues. Therefore its' acronym sometimes means "Never The Same Color". In 1957, Henri de France developed the "Systeme Electronique Coloeur avec Memoire" (SECAM), which tackled the hue error problem. Later in 1961 W. Bruch in Germany developed the "Phase Alternating Line" system (PAL).

The term High Definition has been an old one in the industry. Baird called his mechanical 30-line system a High Definition one. Nowadays, we refer to HDTV as 1000 line systems. There is other work on higher resolution systems like ultra HDTV in Japan with 3000 lines.

The real work on HDTV started in the sixties in Japan when Dr. Takashi Fuijo from NHK started development on a high quality system that is compatible with 35 mm quality and CD quality audio. In the second half of 1970s the first broadcast started with an 1125 lines 60 Hz system. NHK also developed a Multiple Sub-Nyquist Sampling Encoding (MUSE) for satellite service in 1984 with 1125 lines 60 Hz system.

In 1985, the European "Comite' Consultatif International des Radiocommunications CCIR Interim working party (IWP) made a proposal for a 1125 line 60 Hz system similar to the Japanese system. This was not successful. This is when the USA and Europe went into different paths in the HDTV development.

In Europe they developed a system called High Definition Multiplexed Analog component (HDMAC) for satellite transmission using a 1250 lines and 50 Hz system.

In 1987 in the USA, the FCC asked the industry for proposals for a HDTV terrestrial system and received 21 proposals. However, only 4 of those were compatible with the NTSC system. In 1992 the 4 parties involved:

1- General Instruments (later bought by Motorola),

2- AT&T / Zenith

3- David Sarnoff Research Center/ Philips/ Thomson

4- MIT (Massachusetts Institute of Technology)

Formed what is known as the Grand Alliance, with the mandate of developing one HDTV standard for terrestrial TV in the USA. They chose MPEG2 for source coding, Dolby AC-3 for audio coding. The GA was later transformed into the Advanced Television Standards Committee ATSC. The HDTV system they developed in called ATSC for terrestrial broadcast.

In 1993, the European Launching Group with 84 European broadcasters, telecommunications organizations, manufacturers and regulatory authorities signed a memorandum forming the Digital Video Broadcast project (DVB). The DVB system adopted MPEG-2 as well for source

coding for both audio and video. There are DVB standards now for terrestrial, cable as well as satellite broadcast.

Thus there are 2 main standards in HDTV terrestrial systems worldwide:

1- ATSC

2- DVB-T

3.2 Advanced Television Standards Committee (ATSC):

The Advanced Television Systems Committee (ATSC) was formed in the USA in 1982 by the member organizations of:

- The Joint Committee on Inter Society Coordination (JCIC)
- The Electronics Industries Association (EIA)
- The Institute of Electrical and Electronics Engineers (IEEE)
- The National Association of Broadcasters (NAB)
- The National Cable Television Association (NCTA)
- The Society of Motion Picture and Television Engineers (SMPTE)

The ATSC currently has over 200 members. The ATSC standards include standard definition TV, high definition TV, data broadcasting, multi-channel surround sound audio and satellite direct to home broadcasting.

The Federal Communications Commission (FCC) in the USA adopted the ATSC digital TV standard (document A/53) on Dec. 24th/1996. Canada adopted that on Nov. 8th, 1997. South Korea on Nov. 21st, 1997, Taiwan on May 8th, 1998 and Argentina on Oct. 22nd, 1998.

The ATSC standard document A53 was issued in Sept./95 and later revised in April/2001 [www.atsc.org]. It defines the structure of the ATSC transport stream, which is 19.39 Mbits/sec, and based on MPEG-2 source coding standard as specified in ISO/IEC 13818-1 standard.

For terrestrial transmission, the modulation technique used is 8VSB (Vestigial side band with 8 discrete values). The standard also allows for 16VSB, although nothing much has happened in

that direction. For picture formats, the ATSC standard allows for 36 picture formats (including 18 NTSC-friendly formats) . Probably the most 2 popular formats adopted by broadcasters in the USA are the 720 line progressive and 1080 line interlaced formats.

ATSC Picture Formats				
Aspect Ratio	Active H-Pixels	Active Lines	Scanning Mode	Frame Rate
4:3	640	480	Progressive	60(59.94), 30(29.97), 24(23.98)
			Interlaced	30(29.97)
4:3	704	480	Progressive	60(59.94), 30(29.97), 24(23.98)
			Interlaced	30(29.97)
16:9	704	480	Progressive	60(59.94), 30(29.97), 24(23.98)
			Interlaced	30(29.97)
16:9	1280	720	Progressive	60(59.94), 30(29.97), 24(23.98)
16:9	1920	1080	Progressive	30(29.97), 24(23.98)
			Interlaced	30(29.97)

Table 3.1 Popular ATSC picture formats

The FCC in the USA has issued a ruling in 1996 that all broadcasting stations in the USA have to broadcast at least one ATSC station by the year 2006, or they may lose the spectrum they already have. At the moment there are more than 200 stations in the USA broadcasting ATSC terrestrial signal. [www.atsc.org], and are increasing every day. The total number of off the air stations is about 1500 stations in the USA.

3.3 Digital Video Broadcast (DVB):

The Digital Video Broadcasting Project (DVB) is a consortium of over 300 broadcasters, manufacturers, network operators and regulatory bodies worldwide, committed to designing a global standard for the delivery of digital television. Numerous broadcast services using DVB standards are now operational in Europe, North and South America, Africa, Asia, and Australia.

The DVB family of standards include:

DVB-S A satellite system that can be used with any transponder, current or planned (Uses QPSK modulation).

DVB-C	A matching cable system to suit the characteristics of all cable networks (Uses QAM-64 modulation).
DVB-T	A digital terrestrial system (Using COFDM modulation).
DVB-MC/S	A microwave multipoint video distribution systems.
DVB-SI	A service information system, enabling the user to navigate through the DVB environment;
DVB-CA	A common scrambling or Conditional Access system.
DVB-CI	A common interface for conditional access and other uses.

DVB-T which uses the multi-carrier Coded Orthogonal Frequency Division Multiplexing (COFDM) modulation technique, is capable of delivering crystal clear picture to television connected to portable, set-top antennas in hostile reception environments such as city apartments, or even to receivers on the move. In trials in Germany since 1997, DVB-T has been tested in slow moving city trams and on the autobahn at speeds in excess of 275 km/hr.

The vision of DVB is to bring many types of digital video broadcast services to various devices in the home of the future.

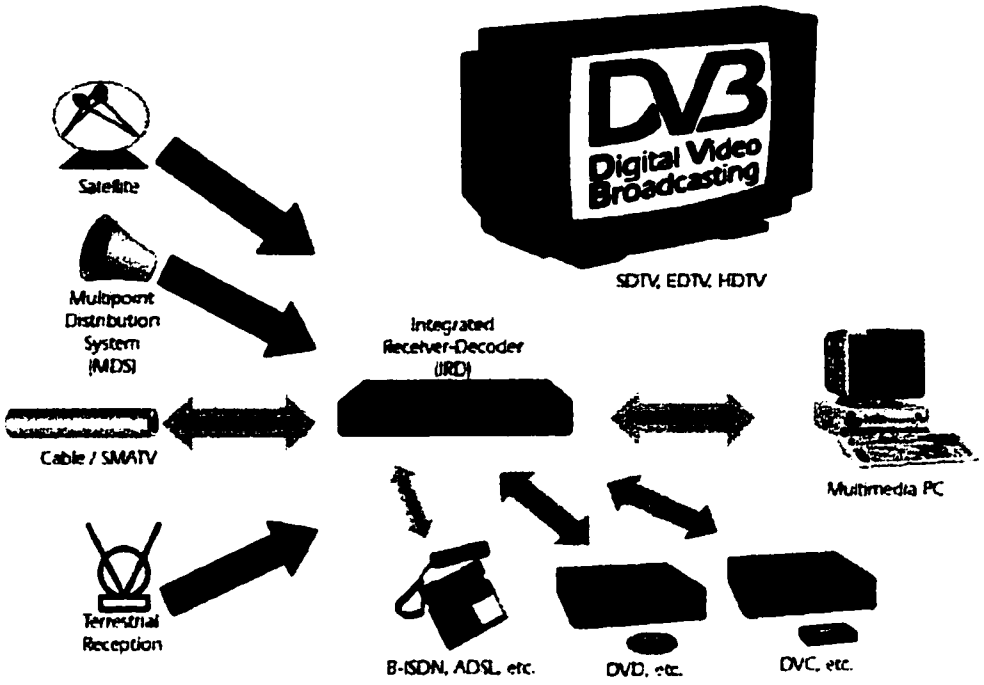


Fig 3.1 Multimedia Interfaces in the Home using DVB

DVB has defined the delivery media (for satellite, cable, terrestrial, etc.), and the interfaces (to ISDN, PSTN, ATM, PDH, SDH... etc.), which enable a cluster of interconnected devices in the home, all receiving and processing digital broadcast services. It all comes together on the set top box (STB) also known as the integrated receiver-decoder (IRD). MPEG-2 is used as a common sound and vision-coding system. In fig. 3.2, four MPEG-2 "data containers" are shown:

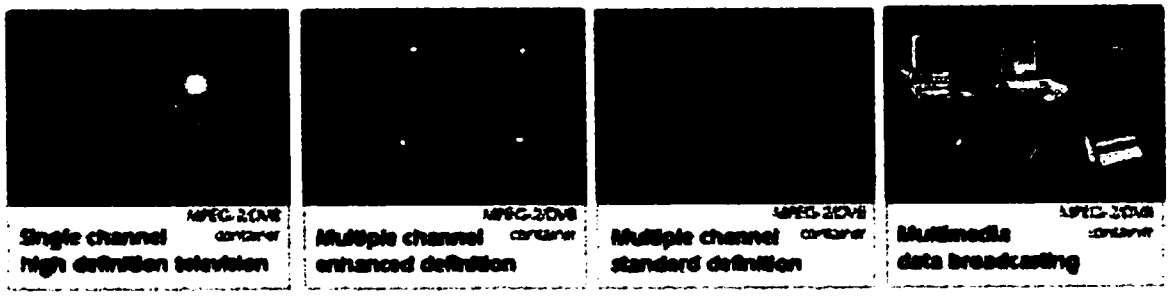


Fig 3.2 The DVB transport Stream

The DVB approach provides great flexibility in terms of transmitted digital information, owing to its data "container" concept.

DVB simply delivers to the receiver "containers" with compressed image, sound or data. No restrictions exist as to the kind of information, which can be stored in these containers. The DVB Service Information acts like a header to the MPEG-container, ensuring that the receiver knows what it needs to decode.

DVB selected the "MPEG-2" approach developed by the Moving Picture Experts Group (MPEG) for the compression of image and sound data prior to transmission. MPEG-2/DVB compliant systems are capable of delivering anything from multiple-channel Standard Definition (SDTV) or Enhanced Definition (EDTV) to single channel HDTV in the same transport stream. Or can also deliver a SDTV signal along with other audio or data services including interactive services. Delivering this service via many mediums, this will open the door for the Convergence era everybody is talking about.

The sound coding system specified for all DVB systems is the MPEG Layer II audio standard (MUSICAM). MPEG Layer II is a digital compression system, which takes advantage of the fact that a sound element will have a masking effect on other nearby lower-level sounds (or on noise). This is used to facilitate the coding of the audio with low data rates. This system can achieve a sound quality, which is very close to Compact Disc. The system can be used for mono, stereo, or multilingual sound, and as well for surround sound.

A DVB "Blue Book" called "Implementation Guidelines for the use of MPEG-2 Systems, Video and Audio in Satellite and Cable Broadcasting Applications in Europe" has been provided which describes in detail the subset of MPEG-2 elements to be used by DVB.

The baseline SDTV decoder is based on MPEG-2 Main Profile at Main Level (MP @ ML). The HDTV baseline decoder uses MPEG-2 Main Profile at High Level (MP @ HL), ensuring backwards compatibility with existing DVB/MPEG-2 bit streams.

While the DVB MPEG-2 Data Container gives a flexible range of service options, the Implementation Guidelines recommend HDTV broadcasters use the so-called Common Image Format (CIF) proposed by the ITU and DAVIC (i.e. 1080 lines by 1920 pixels). They also detail

the implementation requirements necessary for Integrated Receiver-Decoders (IRDs) to be used.

When HDTV broadcasts are directed towards populations of DVB-compliant IRDs of which some are not HDTV-enabled, the DVB Implementation Guidelines recommend multiplexing SDTV program streams into the MPEG-2 bit stream alongside the HDTV programs, in an approach known as Simul-casting.

3.3.1 Coded Orthogonal Frequency Division Multiplexing (COFDM)

COFDM is an un-orthodox, rather sophisticated modulation scheme that has been considered on-and-off for terrestrial broadcasting for over 6 years. It was originally considered as the modulation approach for one of the proposed digital ATV systems but was abandoned for another method because of the time it would have taken to design and build a world-class version for testing. Later it became of personal interest because of the opportunity it offered to support distributed transmission.

Development of COFDM has been largely in Europe, with some work coming later in Japan. In Europe, it was very close to adoption by the Digital Video Broadcasting (DVB) group as the standard method for terrestrial digital transmission. A late-in-the-game effort was mounted by some North American Broadcasters to gain consideration of COFDM as a replacement for the Grand Alliance VSB scheme, but it was too little, too late to succeed.

Coded Orthogonal Frequency Division Multiplexing (COFDM) has been specified for digital broadcasting systems for both audio -- Digital Audio Broadcasting (DAB) and terrestrial television -- Digital Video Broadcasting (DVB-T).

COFDM is particularly well matched to these applications, since it is very tolerant of the effects of multi-path (provided a suitable guard interval is used). It is not limited to 'natural' multi-path as it can also be used in Single-Frequency Networks (SFNs) in which all transmitters radiate the same signal on the same frequency.

A receiver may thus receive signals from several transmitters, normally with different delays and thus forming a kind of 'unnatural' additional multi-path. Provided the range of delays of the multi-path (natural or 'unnatural') does not exceed the designed tolerance of the system (slightly greater than the guard interval) all the received-signal components contribute usefully.
Multi-path can be viewed in the frequency domain as a frequency selective channel response.
Another frequency-dependent effect for which COFDM offers real benefit is the presence of isolated narrow-band interfering signals within the signal bandwidth. Note that conventional analogue television signals (NTSC/PAL/ SECAM) essentially behave like narrow-band interferers to COFDM.
COFDM copes with both these frequency-dependent effects as a result of the use of forward error coding. However, rather more is involved than simply adding coding -- the 'C' -- to an un-coded OFDM system. The coding and decoding is integrated in a way, which is specially tailored to frequency-dependent channels and brings much better performance.

3.3.2 UNCODED OFDM

OFDM spreads the data to be transmitted over a large number of carriers -- typically more than a thousand (1705 carriers in case of 2K systems, and 6814 carriers for 8K systems. The data rate to be conveyed by each of these carriers is correspondingly reduced. It follows that the symbol length is in turn extended. These modulation symbols on each of the carriers are arranged to occur simultaneously.
The carriers have common frequency spacing. This is the inverse of the duration, called the active symbol period, over which the receiver will examine the signal, performing the equivalent of an 'integrate-and-dump' demodulation. This choice of carrier spacing ensures orthogonality (the 'O' of OFDM) of the carriers -- the demodulator for one carrier does not 'see' the modulation of the others, so there is no cross talk between carriers, even though there is no explicit filtering and their spectra overlap.

A further refinement adds the concept of a guard interval. Each modulation symbol is transmitted for a total symbol period, which is longer than the active symbol period by a period called the guard interval. This means that the receiver will experience neither inter-symbol nor inter-carrier interference provided that any echoes present in the signal have a delay, which does not exceed the guard interval.

The addition of the guard interval reduces the data capacity by an amount dependent on its length. The concept of a guard interval could in principle be applied to a single-carrier system, but the loss of data capacity would normally be prohibitive. With OFDM it is possible to protect against echoes with prolonged delay, simply by choosing a sufficient number of carriers that the guard interval need not form too great a fraction of the active symbol period. Both DAB and DVB-T have a guard interval which is no greater than 1/4 of the active symbol period, but can protect against echo delays of the order of 200 μs (depending on the mode chosen).

Fortunately the apparently very complex processes of modulating (and demodulating) thousands of carriers simultaneously are equivalent to Discrete Fourier Transform operations, for which efficient Fast Fourier Transform (FFT) algorithms exist. Thus integrated circuit implementations of OFDM demodulators are feasible for affordable mass-produced receivers.

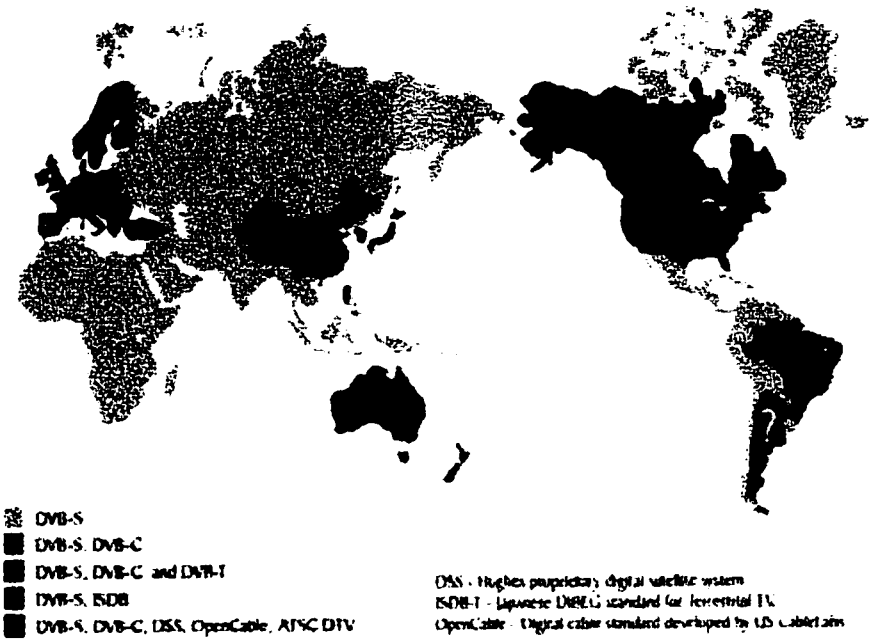


Fig 3.3 DVB implementation Worldwide

As seen from this map, the DVB standard is dominant in the World, except in North America and a few other countries for terrestrial transmission. In Japan they adopted a variation of the DVB called ISDB.

ISDB is the Japanese Digital Broadcasting Experts Group (DiBEG's) variant of the DVB-T system. It differs in one key respect only, the use of an intermediate (software driven) data segmentation system, whereby services such as radio, SDTV, HDTV and Mobile TV can be flexibly allocated pieces of the overall service bandwidth. The Japanese ISDB system is anticipated to launch in Japan no earlier than 2003.

At the moment DVB-T broadcast has already started in the UK, Sweden, Spain and Australia. Many European countries and New Zealand will follow very soon.

3.4 Digital Video Tape Standards:

There are a number of digital tape standards in the market nowadays. And of course there is the concept of DDR or digital disk recorders. These recorders record on a hard drive medium, and

can be without compression. However, many TV stations use one or more of the other tape standards for their materials. The main digital tape standards are:

D1 (introduced 1986)

A format for component digital video tape recording working to the ITU-R 601, 4:2:2 standard using 8-bit sampling. There is no compression in this format, and thus it's one of the best quality systems, however it's expensive. The tape is 19 mm wide and allows up to 94 minutes to be recorded on a cassette. Being a component recording system it is ideal for studio or post-production work with its high chrominance bandwidth allowing excellent chroma keying. Also multiple generations are possible with very little degradation and D1 equipment can integrate without transcoding to most digital effects systems, telecines, graphics devices, disk recorders, etc. Being component there are no color framing requirements. Despite the advantages, D1 equipment is not extensively used in general areas of TV production, at least partly due to its high cost. (Often used incorrectly to indicate component digital video.)

D2 (Introduced 1988)

The VTR standard for digital composite (coded) NTSC or PAL signals that uses data conforming to SMPTE 244M. It uses 19 mm tape and records up to 208 minutes on a single cassette. Neither cassettes nor recording formats are compatible with D1. D2 has often been used as a direct replacement for 1-inch analog VTRs. Although offering good stunt modes and multiple generations with low losses, being a coded system means coded characteristics are present. The user must be aware of cross-color, transcoding footprints, low chrominance bandwidths and color framing sequences. Employing an 8-bit format to sample the whole coded signal results in reduced amplitude resolution making D2 more susceptible to contouring artifacts. (Often used incorrectly to indicate composite digital video.)

D3 (1990)

A composite digital video recording format that uses data conforming to SMPTE 244M. Uses 1/2-inch tape cassettes for recording digitized composite (coded) PAL or NTSC signals sampled at 8 bits. Cassettes are available for 50 to 245 minutes. Since this uses a composite signal the characteristics are generally as for D2 except that the 1/2-inch cassette size has allowed a full family of VTR equipment to be realized in one format, including a camcorder.

D4

A format designation never utilized due to the fact that the number four is considered unlucky (being synonymous with death in some Asian languages).

D5 (1994)

A VTR format using the same cassette as D3 but recording component signals conforming to the ITU-R BT.601-2 (CCIR 601) recommendations at 10-bit resolution. With internal decoding D5 VTRs can play back D3 tapes and provide component outputs. Being a non-compressed component digital video recorder means D5 enjoys all the performance benefits of D1, making it suitable for high-end postproduction as well as more general studio use. Besides servicing the current 625 and 525 line TV standards the format also has provision for HDTV recording by use of about 4:1 compression (HD D5).

D6 (1996)

A digital tape format which uses a 19mm helical-scan cassette tape to record uncompressed high definition television material at 1.88 GBps (1.2 Gbps). D6 is currently the only high definition recording format defined by a recognized standard. D6 accepts both the European 1250/50 interlaced format and the Japanese 260M version of the 1125/60 interlaced format, which uses 1035 active lines. It does not accept the ITU format of 1080 active lines. ANSI/SMPTE 277M and 278M are D6 standards.

D7 (1996)

DVCPRO. Panasonic's development of native DV component format, which records an 18-micron (18x10-6m, eighteen thousandths of a millimeter) track on 6.35 mm (0.25-inch) metal particle tape. DVCPRO uses native DCT-based DV compression at 5:1 from a 4:1:1 8-bit sampled source. It uses 10 tracks per frame for 525/60 sources and 12 tracks per frame for 625/50 sources, both use 4:1:1 sampling. Tape speed is 33.813mm/s. It includes two 16-bit digital audio channels sampled at 48 kHz and an analog cue track. Both Linear (LTC) and Vertical Interval Time Code (VITC) are supported. There is a 4:2:2 (DVCPRO50) and progressive scan 4:2:0 (DVCPRO P) version of the format, as well as a high definition version (DVCPROHD).

D8

There is no D8. The Television Recording and Reproduction Technology Committee of SMPTE decided to skip D8 because of the possibility of confusion with similarly named digital audio or data recorders (DA-88).

D9 (Formerly Digital-S) (Introduced 1996)

A 1/2-inch digital tape format developed by JVC, which uses a high-density metal particle tape running at 57.8mm/s to record a video data rate of 50 Mbps. The tape can be shuttled and search up to 32x speed. Video sampled at 4:2:2 is compressed at 3.3:1 using DCT-based intra-frame compression (DV). Two or four audio channels are recorded at 16-bit, 48 kHz sampling; each is individually editable. The format also includes two cue tracks. Some machines can play back analog S-VHS. D9 HD is the high definition version recording at 100 Mbps.

D9 HD

A high definition digital component format based on D9. Records on 1/2-inch tape with 100 Mbps video.

DV

This digital VCR format is a cooperation between Hitachi, JVC, Sony, Matsushita, Mitsubishi, Philips, Sanyo, Sharp, Thomson and Toshiba. It uses 6.35 mm (0.25-inch) wide tape in a range of products to record 525/60 or 625/50 video for the consumer (DV) and professional markets (Panasonic's DVCPRO, Sony's DVCAM and Digital-8). All models use digital intra-field DCT-based "DV" compression (about 5:1) to record 8-bit component digital video based on 13.5 MHz luminance sampling. The consumer versions, DVCAM, and Digital-8 sample video at 4:1:1 (525/60) or 4:2:0 (625/50) video (DVCPRO is 4:1:1 in both 525/60 and 625/25) and provide two 16-bit/48 or 44.1 kHz, or four 12-bit/32 kHz audio channels onto a 4 hour 30 minutes standard cassette or smaller 1 hour "mini" cassette. The video recording rate is 25 Mbps.

Digital Betacam (1993)

Introduced by Sony. It has a compression of 2:1.

Betacam SX (1993)

Introduced by Sony. It has a compression of 10:1.

Chapter 4

Basics of Video Quality

4.1 The Human Visual System (HVS)

The eye perceives images with a very unique perception. Fig. 4.1 shows the eye structure.

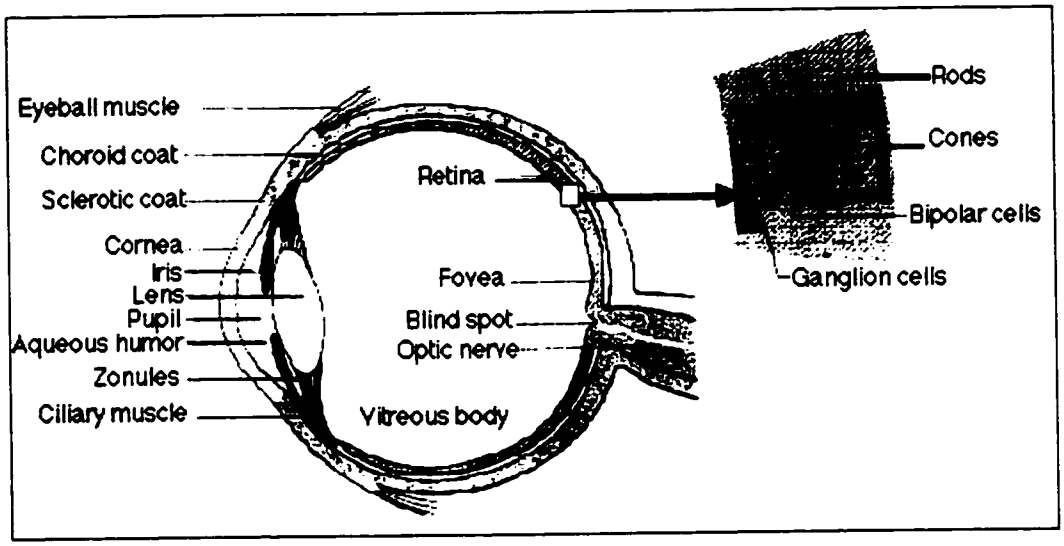


Fig 4.1 Anatomy of the Human Eye

The light bounces off different objects, and then is refracted by the cornea toward the pupil. This is the opening of the iris where the light enters the eye. The light is then refracted by the lens onto the back of the eyeball, forming an image on the retina. The retina consists of receptors sensitive to light called photoreceptors, and those are connected by nerve cells. The photoreceptors contain chemical pigments, which absorb light and create a neural response. "The light absorbed by photoreceptors initiates a chemical reaction that bleaches the pigment, which reduces the sensitivity to light in proportion to the amount of bleached pigment" [Has97]. The amount of bleached pigment corresponds to the amount of light.

There are 2 types of photoreceptors:

1- Rods, which are responsible for low light vision, like night (scotopic) colorless vision.

2- Cones are responsible for chrominance (color) and detail under normal light conditions. There are 3 types of cones that are individually sensitive to Red, green and blue wavelength lights, and the combination of those create all the other different colors.

The light of various wavelengths causes the sensation of colors, or also called hue. In the retina, there are roughly between 110-130 million rods, and about 6 and 7 million cones [Rob98]. That's one of the reasons the human eye is more sensitive to light or luminance than chrominance or color. Fig 4.2 shows the rods and cones cell structure. [http://www.ultranet.com/~jkimball/BiologyPages/V/Vision.html]

Figure 2

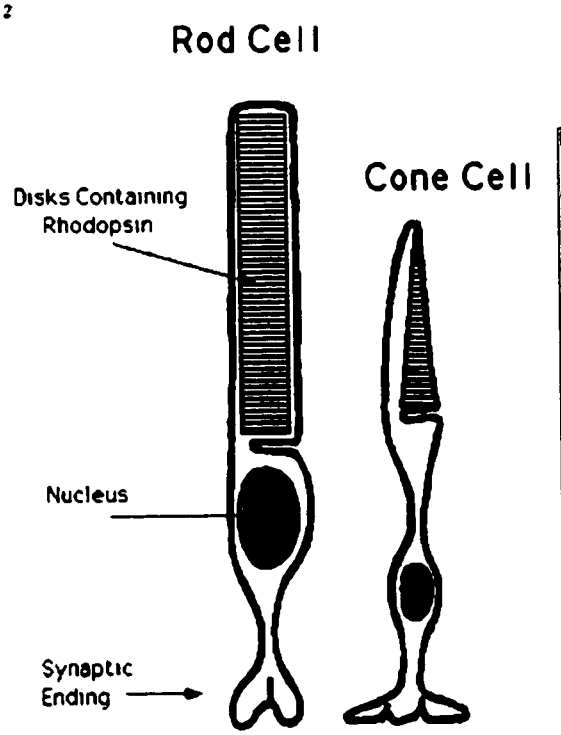


Fig. 4.2 a The Rod and Cone cells

Fig. 4.2.b. Rods and Cones

The information from the retina is transformed through the optical nerve (which has about 800,000 fibers) to the brain. Beyond the retina, the visual information takes place before reaching the brain in places called "The lateral geniculate", and "visual cortex" [Rob98]. A "ganglion cell" feeds each fiber in the optical nerve. Each ganglion cell is connected to hundreds of rod cells, and tens of cone cells. Fig. 4.3 shows the eye-brain connection.

The cones are in the middle of the retina in an area called (Fovea). At high light intensity, the cones have a high colorless visual acuity, and degraded color acuity. When light intensity reduces, the perception is shifted more towards areas with rods.

In the fovea area, every cone has a direct connection to a dedicated ganglion cell, and as well as sharing other ganglion cells with groups of cones and rods. This accounts for the high acuity of vision in the center of the visual field. This acuity degrades as the light intensity decreases [Rob98]. The eye perceives color as the various wavelengths or hues. The RGB wavelengths are: R: 700nm, G: 546.1 nm, B: 435.8 nm. Fig 4.4 shows the visible light wavelength.

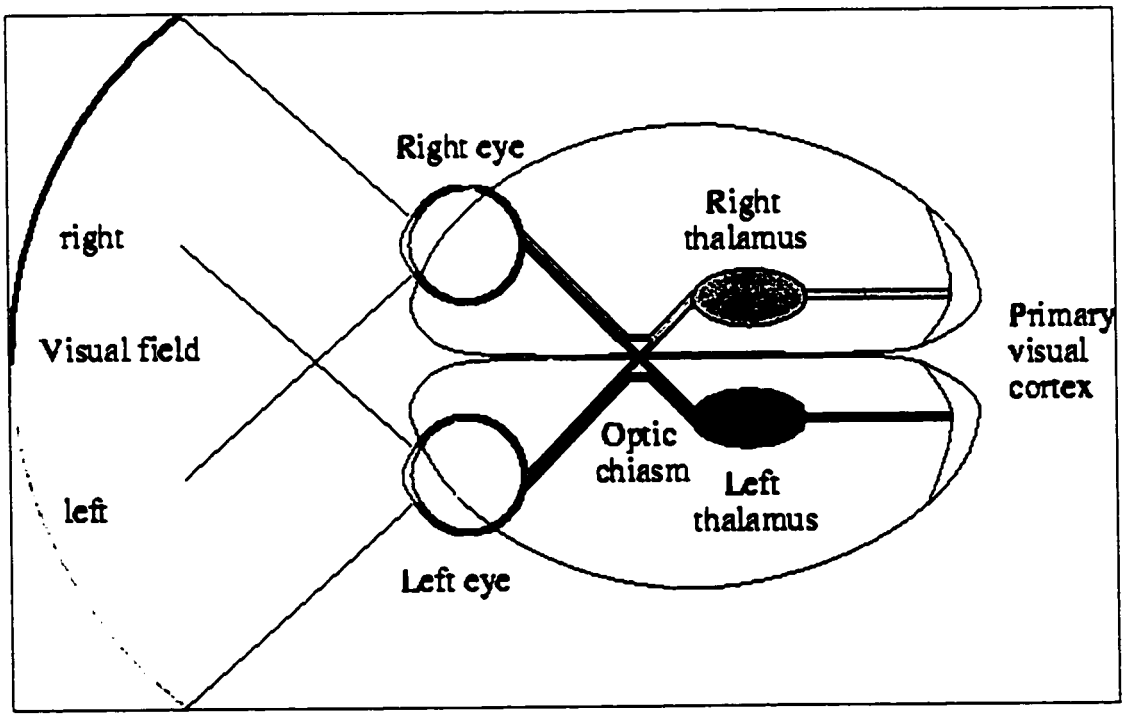


Fig. 4.3 The Human Visual System (HVS)

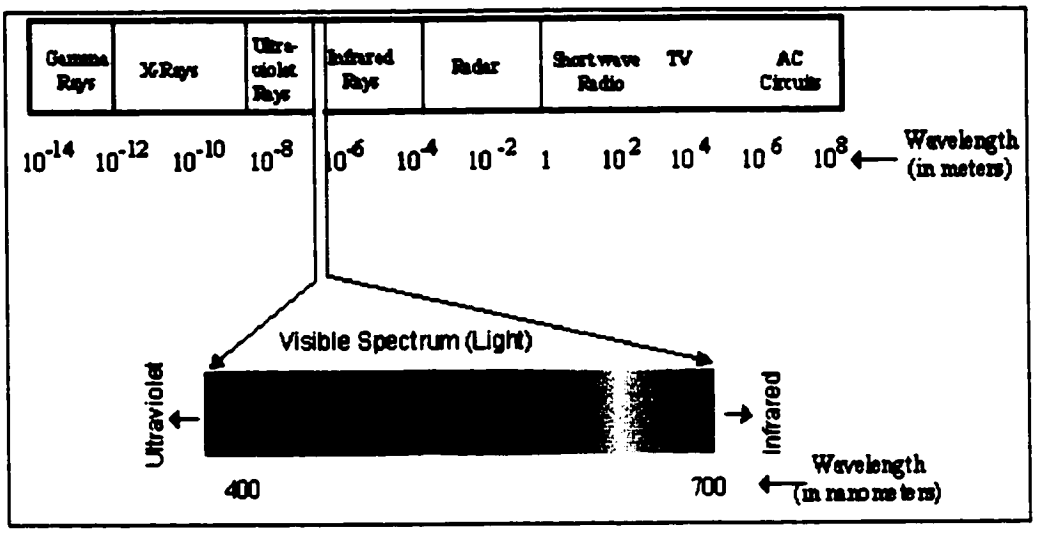


Fig 4.4 The Spectrum of Electromagnetic waves including Visible light

The human eye and the brain's visual cortex make up the Human Vision System (HVS), which can operate over a wide range of light intensities, detect color differences and understand picture contrast as a function of spatial frequency and light intensity.

Since the picture width and viewer distance influence the visibility of pictures, then the image frequency contents is depending on the viewer position. Therefore, it's important for specifying viewing systems to state the viewing distance, and resolution of display system (No. Of lines). The viewing distance for normal TV displays is 6H (where H is the picture height) [Rob98, Has97, Meer97]. This causes the full visual resolution of the details of the picture. However, for high definition systems, this distance is reduced to 3H only, and still maintaining the same level of spatial frequency eye discrimination.

The term "Visual Acuity" means the ability of the human eye to perceive and resolve details. It's measured by "the angle subtended by the smallest visible detail in an object" [Rob98]. In TV systems it's customary to calculate a one-minute of arc as the human acuity. To explain details in a picture, this is called resolution. Television resolution is equal to the number of alternating black and white horizontal lines that can be resolved vertically over the height of the picture.

Assuming normal viewing condition (viewing distance 6H, the vision acuity calculation formula is:

$$N = \frac{1}{\alpha n} \qquad [4.1]$$

Where N = Total number of elements to be resolved in the vertical direction.

Alfa α = Minimum resolvable angle of the eye (in radians)

$N = \frac{D}{H}$ (Viewing distance divided over picture height) [4.2]

If we assume α= 1 minute of arc or 2.91×10^{-4} radians, and n=6, then N becomes approx. 572 lines. This is the origin of why the 2 systems in the world picked numbers of 525 and 625 lines. Another characteristic of the HVS is that it doesn't necessarily perceive equally the digitized samples of a picture [Rob98], thus, the errors in this area do not affect the judgment of perceived quality by a viewer. That's the reason that removing some sample values could be done without affecting the quality of picture from a human perception.

The acuity of the human visual system depends on the following factors (Rob98):

1- The Luminance of the background. The visual acuity increases with luminance, to a maximum level of 100 foot/Lambert or 340 cd/m2 (candles per square meters).

2- The contrast of the picture luminance and chrominance. The image contrast is defined as the difference between the maximum and minimum light intensity to the sum of those two intensities.

$$Contrast = \frac{I_{max.} - I_{min}}{I_{max.} + I_{min.}} \qquad [4.3]$$

Thus picture details are only visible when there is a difference between them and the background. Fig. 4.5 shows the eye contrast sensitivity to spatial frequency in cycles /degree. As well, the eye contrast sensitivity varies with the temporal frequency of the picture (Fig 4.6).

The characteristics of the HVS in relation to spatial and temporal redundancies are summarized as follows [Rob98]; For Spatial Redundancy:

1- The eye is more sensitive to lower spatial frequencies, and less sensitive to higher frequencies. Actually, when the picture is electrically or optically transformed from one stage to another, the different spatial frequency components suffer from different responses. The amplitude of the high frequency components is reduced compared to the low frequency ones.

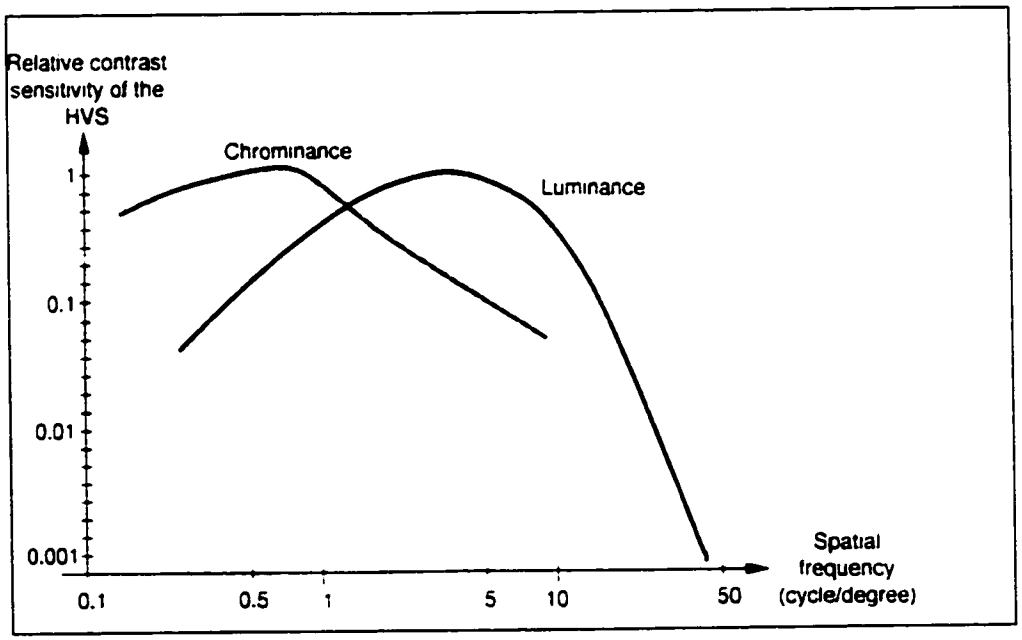


Fig 4.5 Relative contrast Sensitivity of the HVS with spatial frequency [Rob98]

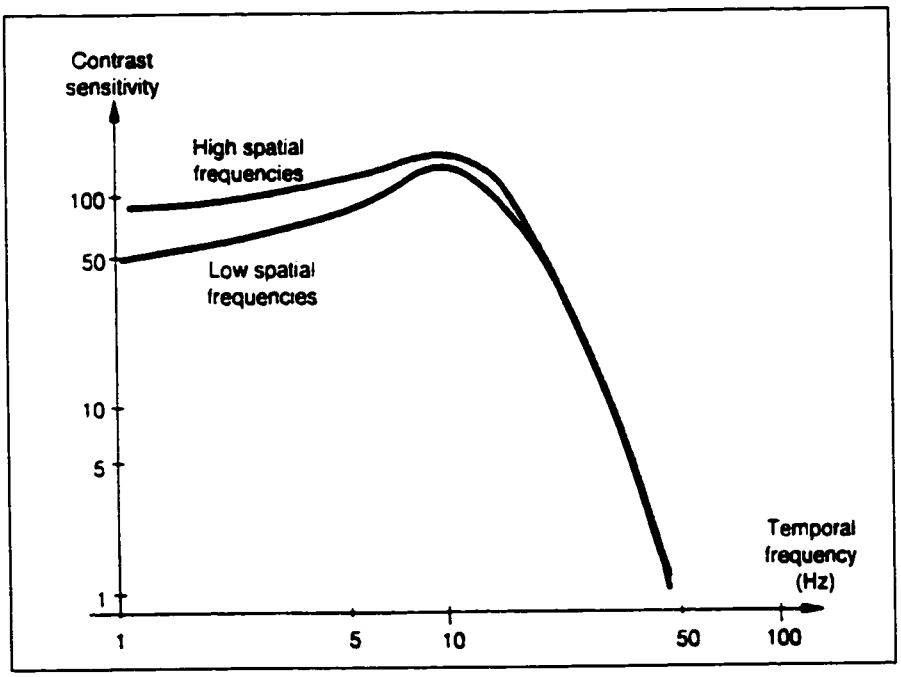


Fig 4.6 Variation of the HVS contrast Sensitivity to temporal frequency [Rob98]

2- Texture masking: The eye may not distinguish errors in textured areas or regions of the picture. And is very sensitive to distortions in uniform areas. The eye can see blocks in a blue-sky area, rather than the same blocks in a sports game with lots of details. In a textured region of the picture, the distortion from a DCT block is spread over many edges, which in turn prevents the detection of distortion a few pixels away from an edge.

3- Edge masking: This means that the contrast sensitivity is smaller in the neighborhood of an edge. Or in other words, errors are harder to see near the edges. This effect depends on the size, shape and duration of a stimuli [Meer97, Vass73]. Van der Meer [Meer97] states that the human visual system treats stimuli of different orientation or frequencies differently. The human visual system contains band-pass filters with a bandwidth of 1 octave, and an orientation of about 30 degrees. The visibility of distortion in a frequency band is independent from the distortion in other bands. [Olz86]. Van der Meer adds that the sensitivity of the human eye, which depends on the picture content, is called masking. The masking effects most often referred to are luminance, edge and temporal masking.

"Although edge masking was first introduced for true edge patterns, it has been shown to exist in each frequency band. It expresses itself as an elevation of the visibility threshold of a frequency orientation band when the contrast in the original band exceeds a certain value" [Daly93]. This is shown in fig 4.7.

4- Luminance masking: Visual threshold increase with background luminance. Van der Meer [Meer97] states that the contrast sensitivity increases with increased luminance. The important value here is the luminance as a function of the Grey background value. Fig. 4.8 shows luminance sensitivity as a function of background Grey value. Assuming a Grey level of 0 to 255 into a screen luminance L, and a correction for the characteristics of the display have been applied, then the relation between L and K can be expressed as:

$$L = 0.00025(K + 15)\ cd/m^2 \qquad [4.4]$$

But of course as different displays have different conversions, there can be different curves for luminance masking for different displays. As well, Van der Meer says "The sensitivity of the Human eye has a peak at mid-Grey levels".

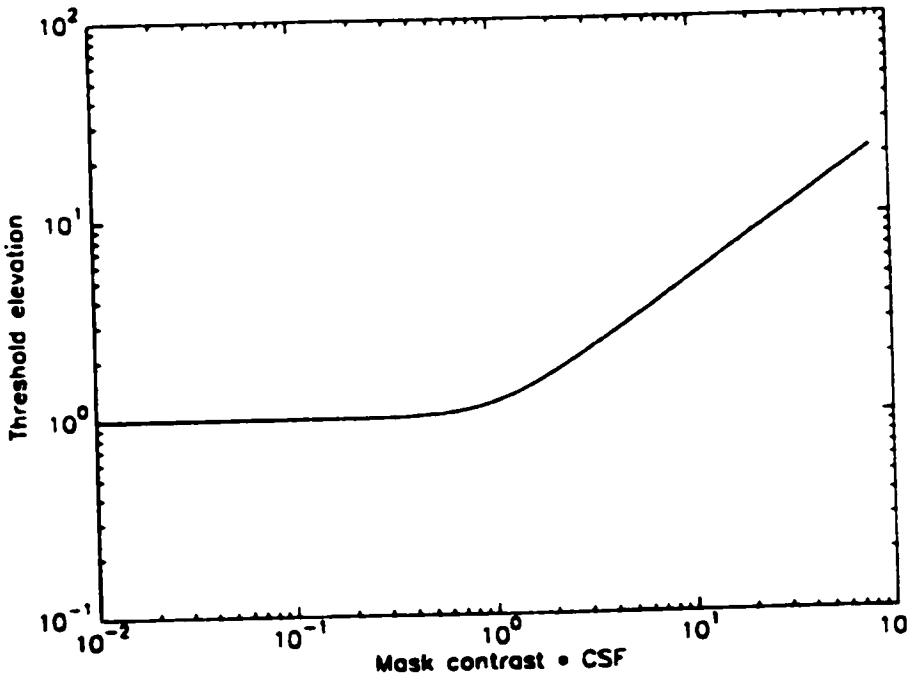


Fig 4.7 Threshold Elevation as a function of local contrast [Daly93]

5- Contrast masking: Where errors and noise in light regions of the picture are more difficult to detect. This refers to the reduction of visibility of one picture detail because of the presence of another.

As for Temporal Redundancy, the HVS is characterized by:

1- Temporal frequency sensitivity: Temporal masking results in a drop of sensitivity near temporal discontinuities"[Meer97]. Where flicker could be visible if frequency below 50Hz. Detection threshold of narrow lines increase four folds at temporal discontinuities [Car96].

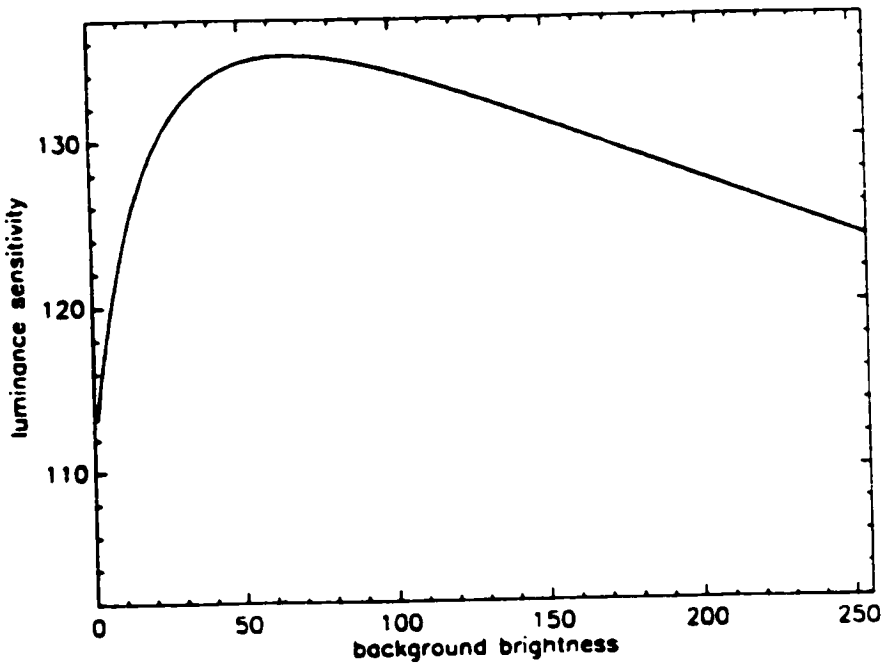


Fig 4.8 Luminance sensitivity as a function of Grey Value [Meer97]

3- Luminance masking: This refers to the fact that high luminance levels increase flicker.

4- Spatial Frequency content: Low spatial frequencies reduce the eye's sensitivity to image flickering.

The response of a system to different frequencies is called "Modulation Transfer Function"(MTF). Therefore, it's desired to characterize any HVS model response in an MTF. But it's not easy to do that. What can be measured is the "Contrast Discrimination", or the ability of the eye to observe differences in contrast. These measurement lead to "Contrast Sensitivity Function (CSF) characterizing the properties of the Human visual system. (See Fig. 4.7)

4.2 Video Artifacts

4.2.1 Analog Signal Artifacts:

Recommendation CCIR567 [Rob99] specifies the testing methodologies for composite analog television signal. The transmission, distribution and processing creates a number of impairments. The different elements making a complete distribution system including Amplifiers, D/A converters, VTRs, cabling, links, ...etc create additive impairments, and degrade the performance of the whole network. The analog impairments are categorized into 3 areas:

1- Linear Distortions

2- Non-Linear Distortions

3- Noise

<table>
<tr><td rowspan="12">Analog Composite Impairments</td><td rowspan="10">Linear Distortion</td><td rowspan="8">Frequency Domain</td><td colspan="2">Frequency Response</td></tr>
<tr><td colspan="2">Group Delay vs. Frequency</td></tr>
<tr><td rowspan="6">Time Domain</td><td rowspan="4">Luminance</td><td>Long Time</td></tr>
<tr><td>Field Time</td></tr>
<tr><td>Line Time</td></tr>
<tr><td>Short Time</td></tr>
<tr><td rowspan="2">Chrominance</td><td>Chrominance to Luminance Delay</td></tr>
<tr><td>Chrominance to Luminance Delay</td></tr>
<tr><td rowspan="6">Non-Linear Distortion</td><td rowspan="2">Luminance</td><td colspan="2">Luminance Non-Linearity</td></tr>
<tr><td colspan="2">Chrominance to Luminance Inter-modulation</td></tr>
<tr><td rowspan="4">Chrominance</td><td rowspan="2">Amplitude</td><td>Gain Non-Linearity</td></tr>
<tr><td>Differential Gain</td></tr>
<tr><td rowspan="2">Phase</td><td>Phase Non-Linearity</td></tr>
<tr><td>Differential Phase</td></tr>
<tr><td rowspan="2">Noise</td><td colspan="3">Random</td></tr>
<tr><td colspan="3">Coherent</td></tr>
</table>

Fig 4.9 Analog Systems artifacts [Rob99]

There are 2 types of testing done on analog systems:

1- Out of Service testing: These are complete full field test signals, similar to tests done during equipment development.

2- In Service testing: Operational testing using test signals within unseen lines, called Vertical Interval test signals (VITs).

These measurements can be done using signal generators, VITs inserters, and analyzing test equipment, waveform monitors & vectroscopes.

4.2.2 Digital Testing Concepts:

Testing of the new digital compressed MPEG system is a totally different challenge than the analog one. The quality of compressed video depends on the distortion introduced during the compression stage.

The test equipment used for analog systems is useless for compressed video systems. Among the MPEG elements affecting the quality of picture are: The data rate (which is one of the most critical elements), The GOP structure (I, B and P frames), field/frame adaptive prediction, motion estimation and compensation methods, slice size and buffer size [Rob98].

The overall compressed video quality is a function of:

1- The Source material: because the perception of materials with high level of details like sports for example is different than plain "talking head" material.

2- Encoder Quality: How well the encoder codes a video stream. This includes the quantization levels and tables, motion and predictive estimation techniques used, as well as buffer size and usage.

3- The Data Rate: The higher the data rate, the more details that can be implemented, and better quality.

The testing of bit reduction systems is divided into 2 parts. The first one is to test the entire MPEG transport stream, and make sure it conforms to the canonical standard stream. To check the existence of synchronization signal, PCR...and different data structures, tables and flags in the stream, and check the integrity of the data after removing all the Reed-Solomon bits, and extracting the data packets.

Then the second level of analysis would be to check the quality of video and audio within that transport stream. Because the MPEG2 transport stream could be perfect, but the DCT, Quantization errors, and the many other errors may have distorted the signal to the extent, it would become annoying to watch.

On another level Van der Meer [Meer97] argues that the base functions of the DCT don't reflect pure frequencies. And for higher spatial frequencies, many DCT coefficients occur in the same band. As well, a DCT coefficient could appear in several different frequency bands of the Human Visual System [Klein92]. The exact visibility threshold for each DCT coefficient can be measured by means of a psychophysical model to measure the smallest value that gives a visible signal. [Pete91, Pete92]

4.2.3 Digital Signal artifacts:

CCIR Report 1089 states the classification of impairments associated with bit- rate reduction systems. Robin [Rob99] classifies the digital artifacts as follows according to CCIR 1089:

Fig 4.11 Original Picture (Susie) (One of the famous MPEG streams)

Fig 4.12 Original Indian Before Compression, with multiple errors from compression

A- Impairments associated with Intra-field coding:

1- Slope Overload: The rise time of the original signal cannot be matched causing blurred edges.

2- Edge Business: The precise continuity of an edge in the original signal cannot be matched and appears noisy.

Fig 4.13 Edge noise effect.

3- Contouring: The uniformity or monotonicity of the original signal cannot be matched causing a layering effect.

4- Granular Noise: Finely detailed portions of the picture (for examples those below threshold level) are not available, and the picture appears noisy.

5- Blocking: The checkered small boxes or underlying block structure sometimes called tiling. This is the most common and visible artifact of DCT transformation.

Fig 4.14 Blocking or Tiling effect due to the DCT

Due to temporal prediction inaccuracy, these artifacts may occur for interfiled and inter-frame coding:

1- Temporal Slope Overload: That happens when edges of fast moving objects cannot be matched and becoming blurred during movement.

2- Granularity and edge business: In movement, where fine detailed areas show granular noise and edge business.

There are other types of impairments caused by temporal sub sampling:

1- Jerkiness: There is a discontinuity in motion for example, where the smoothness of movement cannot be matched by the system.

2- Temporal Aliasing: Where high temporal frequency components are folded back.

Fig 4.15 Aliasing effect

3- Loss of resolution in moving pictures: Where spatial resolution is reduced during movement.

Subjective Evaluation of basic Quality	Intra-field Coding	Slope Overload
		Edge Business
		Contouring
		Granular Noise
		Blockiness
	Inter-field of Inter-frame Prediction Inaccuracy	Temporal Slope Overload
		Edge Business
		Granularity
		Blockiness
	Inter-field of Inter-frame Coding Inaccuracy	Jerkiness
		Temporal Aliasing
		Loss of Resolution in Moving Pictures

Fig 4.10 Bit rate Reduction Impairments according to CCIR1089 [Rob99]

One of the most annoying artifacts in compressed video is called "Mosquito Noise". This is a characteristic of all DCT systems, and appears on sharp edges (such as titles). The edges generate coefficients across the block. The high frequency coefficients are quantized more coarsely than lower frequency components. So, the energy is spread spatially over the block [Meer97, Sym98, Rob98].

Hence, although the block edge can be masked, the distortion from the quantization can be visible a few pixels away from the edge. Another example is around moving objects in front of a homogeneous background.

Back to texture masking, where for a textured region, the distribution of the DCT block is spread over a number of edges, which prevents the detection of distortion a few pixels away from the edges. When the DCT coefficients are not masked by texture, they become visible to the human eye as loss of resolution like un-sharp edges or blur.

And they could also be seen as artifacts like mosquito noise or blocking [Meer97, Sym98].

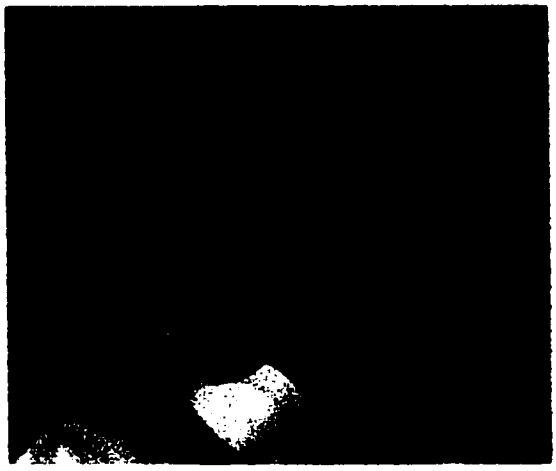


Fig 4.16 Blurring effect

Other artifacts are like:

- Dirty Window [Sym98]: This artifact occurs as streaks or noise that remains stationary, while objects appear to move behind. This looks like viewing

something behind a dirty window. Insufficient bits allocated to the code prediction vectors cause it.

- Wavy Noise [Sym98]: Similar to mosquito noise, this is caused by coarse quantization of high frequency coefficients. This is seen during show pans across a very detailed background scene. As an example, the crowd or spectators at a sporting event, the motion here causes the spreading to vary periodically as the details move across the DCT block.

4.3 The ITU-R.BT-500 standard and Subjective Quality

The first and most reliable technique to evaluate compressed video quality after reconstruction is through means of subjective testing. That is assessing the quality through the evaluation of a sample of viewers.

The CCIR-Rec.500, now known as ITU-R BT-500 sets the standard for all subjective measurements related to compressed video systems. The latest version of that standard is ITU-R BT.500-10 issued on March 2000. This standard describes the different methods and conditions to set up for subjective quality testing. These conditions include:

1- The general viewing conditions like luminance of inactive screen to peak luminance, display brightness and contrast, room illumination, monitor resolution, monitor contrast...etc. There are 2 standards that could be followed, one for lab. environment, and another one for home environment. This creates the potential split of quality standards.

2- The choice of source signals and test material. Particular test material should be used to address particular assessment problems.

3- The choice of observers. ITU-500 states that there should be at least 15 viewers, who are not experts. Prior to the testing, viewers should be screened for normal visual acuity

on the Snellen or Landolt chart [ITU500], and for normal color vision using specially selected charts.

4- The structure of the test session, length of test sequences followed by Grey screen for a number of seconds after each sequence...etc.

However, the most important concept introduced by the ITU-500 standard is the methods used for quality evaluation. They are described as follows:

1- **The Double Stimulus scale (DSIS) or (The EBU method):**

The meaning of double is the evaluation of a reference or clean video with no impairments, and then the same video sequence after being impaired by compression for example. In the DSIS method, the viewers are always shown the clean video source first, then the impaired one. Then they are asked to mark the quality of the second impaired stream, putting in mind the quality of the first over a scale of either 5 level grades or 1-100 scale. In a test session that may last for 30 minutes, the viewers are shown many impaired sequences covering all the required impairment combinations. After the session, the mean scores for each test condition and test picture is calculated.

2- **The Double Stimulus continuous Quality Scale (DSCQS) method**

Another method for evaluating a new system, or the effects of transmission impairments on video quality. This is a bit different than DSIS, in that the viewers are shown pairs of pictures, one is clean, and the other is impaired of the same source material. However, the choice of which comes first is random, so in one time they may see the clean source first, then the impaired one, and in another it could be vice versa. The viewers are asked to evaluate both clean and impaired sequences. The ITU describes a stream sequence of 10 seconds long, with 3 seconds of Grey in between each sequence and the next one.

B- The Single Stimulus (SS) methods:

In this method, a single image or sequence of video is presented, and the viewers give assessment of that one only without any reference stream. SS methods are divided into 3 categories:

B.1 Adjectival categorical judgment methods

In this method, viewers assign a video sequence to one of a set of categories, which are defined in semantic terms. Categories are related to detection of a certain attribute like establish impairment threshold for example. The list of categories is shown in table 4.1.

Five-grade scale			
Quality		Impairments	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

Table 4.1 Quality and Impairment scales [ITU500]

B.2 Numerical categorical judgment methods

This method uses an 11-grade Numerical Categorical Scale (SSNCS). The study on this method is described in report (ITU-R BT.1082).

B.3 Non Categorical Judgment methods:

In this case, the viewers assign a value to each image sequence. There are 2 forms of this method:

B.3.a Continuous scaling: Where viewers assign each image sequence to a point on a line drawn between 2 semantic labels (as in table 4.1).

B.3.b Numerical Scaling: Where viewers assign each video sequence a number reflecting it's value on a specified dimension (image sharpness for example). This scale could be 0-100 for example.

3- Stimulus-comparison methods:

There are 3 stimulus comparison methods used in TV quality evaluation:

3.a Adjectival categorical judgment methods

In this method, viewers assign a relation between members of a pair to one in a set of categories, which are usually defined in semantic terms. The scale is a 7-step scale as shown in table 4.2

-3	Much Worse
-2	Worse
-1	Slightly Worse
0	Same
+1	Slightly Better
+2	Better
+3	Much Better

Table 4.2 Comparison Scale

3.b Non-categorical judgment methods:

Here, viewers assign a value to the relation between the members of the pair. The evaluation could be either in continuous scaling like a point between 2 labels "SAME- SLIGHTLY BETTER". Or it can be expressed in numerical value.

4- Single Stimulus Continuous Quality Evaluation (SSCQE)

In this method, viewers have a slider mechanism unit that is 10 cm, with the 5 grades scale, and a 0-100 numerical scale for example. The viewer would use this scale to continuously change the value depending on the scene, and the unit would capture the readings twice per second. This unit is shown in fig. 4.17

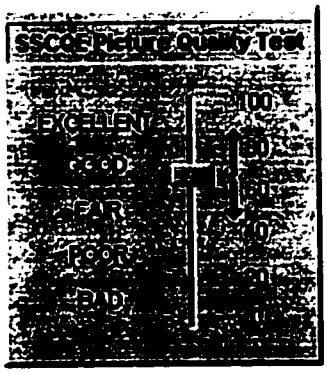


Fig 4.17 The slider mechanism evaluation device

The SSCQE describes conditions for the tests of TV material over longer periods of time that double stimulus systems. The proposed times here are program segments of 5 minutes long at least. The program segments should correspond to specific programming types like sports, news, drama, cartoons...etc. It also suggests a test session length of 30-60 minutes long.

The ITU-500 standard describes in detail the process of data interpretation and analysis to obtain comprehensible information from the data collected from viewers. They are mathematical and statistical techniques calculating the mean average value for each video sequence, then creating sort of confidence level of readings, by eliminating readings within far difference than the other viewer readings.

For each test presentation, the mean score value is calculated, as well as standard deviation and kurtosis coefficient [ITU-500]. From these results further mathematical analysis can be made to create relations between mean scores of streams to certain impairments studied.

4.4 Objective Quality Evaluation:

The objective methods of quality evaluations are made with close correlation to the ITU-500 recommendation in subjective testing. There are 3 agreed upon methods [Rob99]:

4.4.1 Picture Comparison method:

Sometimes this is called double stimulus or Full reference method. This method compares a source video, which is uncompressed (ITU-601 270 Mbits/sec) to the same sequence but impaired from a compression process. After decompression back to a 601 stream, the 2 streams are compared using mathematical processing over each picture. Since the system has the full information of both streams, it could present accurate results.

This system is usually used for MPEG encoders, decoders or Codec evaluations. It's also used in out of service testing of different transmission links like cable, fiber or satellite links. In this case, probably another measuring system must be available at both ends of the link, with the same streams, and synchronizing them to generate valid results. This technique is shown in Fig. 4.18

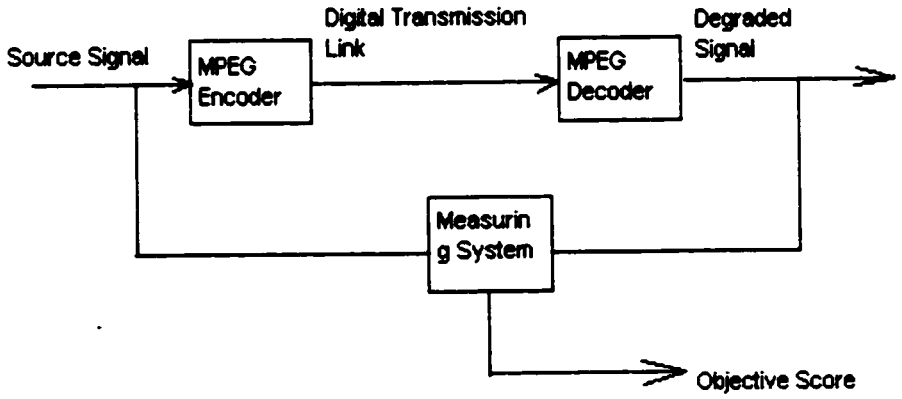


Fig 4.18 Picture Comparison or Full reference System

4.4.2 Feature Extraction Method:

Sometimes called Reduced Reference Method. In this method, the measuring system extracts a reduced amount of features from the degraded signal, and compares that to the same features from the original un-impaired signal. The system then compares the two, and generates an impairment measurement results. This system block diagram is shown in fig. 4.19.

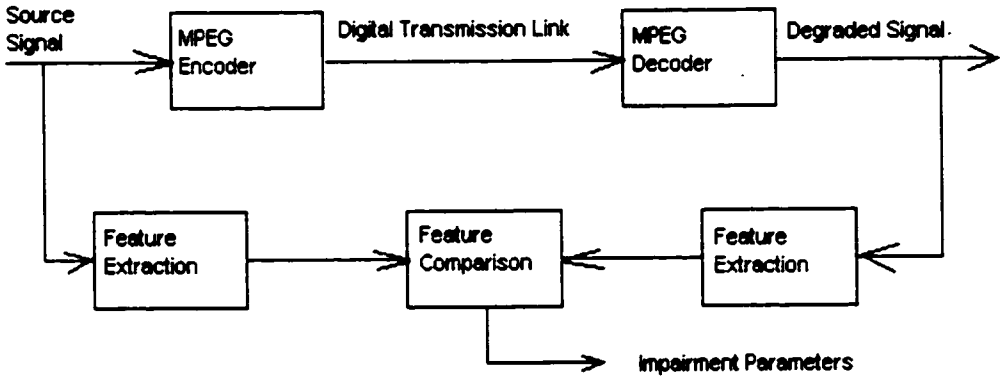


Fig 4.19 Feature Extraction or reduced Reference System

4.4.3 Single ended testing method:

Sometimes called Single Stimulus testing or no reference testing. This system analyzes the received signal based on known impairments or artifacts from MPEG compression, and generates a rating of the quality based on those impairments. This system is shown in Fig. 4.20

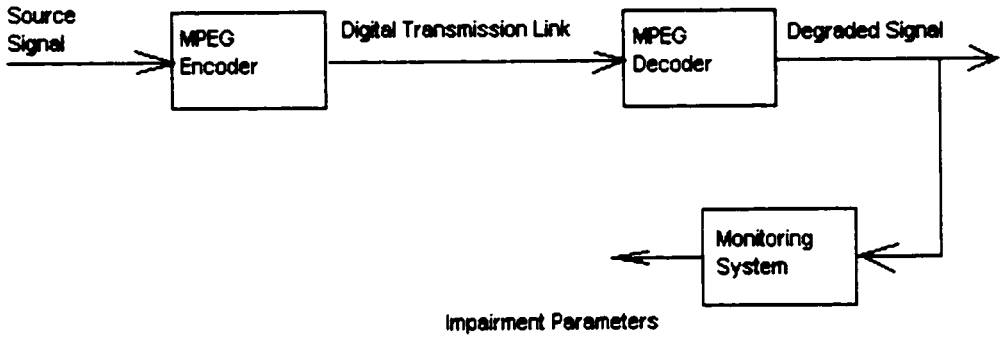


Fig 4.20 Single Stimulus or No Reference System

4.5 Quality Metric

To objectively evaluate the quality of reconstructed video after being compressed, or using equipment rather than subjective human viewers, we need to establish metrics for quality that are a function of both the original non-impaired and the compressed signals. One of the widely used quality measures is called the Mean Squared Error MSE:

$$MSE = \frac{1}{N_r * N_c} \sum_{x=0}^{N_c-1} \sum_{y=0}^{N_r-1} (p(x,y) - p\hat{}(x,y))^2 \qquad [4.5]$$

Where p and p^ are the original and reconstructed video. Nr is number of rows, and Nc is the number of columns. And from this measure, there are 2 Signal-to-Noise ratio SNR measures that can be calculated:

$$SNR_p = 10 \log \frac{(255)^2}{MSE} \qquad [4.6]$$

And $$SNR_v = 10 \log \frac{\sigma_s^2}{MSE}$$ [4.7]

Assuming the picture was quantized by 8 bits per pixel, 255 is the maximum intensity, σ_s^2 is the variance of the original picture. SNRp is called peak-to-peak SNR and SNRv is variance SNR. These MSE and SNR are simple equations that reflect the distortion in the picture. However, they cannot reflect the frequency distribution of the errors, and thus don't reflect the human quality perception [Meer97]. The reason is human visual system is frequency dependent.
A way to compensate for that is using the "Parseval Theorem" and calculating the MSE in the Fourier domain to get the different frequency components [Katt91]:

$$WMSE = \frac{1}{N_r * N_c} \sum_{u=0}^{N_c-1} \sum_{v=0}^{N_r-1} \left(w(u,v)\left(P(u,v) - P\hat{}(u,v)\right)\right)^2$$ [4.8]

Where w (u, v) are the weighing coefficients according to the CSF of the human eye. P (u, v) is the Fourier transform coefficients of the picture for frequencies (u, v). This function can be done in the DCT domain rather than the Fourier one, and in this case, the P (u, v) is replaced by the DCT coefficients c (u, v) and w (u, v), which are the visibility thresholds for the individual DCT coefficients.
The MSE and SNR metric is the simplest of quality metrics. Other metrics that consider the masking of the Human visual system are more complex. There are 2 classes of metrics Quality [Bass96]:

4.5.1 Metrics based on the Human visual system

In this case, the HVS is modeled, splitting the picture into frequency bands like the HVS. We then calculate the contrast from the pixel values from each band. The contrast can be defined

as "Weber contrast" [Olz86] for stimuli that are symmetric relative to the background luminance as:

$$C_w = \frac{\Delta L}{L} \qquad [4.9]$$

Where L is the luminance difference in the pattern, and L is the background luminance. One HVS system normalizes the Local Band-limited Contrast (LBC) to the visibility threshold [Wes95]. Other models express the signal in just noticeable difference (JND) [Jay93]. Adjusting the visibility threshold incorporates the edge and texture masking. The distortion is then measured in each pixel in each of the frequency bands. One of the models referred to by Van der Meer [Wes95] measures JND via the function of Masked LBC difference $MLBC_{K,I}(x, y)$:

$$JND_{K,I}(x,y) = \Delta MLBC_{K,I}(x,y) = \frac{LBC_{K,I}(x,y) - LBC^{*}_{K,I}(x,y)}{TE_{K,I}(x,y)} \qquad [4.10]$$

Where $LBC_{K,I}(x, y)$ and $LBC^{*}_{K,I}(x, y)$ are the local band contrast for the original and impaired video sequence. TE is a threshold elevation function as shown in fig 4.7.

After measuring visibility of difference between original and impaired picture for each pixel each frequency band, we should try to get a measure for the quality of the whole picture. One way to do that is by pooling of errors means (PEM). [Wes95] describes a method for calculating that:

$$PEM = \left(\sum_{x,y} \left| \sum_{k,I} \left| JND_{K,I}(x,y) \right|^{\alpha} \right|^{\beta} \right)^{\gamma} \qquad [4.11]$$

In this case, α, β, γ are constants that affect the relation between JND and the PEM. They could be used to make the correlation between this metric and the subjective measurement values closer.

4.5.2 Metrics based on Visual Artifacts:

This is the other approach for measuring quality, by measuring artifacts in compressed video like blocking, blurring, false contour...etc. The subjective evaluation is based on a combination of those artifacts affecting compressed video.

Miyahara [Miya92] developed a model that extracts five features from a stream of video. A simple model is made after that. Then it's tuned to a maximum correlation to subjective evaluation.

Other models reflected another combination of features. These features could be related to spatial distortions, changes in motion energy. Sometimes these featured are combined linearly in a metric. In other cases like stated by [Meer97]. [Lin95] used a non-linear back-propagation neural network to combine different features in a quality metric. This area of research is still open for new approaches.

Chapter 5

Algorithms of Video Quality

The models for digital video quality evaluation fall under the 3 categories described earlier:

1- Double Stimulus, Full Reference

2- Feature Extraction / Reduced Reference Models

3- Single Stimulus / No Reference Models

There has been extensive work done in this area, and many algorithms introduced. Below is a description of some of the already introduced models. There has been effort done in the past 2 years in the single stimulus or no reference models. This is for the need of TV networks to do on-line monitoring of their programming. Research in this area by [Cav01, Ham01, Kne01 & Tan01].

5.1 Double Stimulus Systems (Full and Reduced Reference FR/RR models:

5.1.1 JND Model (Just Noticeable Difference), Sarnoff Model, USA

The process of digital video reproduction entails the introduction of distortion and artifacts through the whole process. This includes the processes of initial capture (by camera for example), encoding, transmission, decoding and display.

The JND model introduced by Sarnoff makes predictions of perceptual ratings of a degraded video stream compared to a reference or clean one of the same material. It's based on a spatio-temporal human visual model. The difference of the 2 signals is quantified in units of the human Just Noticeable Difference (JND) [Sar97, Lubin95].

The basic block diagram of the system is shown in Fig 5.1.

The reference signal is un-distorted and the test sequence is a degraded signal, which passes through the "System Under test". This in turn could be one of a number of sources of distortion like an encoder, the transmission channel, or the decoding.

The model produces a sequence of JND maps.

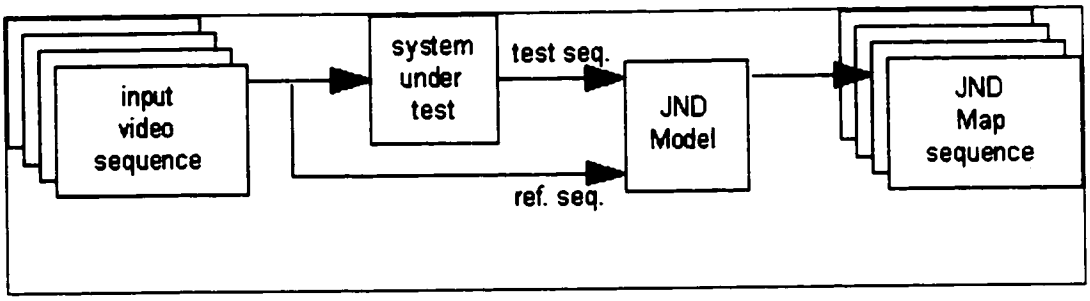


Fig 5.1 Block Diagram of JND Model [Sar97, Lubin]

Fig 5.2 shows the architecture of the model. As mentioned 2 sequences are entered to the system, one reference, and the other processed sequence. In the first stage the processing transforms the Y', Cb', Cr' into R', G' and B' gun voltages. After that point nonlinearity is applied to those values. This models the transform from R', G', B' gun voltages to model intensities (R,G and B) of the display (fractions of maximum luminance) [Sar97]. After nonlinearity, the vertical-electron beam is modeled by replacing the interline values in fields R, G, B by interpolated values from above and below. Then the vector (R, G, B) is subjected to a tri-stimulus coordinates (X, Y, Z). The luminance component Y of that vector is passed to the luminance processing part of the model.

And to ensure perceptual uniformity at each pixel, of the color space to isoluminant color difference, the model maps the pixels into CIELUV (An international standard uniform –color space). The chroma components u*, v* are passed to the chroma processing step.

After that, in Pyramid Decomposition, each luma value is subjected to a compressive nonlinearity. Then each luma field is filtered and down-sampled in a four level Gaussian Pyramid [Burt83] to model the psychophysically and physiologically of visual signals into different spatial-frequency bands. This process efficiently generates a range of spatial resolutions for subsequent filtering operations. After that, there are similar operations like oriented filtering done at each pyramid level.

Next, the Normalization stage sets the overall gain with a time-dependent average luminance, to model the visual system's relative insensitivity to overall light level, and to represent such effects as the loss of visual sensitivity after a transition from a bright to a dark scene or insensitivity to distortion in busy areas. [Nach74].

After normalization, three separate contrast measures are calculated. In each case, the contrast is a local difference of pixel values divided by a local sum, appropriately scaled as a function of pyramid level so that the result is 1 when the image contrast is at the human detection threshold. This establishes the definition of JND, which is passed on to subsequent stages of the model.

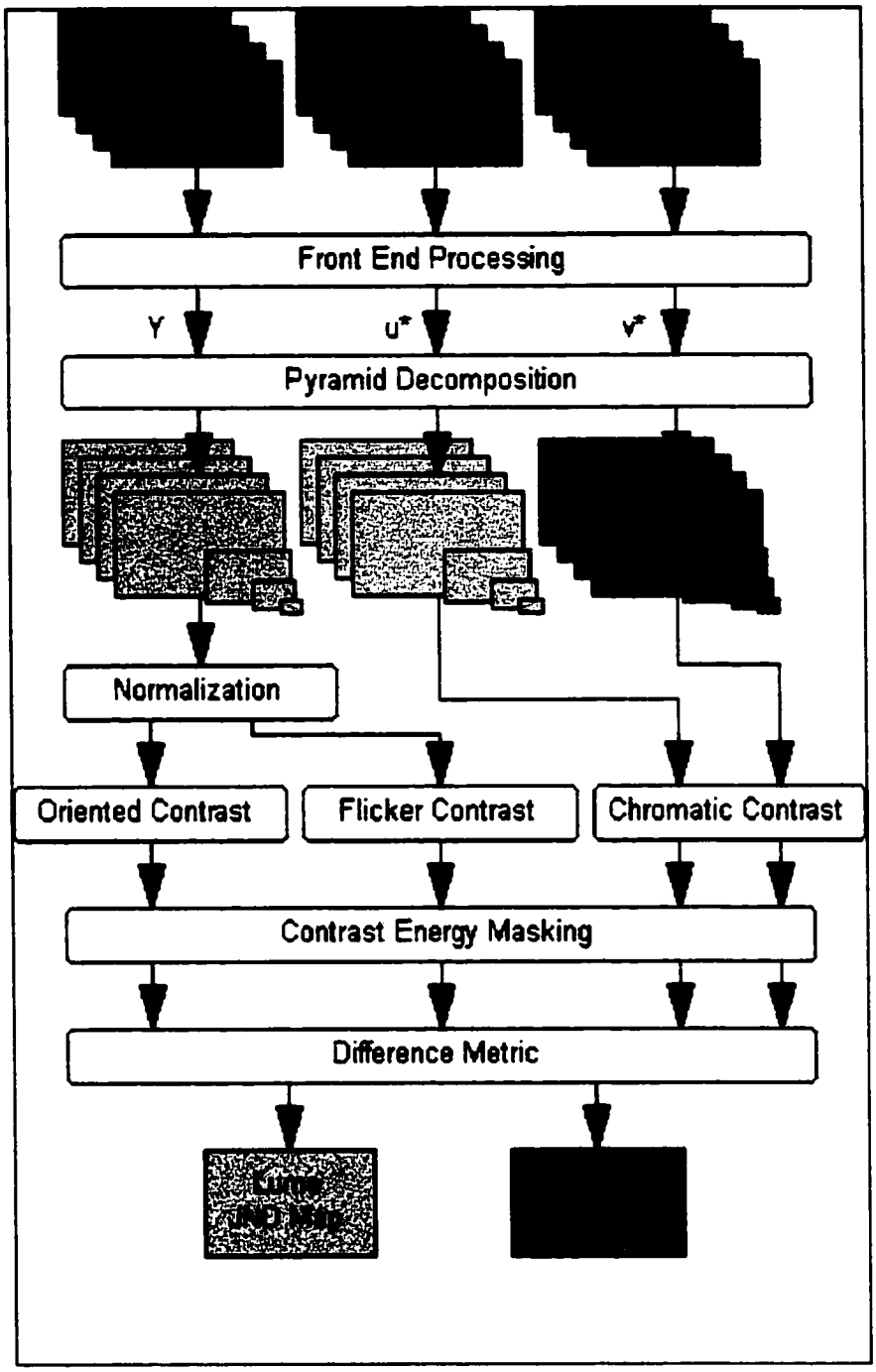


Fig 5.2 JND model Architecture [www.mpeg.org/MPEG/JND]

In the Contrast Energy Masking stage, each contrast image is subjected to a point non-linearity, the gain of which is controlled by the response across other resolution levels and channels. This gain-setting is included to model visual masking effects such as the decrease in sensitivity to distortions in "busy" image regions.

In the Difference Metric stage, outputs from the test and reference sequences are combined via a simple difference operator and then summed across pyramid levels and channels to return the number of JNDs in both luma and chroma.

Separate JND maps for luma and chroma can then be combined into one map. Summary statistics can also be obtained at this point.

Fig 5.3a Reference Image

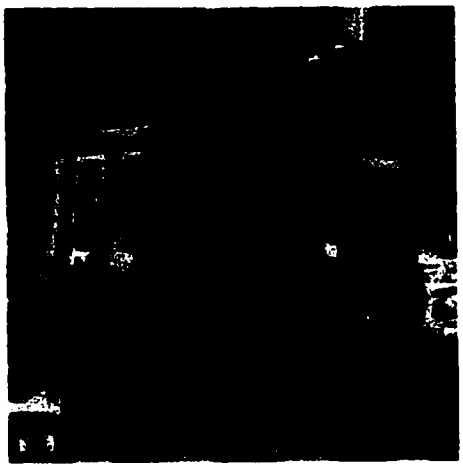


Fig 5.3b Processed Image

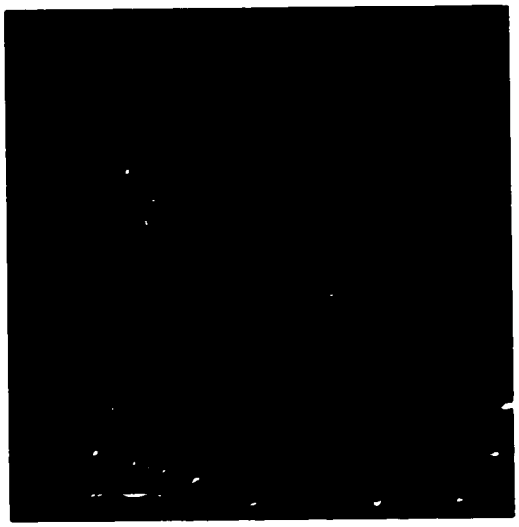


Fig 5.3c The difference map or JND map output according to the model

As well, the luma and chroma JND maps are reduced to one number, luma and chroma Picture Quality Ratings (PQRs). This is done by doing a histogram on the JND values for all the pixels over a threshold, then adopting a 90 percent value as the PQR value [Sar97]. The luma and chroma PQR numbers are combined by a linear combination of a sum and maximum to get the PQR for the field being processed.

5.1.2 The NTIA model, USA

This model is used for in service monitoring of video quality. It extracts some features from the input and output video scenes and compares them concluding the degradation in quality. These features are transmitted from transmission side to the receiving side by an ancillary data channel. This is shown in fig 5.4

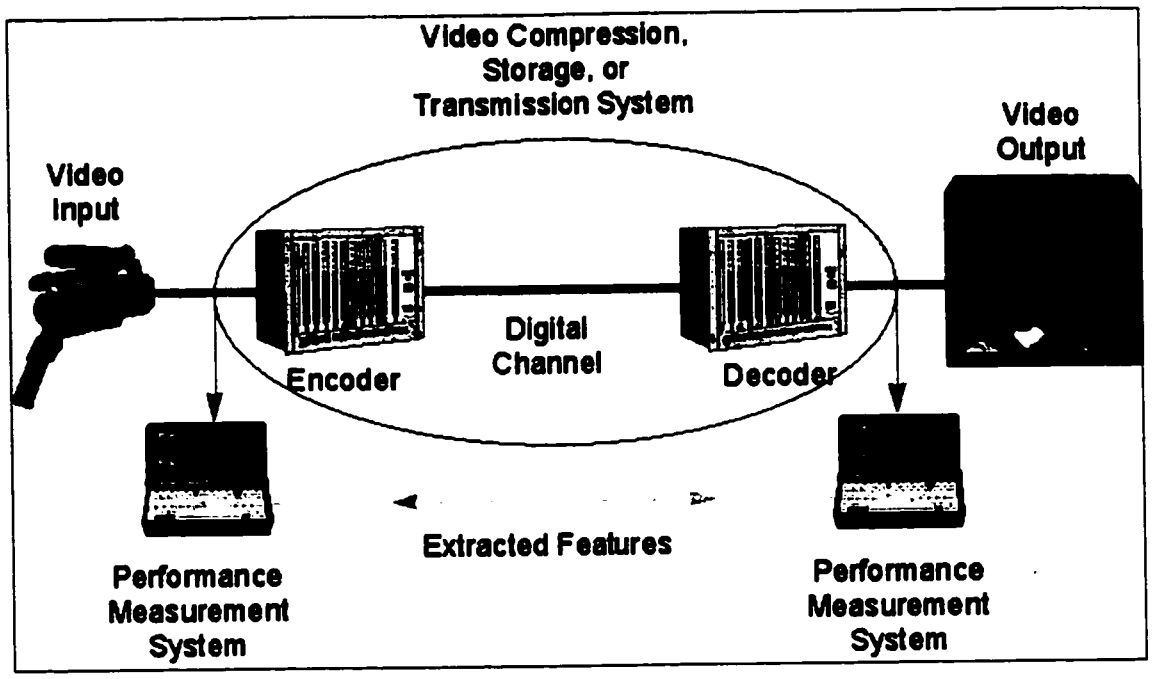


Fig 5.4 The Extracting Feature Approach for Quality Measurement [Wolf98]

For that, it uses a concept of Spatial - Temporal (S-T) regions from the pictures. Fig 5.5 gives a region of 3 pixels by 3 lines in a consecutive 3 frames as the S-T region considered here. The S-T region can vary. Large S-T regions can be used when the ancillary data channel bandwidth is small, while small regions are used when the ancillary data channel bandwidth is large. As more features are associated with smaller regions, this will need more bandwidth.

The way the system works is very straightforward. There is an input and an output calibration processor. The input processor estimates and adds the video transmission system delay to the input stream to synchronize with the output stream. This is achieved by correlating low bandwidth temporal features like motion features between the input and output calibration processors [Wolf98].

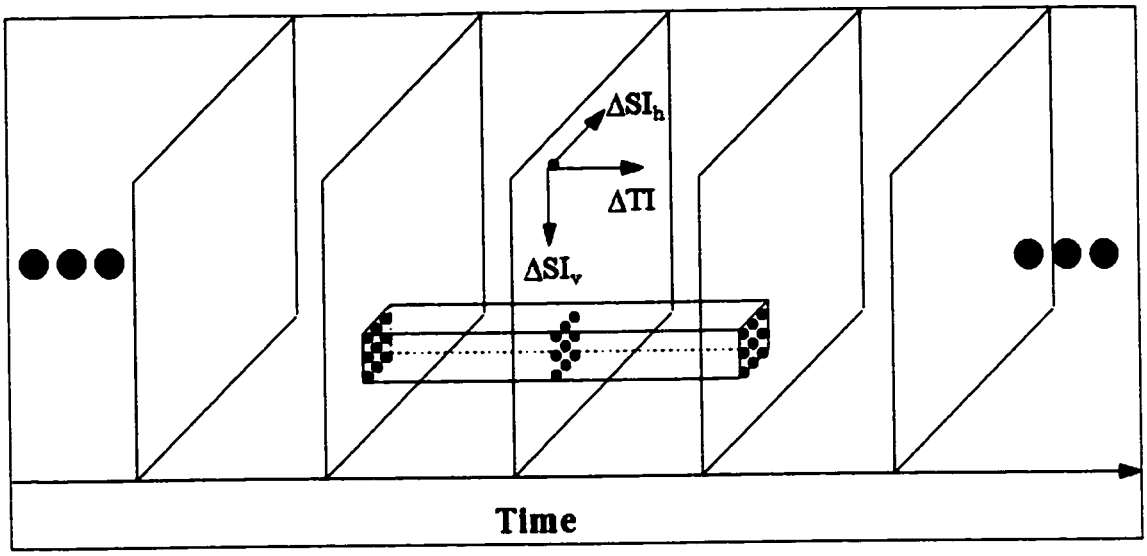


Fig 5.5 Sample Spatial Temporal region (S-T)

After this time alignment is done, the output calibration processor estimates the spatial shifts, gains and level offset of the transmission system, and applies the corrections to the output video stream.

After that, there are 4 programmable filters to extract features from the calibrated input and output video. They are spatial, temporal, spatial-temporal and chroma. The filters are programmable; meaning they can extract the features from variable S-T regions. The size of the S-T region can be changed based on the bandwidth of the ancillary data channel. The features chosen after extensive research to characterize the activity of motion, edges and colors.

The Spatial feature gives the activity of image edges. The digital video system can add edges creating edge noise, or reduce edges causing blurring. The temporal feature gives the temporal difference. The video system here can add motion causing blocking, or reduce motion causing frame repeats or picture freeze. In the chrominance feature, the video system can add color causing cross color or color artifacts, or reduce color causing color sub-sampling. The spatial temporal feature is a cross between the 2. An example of added spatial temporal artifact is mosquito noise, which is visible in stationary background around moving objects [Wolf98].

The benefit of this system is that it can do in-service measurement of quality without knowing the input video scenes, and needs only an ancillary data channel for the transmission of features.

5.1.3 Image Segmentation Model- CPqD-Brazil

This model assesses video quality using objective parameters based on image segmentation. Natural scenes are segmented into plane, edge and texture regions. A set of objective parameters are assigned to each of these contexts. Fig 5.6 shows the configuration of the objective parameters computation used.

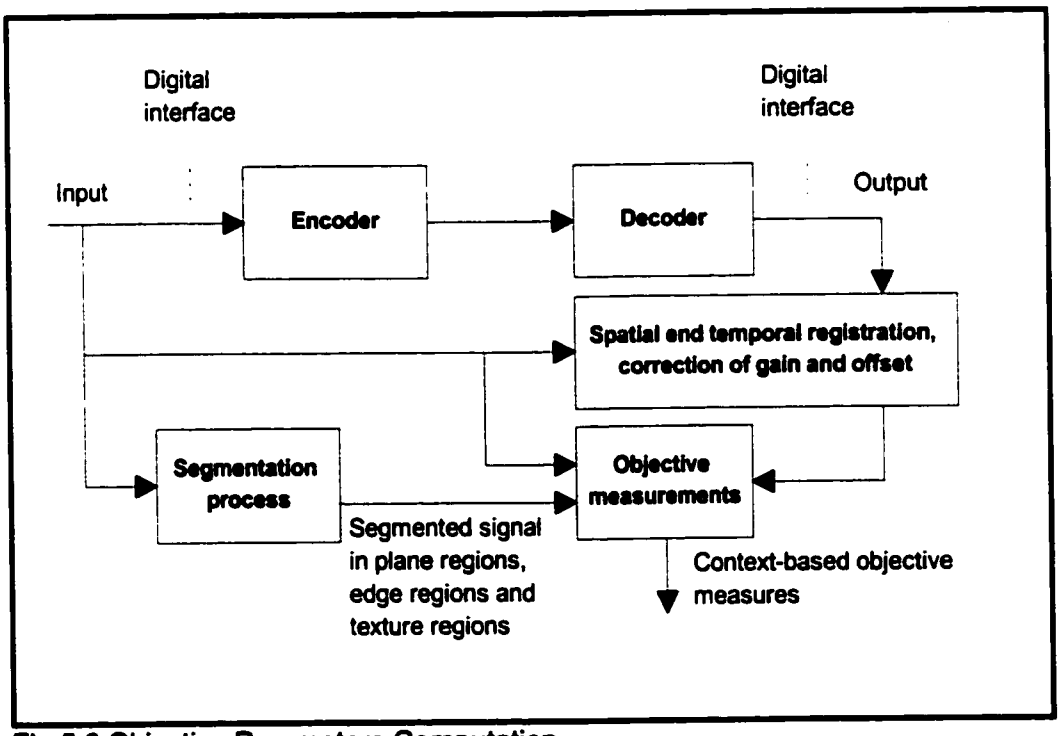


Fig 5.6 Objective Parameters Computation

Each objective parameter is computed separately within the context of plane, edge and texture regions. For example, a blocking distortion can be measured by an edge detector applied to the plane regions of the video scenes.

This type of algorithm usually needs a high computational complexity. However, this is reduced by using low complexity estimators, and by limiting their computation within the context of the scenes [Pess97].

There is a process of alignment needed, where the spatial and temporal registration between the input and output video, and the correction of gain and offset are estimated.

The objective parameters are computed by comparing the original to the impaired video sequences. The estimators are applied to fields rather than frames to ensure the reliability of the measurement in high level of motion scenes.

A perceptual based model, that predicts subjective rating, is defined by computing the relationship between objective measures and the results of subjective assessment tests applied on a set of natural scenes and MPEG-2 codecs. The scene-dependent perceptual models are defined in 2 steps:

A- The relationship between each parameters and the subjective impairment level is approximated by a logistic curve.

B- The final result is achieved by linearly combining the estimated impairment levels, where the weight of each impairment level is proportional to its' statistical reliability.

For spatial segmentation, there has been 3 algorithms developed:

1- Image segmentation based on edge detection using recursive filtering

This algorithm classifies each pixel based on the brightness variance computed within a neighborhood of the pixel. The resulting image is then smoothed by a median filter. The algorithm applies an edge detector on the image based on recursive filtering. The edge on the boundary of the plane regions is classified as belonging to the edge regions. The texture regions are the remaining regions of the picture.

Fig 5.7a,b show the mobile and calendar scene after segmentation. The plane regions are represented by white pixels, edge regions by gray pixels, and texture regions by black regions.

Fig 5.7a Part of Mobile & Calendar

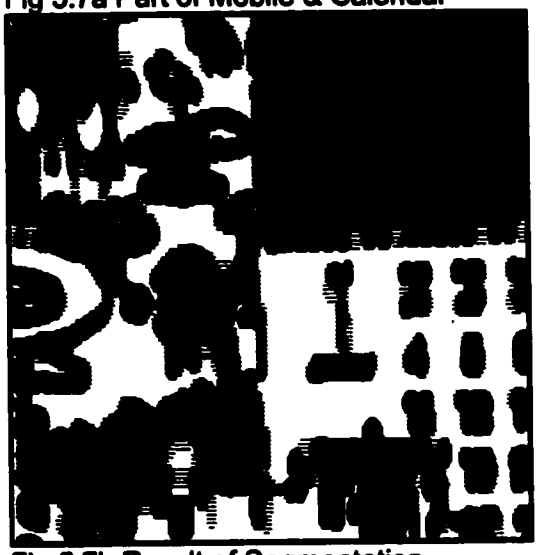


Fig 5.7b Result of Segmentation

2- Fuzzy Image segmentation based on spatial features:

This algorithm is divided into 2 steps. In the first step, the algorithm assigns a membership function to each of the 3 contexts under classification. For plane regions, the value of a pixel is defined inversely to the brightness variance within a neighborhood of the pixel. There is a "Morphological gradient" [Pess97] applied to define the membership function of the edge regions. The compliments of the fuzzy union between these 2 functions define the membership function of the texture regions. In the second step, each pixel is classified as belonging to context with highest value of membership among its' 3 membership values.

3- Image segmentation based on watershed:

In this algorithm, the luminance component is simplified by applying edge smoothing filter, increasing its' homogeneous regions. Then a "watershed algorithm" is applied to the morphical gradient of the simplified image. The watershed detects homogeneous regions, called catching basins. The plane regions are the catching basins with area greater than a threshold. The texture regions are given by the erosion of the complement of the plane regions, and the edge regions are the remaining ones.

Pessoa [Pess97] states that by using region-based objective measurements, more accurate predictions are obtained, compared on predictions based on global parameters.

5.1.4 DVQ model- NASA, USA

This algorithm accepts a pair of digital video sequences, and computes a measure of the magnitude of the visible difference between them. The metric is based on the Discrete Cosine Transform. It incorporates aspects of early visual processing, including light adaptation, luminance and chromatic channels, spatial and temporal filtering, spatial frequency channels, contrast masking, and probability summation. It also includes primitive dynamics of light adaptation and contrast masking. Fig. 5.8 shows a block diagram of the model.

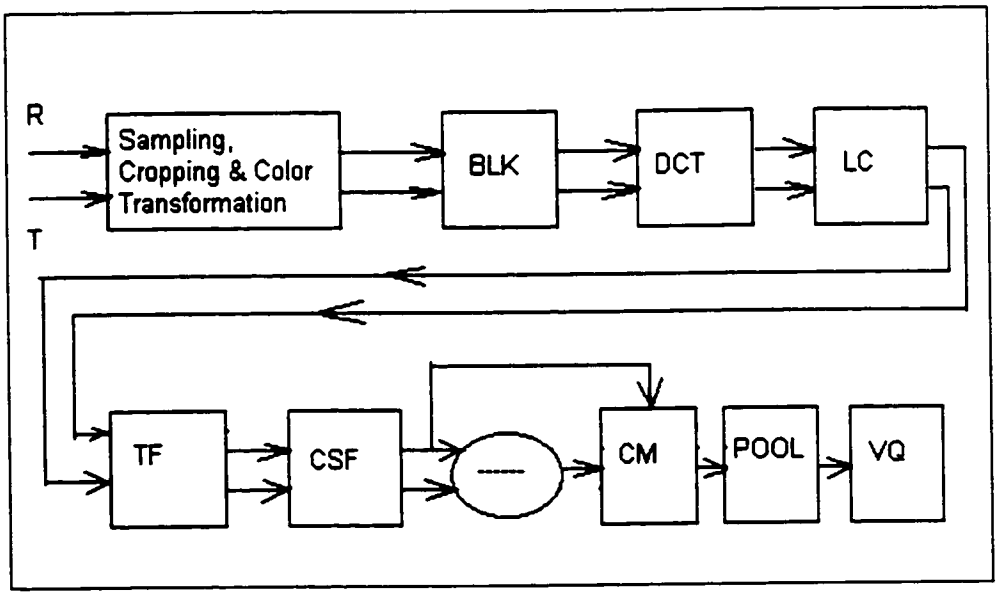


Fig 5.8 Overview of the DVQ Algorithm processing steps [Watson99]

The DVQ (Digital Video Quality) metric incorporates many aspects of human visual sensitivity in a simple image processing algorithm. One of the most complex and time consuming elements are the spatial filtering operations employed to implement the multiple, band-pass spatial filters that are characteristic of human vision.

This algorithm uses the Discrete Cosine Transform (DCT) for this decomposition into spatial channels. This provides a powerful advantage since efficient hardware and software are available for this transformation, and because in many applications the transform may have already been done as part of the compression process.

Watson [Watson99] describes the model as shown in fig 5.8 as follows; The input to the metric is a pair of color image sequences: reference (R), and test (T). The first step consists of various sampling, cropping, and color transformations that serve to restrict processing to a region of interest and to express the sequences in a perceptual color space. This stage also deals with de-interlacing and de-gamma-correcting the input video. The sequences are then subjected to a blocking (BLK) and a Discrete Cosine Transform (DCT), and the results are then transformed to local contrast (LC). Local contrast is the ratio of DCT amplitude to DC amplitude for the corresponding block.

After that, a temporal filtering operation (TF) is applied, which implements the temporal part of the contrast sensitivity function. This is accomplished through a suitable recursive discrete second order filter. The results are then converted to just-noticeable differences by dividing each DCT coefficient by its respective visual threshold. This implements the spatial part of the contrast sensitivity function (CSF).

After that, the two sequences are subtracted. The difference sequence is then subjected to a contrast masking operation (CM).

Finally the masked differences may be pooled in various ways to illustrate the perceptual error over various dimensions (POOL), and the pooled error may be converted to visual quality (VQ).

5.1.5 The Perceptual Distortion Metric (PDM), EPFL, Switzerland

This metric is based on a spatio-temporal model of the human visual system. It consists of 4 stages as shown in fig 5.9. The reference and processed or impaired sequence pass through those 4 stages. The first one converts the input to an opponent-color space. The second stage implements a spatio-temporal perceptual decomposition into separate visual channels of different temporal frequency, spatial frequency and orientation. The third stage models effects of pattern masking by simulating excitatory and inhibitory mechanisms according to a model of contrast gain control. The final stage processes the pooling and detection stage and calculates a distortion measure from the difference between the sensor outputs of the reference and the processed sequence [Wink99].

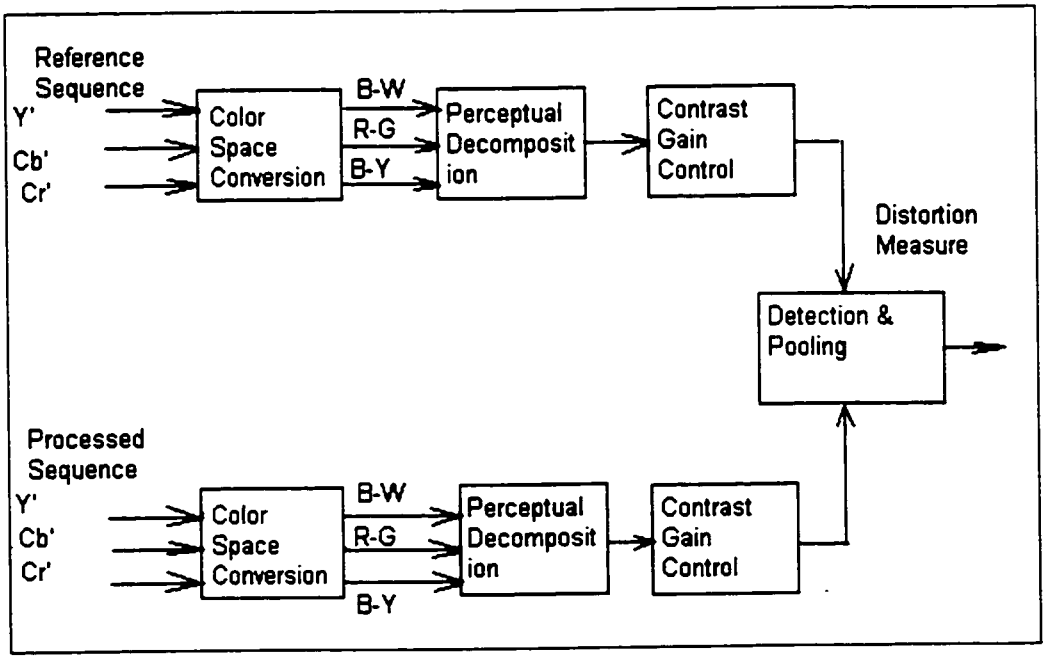


Fig 5.9 Block diagram of Perceptual Distortion Metric [Wink99]

5.1.6 Criticality Model, NHK, Mitsubishi, Japan

This model introduces a new concept called "Criticality". A definition of "criticality" is proposed as a quantitative measure of difficulty for MPEG-2 video coding, to analyze the picture quality of television programs statistically, depending on the complexity of the content, since picture quality in digital coding depends on picture characteristics as spatial detail and motion. [Nish97] Criticality is useful not only for estimating picture quality distribution characteristics but also for selecting a set of test sequences. By measuring frequency of occurrence of criticality and relating it to subjective picture quality, it becomes possible to obtain the statistical quality distribution of television programs in digital broadcasting.

The definition of criticality and the procedure to derive statistical picture quality distribution have been included in the latest revised ITU-R Recommendations BT.1210 (Test materials to be used in subjective assessment) and BT.1129 (Subjective assessment of standard definition digital television (SDTV) systems) [Nish00].

Criticality is defined as "the number of output bits per pixel from a hybrid DCT encoder with a fixed quantizer". Criticality is averaged over each whole frame. Figure 5.10 shows the configuration of the criticality measurement equipment based on the definition. This definition corresponds to the bit rate required to obtain an almost constant picture quality for various sequences, because a critical sequence requires a higher bit rate to maintain picture quality to the level of the non-critical sequences.

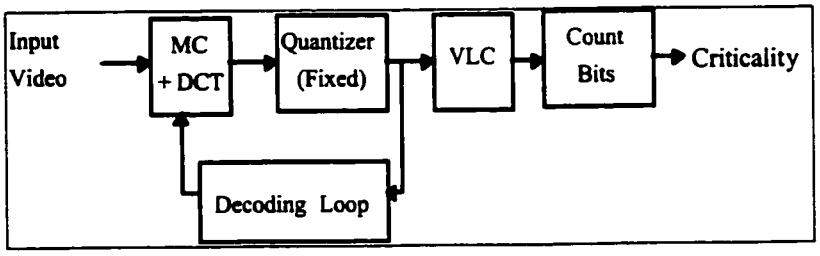


Fig 5.10 Configuration of criticality measurement [Nish00]

The procedure suggested by Nishida to evaluate picture quality distribution characteristics is shown in fig 5.11, and consists of 5 steps:

1- Measure criticality of test sequences used in subjective assessment.

2- Measure criticality distribution of broadcast programmes for a long time period.

3- Conduct a subjective assessment of picture quality of the system under test.

4- Derive a relationship between criticality and subjective picture quality for the test sequences.

5- Derive picture quality distribution characteristics (quality vs. frequency of occurrence) by combining the results of Step 4 (criticality vs. quality) and Step 2 (criticality vs. frequency of occurrence).

This method enables to evaluate the performance of coding systems in terms of frequency of occurrence of certain picture quality for overall broadcast programs. [Nish97, Nish00]

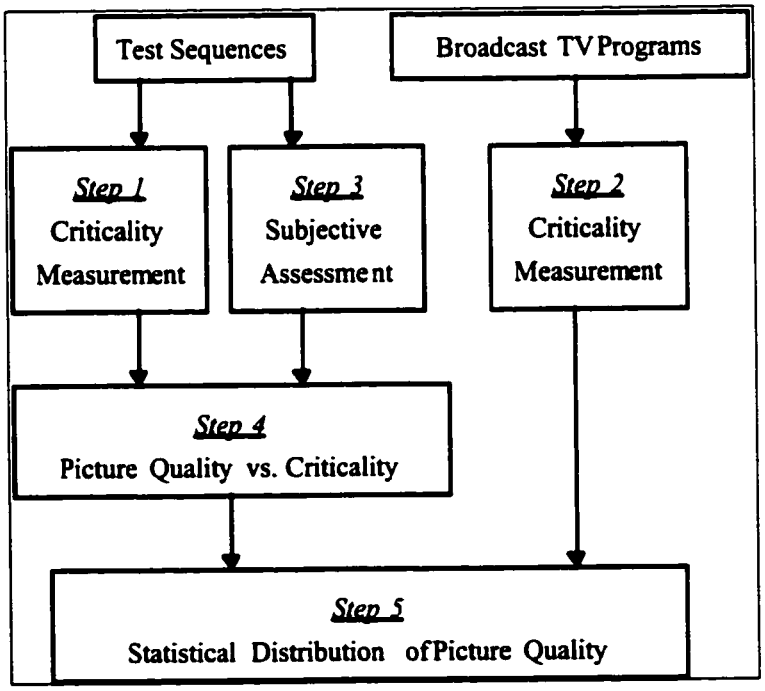


Fig 5.11 Procedure to derive picture quality distribution characteristics

5.1.7 KDD/Pixelmetrix Model, Japan

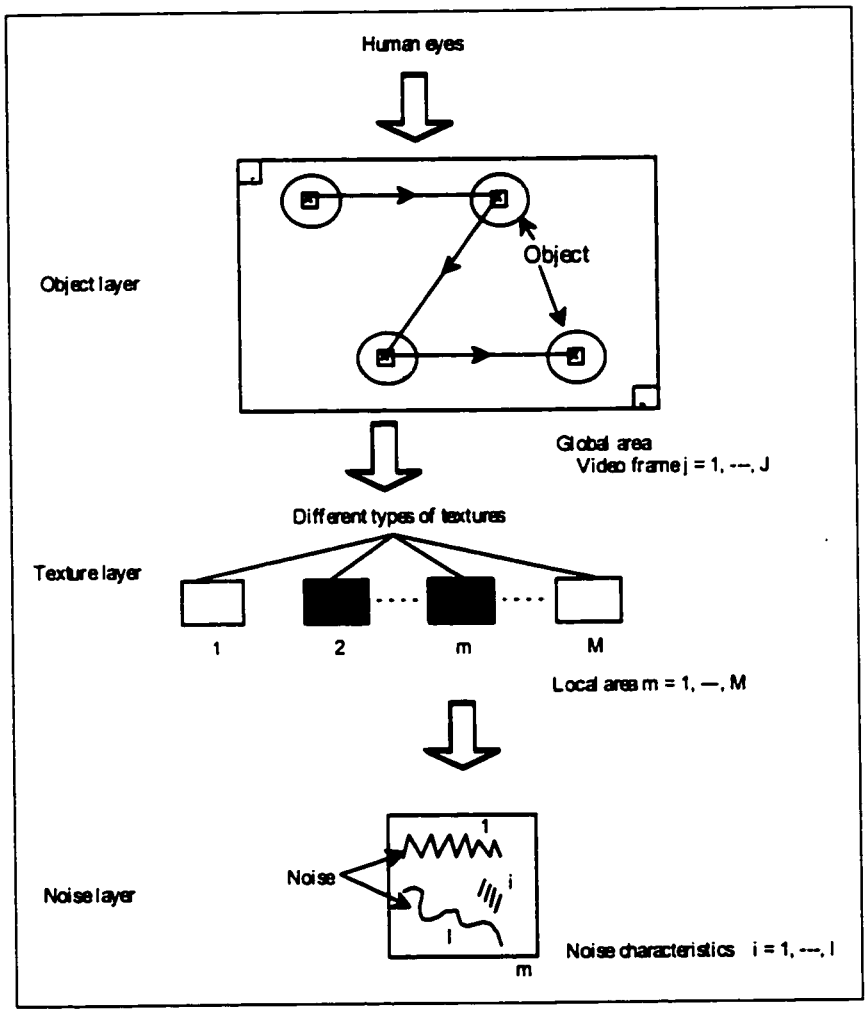


Fig 5.12 The 3 layered model for quality evaluation

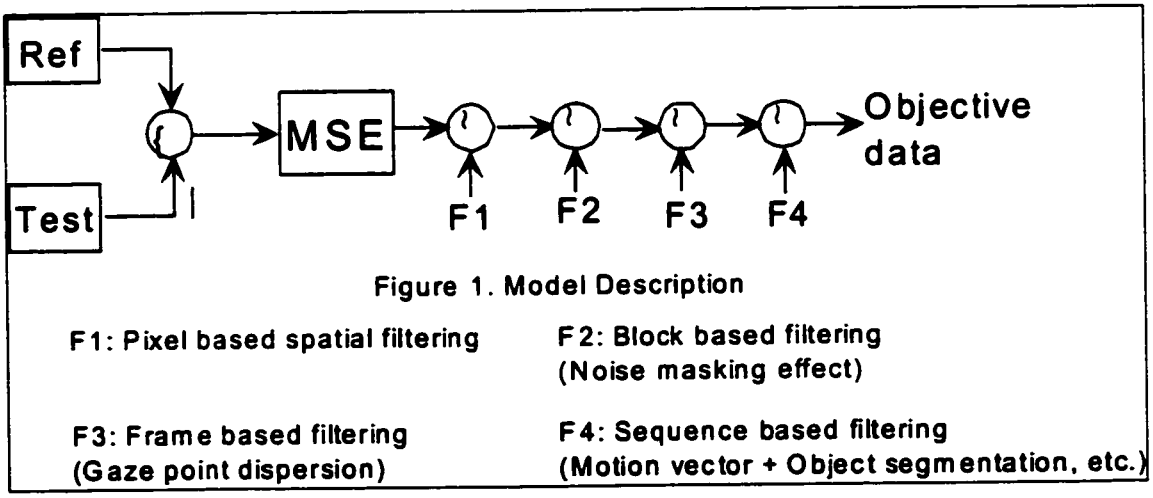


Fig 5.13a The KDD model

Figure 5.12, fig 5.13 shows the three-layered picture quality model as seen by the human eye. The human eye cannot watch a whole frame at a glance, but watch only a local spot area in a frame, which is around the gaze point of the human eye, and recognizes the texture and also quality of the area depending on the degree and characteristics of noise mixed in this texture [Ham00]. The whole frame is understood by moving the gaze point among objects for the whole frame at the same time. In this process, picture quality is determined by the noise over a frame. Therefore, to perform objective measurement of subjective picture quality, the macro to micro three-layered picture structures (object, texture and noise layers) are used, and a bottom-up noise weighting scheme is proposed which uses a particular weighting function at each layer taking into account human visual perception.

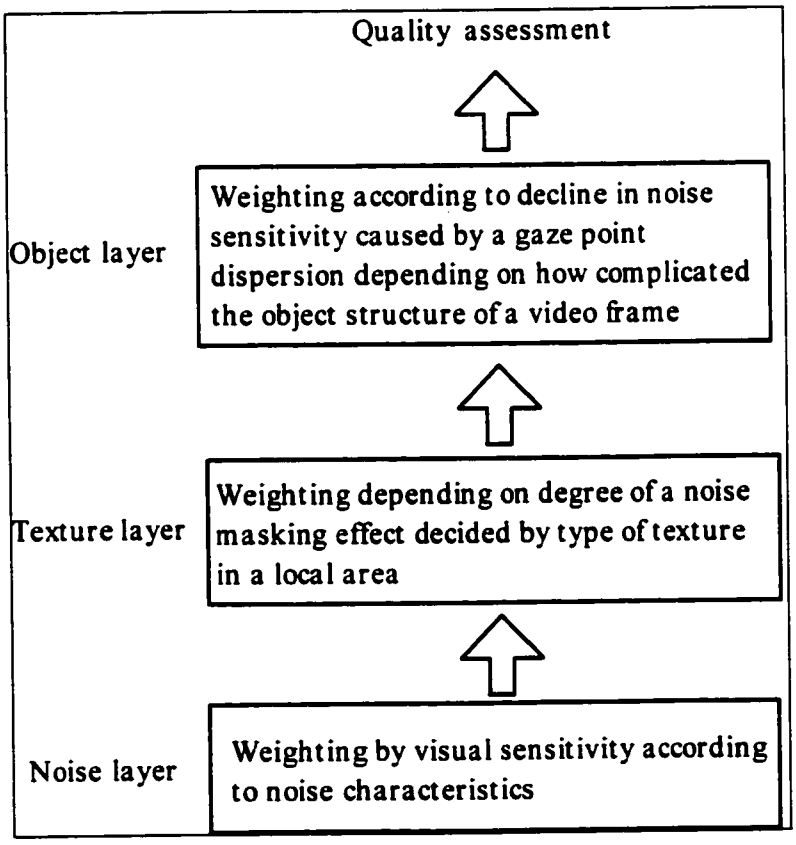


Fig 5.13b The 3 layer model of noise weighting [Ham00]

First, at the noise layer, common noise in a video compression process such as high frequency noise, low frequency noise, chroma noise, jerkiness, flicker ...etc are weighted depending on

their degrees and characteristics. After that, local spot areas are classified into several groups by their texture types. These groups include for example, forest, trees and a sports stadium in which noise are strongly masked, and "flat texture" such as a human skin and a sky in which noise are easily recognized. Thus, noises are weighted more or less according to their texture types.

Finally, at the object layer, the dispersion degree of the gaze point is predicted by measuring how complicated the structure is of objects in the video frame. Then, noises in the whole frame are weighted corresponding to a decline in noise sensitivity caused by this dispersion. [Ham00]

5.1.8 The KPN/Swisscom CT model, Netherlands

The Perceptual Video Quality Measure (PVQM) as developed by KPN/Swisscom CT uses the same approach in measuring video quality as the Perceptual Speech Quality Measure [Bee94] in measuring speech quality. The method was designed to cope with spatial, temporal distortions, and spatio-temporally localized distortions like found in error conditions. It uses ITU-R 601 input format video sequences (input and output) and re-samples them to 4:4:4, Y, Cb, Cr format.

A spatio-temporal-luminance alignment is included into the algorithm. Because global changes in the brightness and contrast only have a limited impact on the subjectively perceived quality, PVQM uses a special brightness/contrast adaptation of the distorted video sequence. The spatio-temporal alignment procedure is carried out by a kind of block matching procedure. The spatial luminance analysis part is based on edge detection of the Y signal, while the temporal part is based on difference frames analysis of the Y signal.

Since the Human Visual System (HVS) is much more sensitive to the sharpness of the luminance component than that of the chrominance components. As well, the HVS has a contrast sensitivity function (CSF) that decreases at high spatial frequencies. These basics of the HVS are reflected in the model. The system has 4 steps:

1- The PVQM algorithm provides a first order approximation to the contrast sensitivity functions (CSF) of the luminance and chrominance signals.
2- The edginess of the luminance Y is computed as a signal representation that contains the most important aspects of the picture. This edginess is computed by calculating the local gradient of the luminance signal (using a Sobel like spatial filtering) in each frame and then averaging this edginess over space and time.
3- The chrominance error is computed as a weighted average over the color error of both the Cb and Cr components with a dominance of the Cr component.
4- In the last step the three different indicators are mapped onto a single quality indicator, using a simple multiple linear regression, which correlates well the subjectively perceived overall video quality of the sequence. [VQEG00, Bee97]

5.1.9 Tapestries Model, UK

This model used a different approach by designing separate modules specifically tuned to certain type of distortions, the select one of the results reported by these modules as the final objective quality score.

It consists of a perceptual model and a feature extractor. The perceptual model simulates the human visual system, weighting the impairments according to their visibility. It involves contrast computation, spatial filtering, orientation-dependent weighting, and cortical processing. The feature extractor is tuned to blocking artifacts, and extracts this feature from the video. The perceptual model and the feature extractor each produce a score rating the overall quality of the HRC video.

Since the objective scores from the two modules are on different dynamic range, a linear translation process follows to transform these two results onto a common scale. One of these transformed results is then selected as the final objective score, and the decision is made based on the result from the feature extractor. [VQEG00]

5.2 No Reference Models

5.2.1 The IFN/R&S Algorithm, Germany

This model which was developed by the Institut für Nachrichtentechnik (IFN), Braunschweig Technical University and Rohde & Schwarz, Germany, is a single ended system that evaluates the quality of degraded video and measuring DCT artifact (i.e. blocking). The typical application of such an algorithm is in the on-line monitoring of MPEG2 video quality.

Woerner and Lauterjung [Woe01] state that during a research at the IFN by Prof. Ulrich Reimers; "It became apparent that the visual impression of the MPEG-2 inherent blocking structure has the greatest impact on the picture quality in an otherwise normal video stream". Other kinds of artifacts like edge blurriness and mosquito noise that are also visible have been found to be of lower importance.

The model consists of four main processing steps. The first one is the detection of the coding grid used. In the second step based on the given information the basic parameter of the method is calculated. In the third step, the result is weighted by some factors that take into account the masking effects of the video content like spatial activity and temporal activity. Because of the fact that the model is intended for monitoring the quality of MPEG-coding, the basic version produces two quality samples per second, as the Single Stimulus Continuous Quality Evaluation method (SSCQE, ITU-R BT rec. 500) does. The submitted version produces a single measure for the assessed sequence in order to predict the single subjective score of the DSCQS test used in this validation process. [VQEG00]

The theory of the algorithm is based on the values associated with the 8x8 blocks and 16x16 macroblocks of the DCT transform. The digital video quality level is calculated from vectors, which contain information about the averaged difference between adjacent pixels. Comparing all the pairs in a block as shown in fig 5.14, the encoding process usually reduces the difference between adjacent pixels. The exception of that are the pixels on the edges of the blocks or macroblocks.

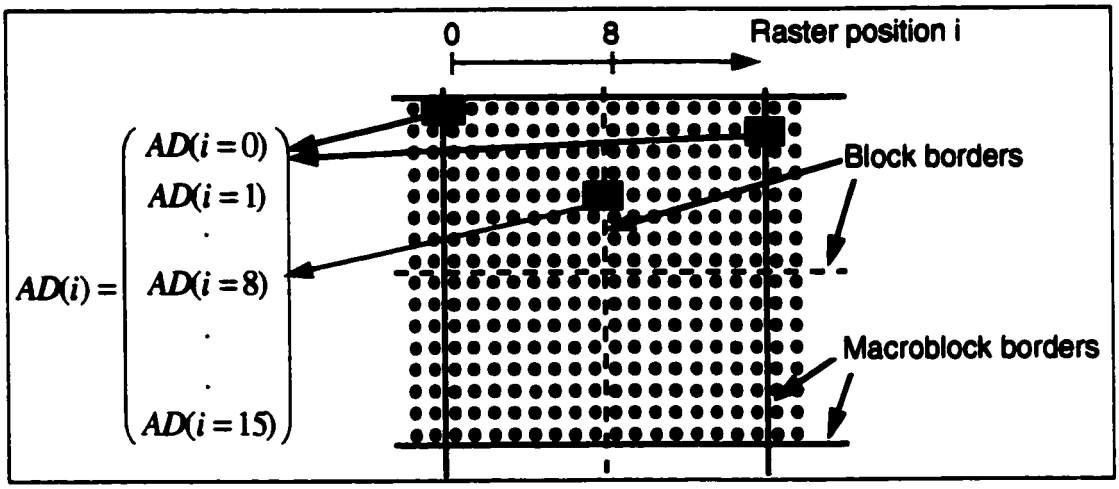


Fig 5.14 Sample calculation process for the amplitude differences of adjacent pixels [Lau98]

There are 16 vector elements built, which represent the average amplitude difference of all pixel pairs with the same relative distance within the macroblock grid. Fig 5.15 shows the result of the results of the calculation of the average differences on the original uncompressed stream of "Flowergarden". The values of all the pixel pairs are very close. The overall average represents the spatial activity of each frame.

After encoding the same sequence at a rate of 2 Mbits/sec then decoding it, the result is shown in fig 5.16. In this case, it's obvious that there are 2 values for the elements AD0, AD8 that are different from the rest of the block values. These values are at the edges of the block and macroblock. This affects the visibility of the blocking effect. In case the video was encoded at a higher rate for example, the difference between those 2 values and the rest of the block values will become less, hence reducing the blocking effect.

The spatial activity is calculated by averaging all pixel amplitudes differences regardless of their position within the frame. The temporal activity is calculated by averaging the amplitude difference of the same pixels in subsequent frames. In version 2.0 of the algorithm, this has been updated to use regions of pixels in subsequent frames. This made more accurate temporal

activity measurement, especially for slow moving camera or slow moving background scenes. For those scenes the previous system gave higher values for temporal activity.

There is masking applied to the values afterwards to come up with the objective weighted video quality measure that corresponds to the subjective evaluation.

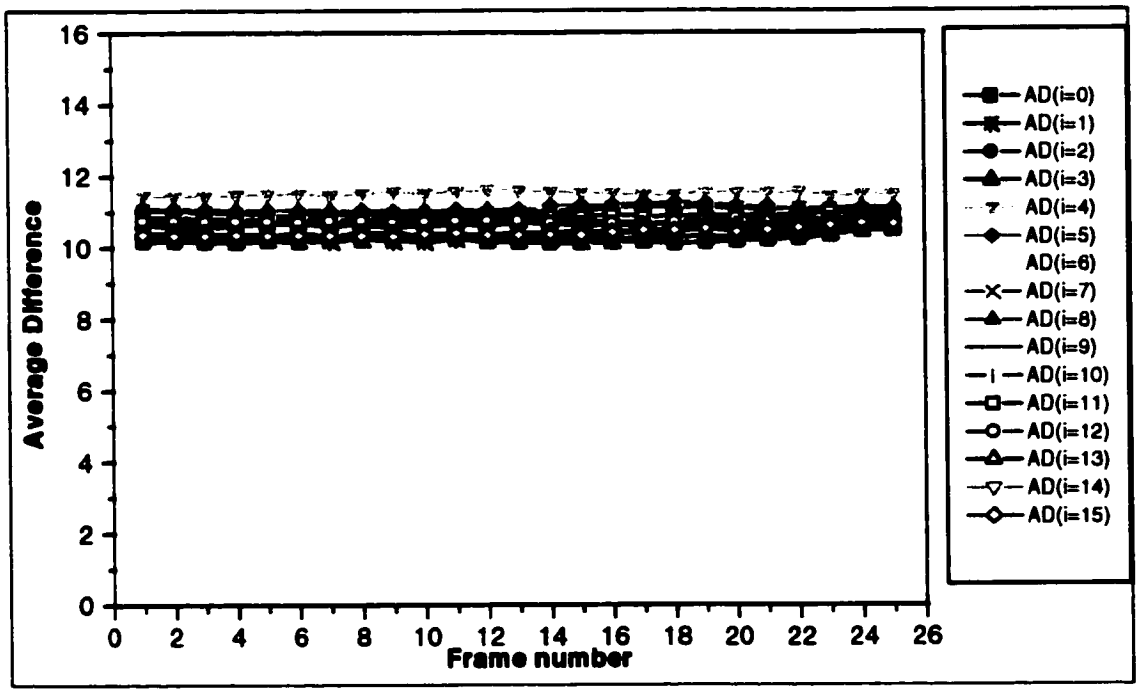


Fig 5.15 Averaged pixel amplitude differences, (Flowergarden stream uncompressed)

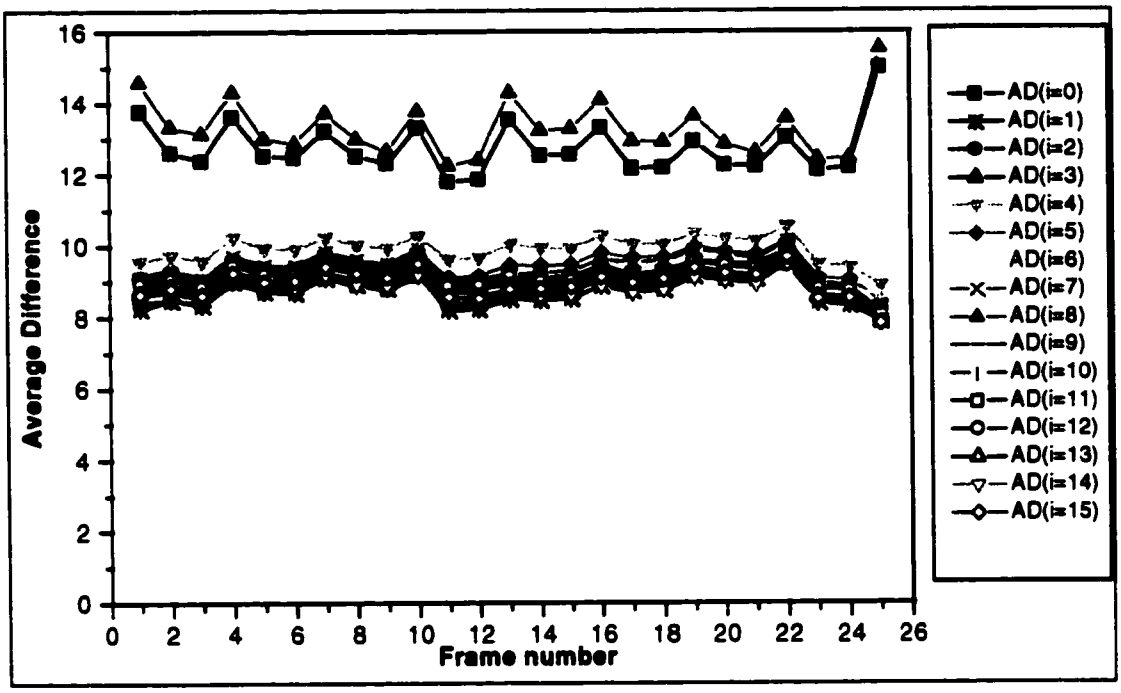


Fig 5.16 Averaged pixel amplitude difference (Flowergarden Sequence coded/decoded at 2 Mbits/sec)

The major benefit of this algorithm is that it doesn't need any reference signal to operate, and thus can be used in on-line monitoring of video networks, besides it can be put in any point in the network.

Chapter 6

Calibration of Objective to Subjective Values

6.1 The VQEG Work

One of the most exhaustive efforts done in the evaluation of digital video quality has been performed by the Video Quality Expert Group (VQEG). VQEG was formed in Oct. 1997 (CSELT, Turin, Italy) to make a framework for the evaluation of new objective methods for video quality. This would ultimately help the different ITU study groups develop new recommendation on the subject. A framework test plan was made in 1998 at the National Institute of Standards and technology, Gaithersburg, USA. The evaluation was made for 9 double-ended systems (mentioned earlier), and one single ended system (The IFN/R&S) system. They added one more metric which was the PSNR as a proponent.

The subjective testing was made according to the ITU-R BT.500 standard. Sessions where made into 20 minutes, with 2 sessions per day. Each stream was 8 seconds long with a 2 second Grey level in between different sources.

They followed the Double Stimulus Continuous Quality Scale (DSCQS). So, The viewer was shown each sequence then he would mark on a sheet on a scale 1-5 or 0 -100 the mark he felt corresponded to the quality value. The sequence of scenes always composed of a pair of the same scene. One of the streams in this pair would be the reference or clean signal, and the other would be the impaired signal. The viewer was not told which of the pairs was the clean and which was the impaired. This procedure is unlike the DSIS procedure, where the first scene of the pair was always the reference signal, and the second was the impaired one.

The score of each group was averaged for each sequence after excluding readings that were twice the deviation level from the mean value.

The Mean Opinion Score were calculated and DMOS or Differential Mean Opinion Score between reference and processed video quality values.

The VQEG divided their test material into 4 groups:

1- 50Hz or 625 line streams

2- 60Hz or 525 line streams

3- Low quality streams (768KB/sec – 4.5 MB/s)

4- High quality streams (3 MB/s – 50 MB/s)

This covered the different categories they considered. They used 20 test sequences (10- 50Hz 625 line, 10- 60Hz 525 line streams. The picture formats were different from MPEG2 with different profiles (MP@ML, SP@ML, 422p@ML and H263. They used 16 Hypothetical Reference Circuit) HRC to insert artifacts into each reference sequence.

6.1.1 Objective Model Evaluation Criteria:

The VQEG estimates a number of attributes are responsible for characterizing the performance of a video quality model. These are:

1- Prediction Accuracy

2- Prediction Monotonicity

3- Prediction Consistency

The outputs by the objective video quality rating model (the VQR's) are correlated with the viewer DMOS's (Differential Mean Opinion Score) in a predictable and repeatable fashion. The VQEG group in their objective test plan explored the possibility of the non-linearity in the relationship between predicted VQR and DMOS, as subjective testing can have nonlinear quality rating. However, later on it was accepted by VQEG and others like Woerner and Lauterjung [Woe01] that linear regression is acceptable as index value of the correlation.

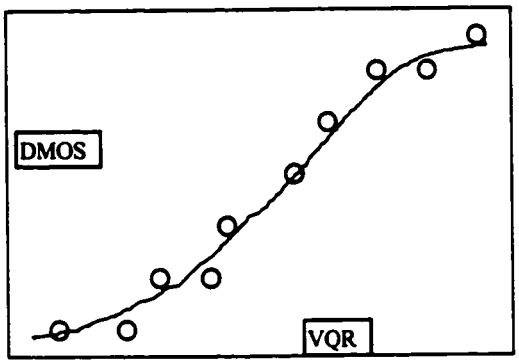


Fig 6.1 Relation between subjective and objective Measurement [VQEG-Ob98]

Fig 6.1 shows the relation graph between a subjective measured DMOS value and the predicted DMOS of the objective model. The scale could be 0-100 for both. This curve represents the correlation value for one specific objective model, and the points on the curve are the different video sequences used in both subjective and objective tests.

The nonlinear regression can be fitted to the [subjective/objective] data set. The 2nd, 3rd and 4th order polynomial nonlinear regression are discussed, besides the linear regression. Where the set of differences between measured and predicted DMOS is defined as the quality-error set Qerror[ ]:

$$Qerror[i] = DMOS[i] - DMOS_p[i]$$

The DMOS[I] is the measured differential mean opinion score for stream "I", and DMOSp[I] is the predicted DMOS by a specific objective model for the same stream i.

There were a number of metrics chosen as a measure for the above mentioned attributes [VQEG-Ob98].

5.2 Evaluation Metrics:

5.2.1 Metrics relating to Prediction Accuracy of a model

This attribute measures the ability of the model to predict the viewers' DMOS ratings with a minimum error "on average". [VQEG-Ob98]. There are a few metrics that can be used to measure the average error, with root-mean-square (RMS) error being a common one. To

include the known variance in subjective DMOS data, the simple RMS error can also be weighted by the confidence intervals for the mean DMOS data points. ITU500 describes a procedure to get 95% confidence level in the subjective measurements [ITU500]. The pearson linear correlation coefficient is a common metric that is related to the average error because lower average errors lead to higher values of the correlation coefficient, and it's a very simple relation to realize by mathematical or spreadsheet programs.

Metric1: The Pearson linear correlation coefficient between DOS_p and DOS, including a test of significance of the difference. Where DOS is differential opinion score for subjective measurement, and DOSp is for predicted objective measurement.

Metric2: The Pearson linear correlation coefficient between $DMOS_p$ and DMOS.

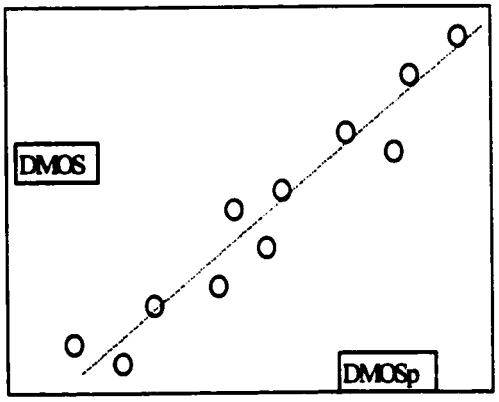


Fig 6.2 Model with greater accuracy [VQEG-Ob98]

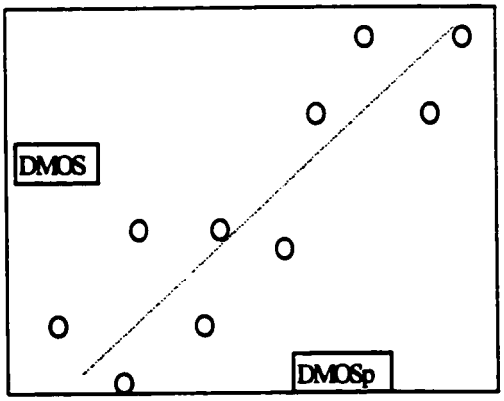


Fig 6.3 Model with lower accuracy [VQEG-ob98]

6.2.2 Metrics relating to Prediction Monotonicity of a model

An objective model's $DMOS_p$ values should ideally be completely monotonic in their relationship to the matching DMOS values [VQEG-Ob98]. The model should predict a change in $DMOS_p$ that has the same sign as the change in DMOS. Figures 6.4 and 6.5 show the hypothetical relationships between $DMOS_p$ and DMOS for two models of varying monotonicity. Both relationships have approximately the same prediction accuracy in terms of RMS error, but the model of Figure 6.4 has predictions that monotonically increase. The model in Figure 6.5 is less monotonic and falsely predicts a decrease in $DMOS_p$ for a case in which viewers actually see an increase in DMOS.

The Spearman rank-order correlation between $DMOS_p$ and DMOS is a sensitive measure of Monotonicity. [VQEG-Ob98].

Metric3: Spearman rank order correlation coefficient between $DMOS_p$ and DMOS.

A pair-wise comparison of pairs of HRC's on a scene by scene basis has also been proposed for examining the correlation between subjective preferences and objective preferences, and merits further investigation by the VQEG for inclusion in these tests.

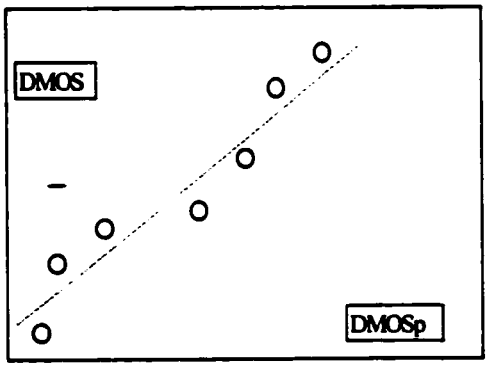


Fig 6.4 Model with more Monotonicity [VQEG-Ob98]

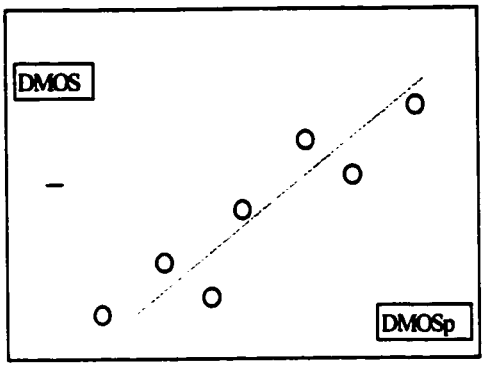


Fig 6.5 Model with less Monotonicity [VQEG-Ob98]

6.2.3 Metrics relating to Prediction Consistency of a model

This attribute means that the objective quality model's ability to provide consistently accurate predictions for all types of video sequences and not fail for a subset of sequences.

Figures 6.6 and 6.7 show models with approximately equal RMS errors between predicted and measured DMOS, but in Figure 6.6 it has accurate predictions for the majority of sequences but has large prediction error for the two points in the middle of the figure. In the second Figure 6.7 this model has a balanced set of prediction errors, although it is not as accurate as the first model, but it performs "consistently" by providing reasonable predictions for all the sequences. The model's prediction consistency can be measured by the number of outlier points, like having an error greater than a given threshold such as one confidence interval, [ITU500]. A smaller outlier fraction means the model's predictions are more consistent. Another metric that relates to consistency is Kurtosis [VQEG-ob98], which is a dimensionless quantity that relates only to the shape of the error distribution and not to the distribution's width. Two models may have identical RMS error, but the model with an error distribution having larger "tails" to the distribution will have a greater Kurtosis

Metric4: Outlier Ratio of "outlier-points" to total points N.

$$OutlierRatio = \frac{(Total\ number\ of\ Outliers)}{N}$$

where an outlier is a point for which:

$$Abs[Qerror[i]] \rangle 2 * DMOSS\tan dardError[i]$$

Twice the DMOS Standard Error is used as the threshold for defining an outlier point.

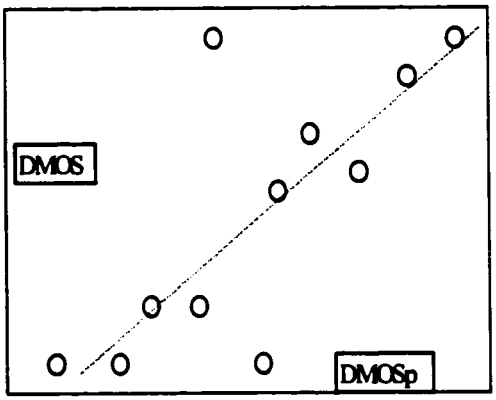


Fig 6.6 Model with large outlying errors [VQEG-Ob98]

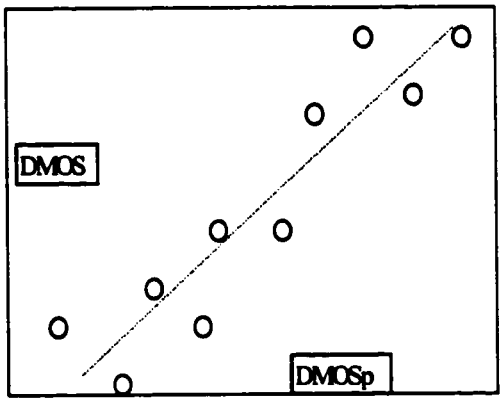


Fig 6.7 Model with consistent errors [VQEG-Ob98]

6.3 The VQEG Test and Conclusions:

The VQEG initially adopted these metrics and another one for the Analysis of variance ANOVA [VQEG00]. In the testing phase, they used software models on workstations like Sparc, silicon graphics and the streams were entered as a luminance and chrominance ITU601 component

stream on the algorithms, after being normalized and aligned in a special way [VQEG00, VQEG-ob98].

There were 2 other double-ended algorithms that reported poor results due to some technical problems. However, it's probable that these problems have been remedied afterwards with new releases of the algorithms.

The most important finding of the report was that there was "No conclusive Result"! The aim of VQEG was to get the best system performance. However, most of the algorithms gave very close results and correlation values (around 70s-80s) with no one to be a clear best. Another finding was that the different quadrants they used (525 line, 625 line, low and high quality sequences), results gave a wide range of values for the same model. Thus there was low consistency of results.

There was another observation in one of the correlation tables in the VQEG report [VQEG00 pp106] that for some models, the correlation for any 2 quadrants for a model gave better results when the values were combined to get the (all) data sets results. For example if we see the 50Hz (625 line) sequences correlation values and 60 Hz (525 line streams) correlation values, the correlation value for all the sequences should give a value somewhere in between the 2. However in a few models, it showed that it gave "Better" correlation. Thus the correlation value was sometimes affected by the number of points in the curve. And with more points they got better correlation rather than the more accurate one.

The other very interesting finding that the simplest model added, which was the PSNR reported remarkably good results overall (around 80%).

6.4 The Suggested In-Field Calibration standard

The excellent work done by groups like the VQEG has created a framework for development of such solutions and answers for broadcasters. And one solutions possible is to suggest a form of In-Field calibration performed by the TV networks or organizations involved in the transmission

and distribution of digital MPEG2 video to adopt. The [ITU-500] standard has suggested 2 levels of subjective tests to be done. One for the laboratory level and a lower less stringent level for home viewer level. The second one can be adopted by this standard to stand for the quality level expected in the customer's environment as their benchmark.

The most critical part in this new standard is the use of test sequences for the evaluation. At this age of hundreds of TV stations, the concept of specialized networks is becoming more dominant. Thus the material varies considerably between the following classes of networks:

1- Cartoon Networks

2- Drama Networks

3- News Networks

4- Sports Networks

5- Talk Show Networks ...etc.

The type of audience for these types of programming maybe different. Their judgment differs based on age group, psychological, mood and other factors that are not measured in a very straightforward manner.

The other but more relevant issue is that the complexity of spatial, temporal activity, luminance and chrominance complexity of these programs. The data rates for these kinds of programs are a critical factor for the quality.

One difference demonstrated by the VQEG report that the 50Hz vs. 60Hz programs may give different results, so this is another factor to be considered.

The other factor that was discounted as a variable, thus making the modeling a bit easier, was the type of MPEG2 encoders used to perform the compression of programs.

As mentioned before, video quality objective techniques are used for some applications including as one to evaluate the differences in CODECs and encoders performance. ITU-R BT.813 recommendation "Methods for Objective Picture Quality Assessment in Relation to Impairments from Digital Coding of Television Signals" [ITU-813] demonstrates the differences

in encoder types and testing methodologies. Thus this is another factor for TV networks to consider. Different networks will have different encoders, which yield different quality of video. As well, with time, those networks will change those encoders over time to newer technology encoders that will be more efficient in addressing critical areas within the video stream to give better data bandwidth in those areas, without exceeding the constraint of bit rate allocated. This would be in either constant bandwidth or statistical multiplexing [Kuhn00] systems.

Coombs [Coom72] states that: "It's natural that an echelon or hierarchy of standards will evolve if a measurement system is to show lineage or traceability to a common source". There will be standards for higher end reference entities like laboratory and R&D, manufacturers for example. And there will be lower working standards for operational units. In our case would be TV stations and networks.

Therefore a suggested calibration standard or validation for networks would include:

1- Subjective Testing of streams according to the condition of the Network or station (Data rates, test material, encoders)
2- Objective testing of the same streams by the model they have. The IFN/R&S in this case.
3- Adapting the results to create the correlation relationship.

This is shown in fig. 6.8. And if there was another double stimulus system used, the same is shown in fig. 6.9.

6.4.1. Subjective Measurement of quality for video sequences.

The choice of material should be a simulation of the regular programming material that this Network broadcasts. For example if it's a sports network, emphasis should be made on the different artifacts caused in the high spatial, temporal, luminance and chrominance fields in different sports, when constructing the HRCs and video sequences used.

The test should follow one of the standard listed in the ITU-R BT.500 (ITU-500 version 10 at the moment issued on March/2000), when performing this.

Following the second standard that addresses viewer rather than the lab test, may address the customers of network better.

For the continuous monitoring of video quality, if a single stimulus standard were followed, the duration of the sequences would be a critical issue. This should be chosen somewhere between 5 minutes and 10 seconds, which are the times for double and single ended tests respectively [ITU-500], if this system would be used for both single and double ended system.

The reading of values may need to be changed accordingly. At this time DSCQS, the measurement is taken once for the whole stream of 8 or 10 seconds, where as for single ended systems, the reading is taken from the evaluation unit twice per second. An optimum time is probably around 40 sec.- 1 minute. This way it would be long enough to do the single stimulus system and at the same time can offset any time lost in the beginning of the sequence when the system starts the measurement. It will have adequate results for the double stimulus systems as well.

A Network can perform these subjective tests at a number of facilities or independent labs available. There are many of those in each country or at least continent. This is probably the most expensive part of the test. As subjective tests take some days, and need quite a lot of resources and cost. However this will not be done on a regular basis. One set of such measurements may be used for years.

Of course when using subjective test sequences with HRCs representing the different artifacts and issues to be addresses, it should played from a digital machine that doesn't have compression inside like a D1, D2, D3 or D5 tape machines. The other solution is to use a digital disk recorder that doesn't have compression. The reason for that is that we shouldn't let the test itself affect the quality measurements.

6.4.2- Objective testing:

The same streams used could be run at the objective quality evaluation system under evaluation. This needs to be very simple by just running the streams from the tape or disk machine into the unit under test, and acquiring the results. This will be similar to the usual way they would get their quality monitoring results. Representing the correlation curve afterwards using one of the standard metrics would give the correlation for that system. A very simple correlation relation that can be used is the Pearson linear regression technique. Needless to say, the same kind of tape or disk machine should be used for this test as well to standardize the same conditions for both subjective and objective tests.

The process of calibration is shown in the following graphs. For single stimulus systems, which would be used for the continuous monitoring of quality, this is shown in fig. 6.8. Fig 6.9 shows the same for double-ended systems.

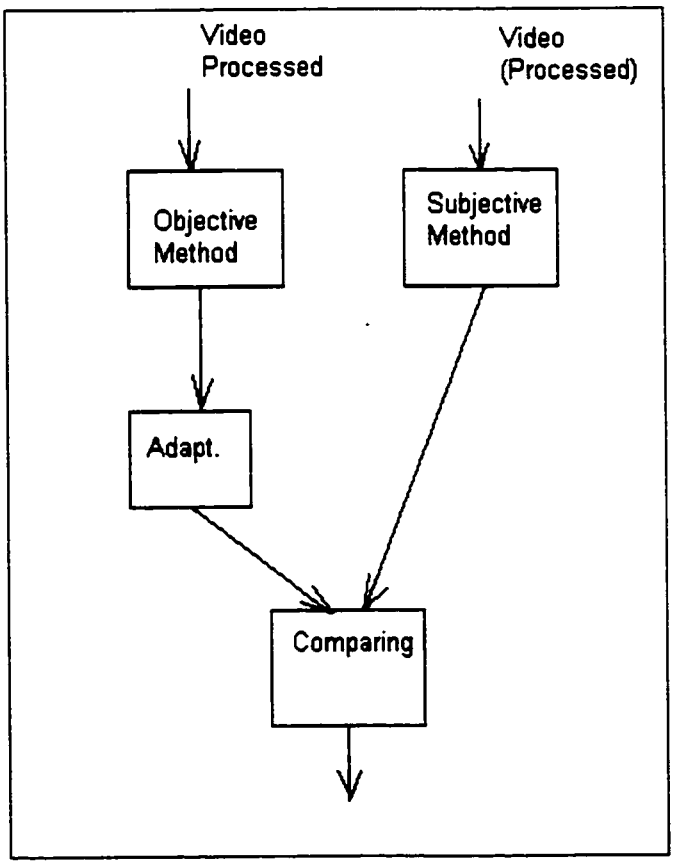


Fig 6.8 Single ended systems evaluation

This kind of procedure would be useful for documenting quality and what quality values obtained really mean in relation to subjective values. This knowledge could be helpful in a number of ways.

1- In case there is an update on the video quality system algorithm itself, and the network would run those objective tests on the updated system, and hopefully get a better correlation result.

2- In case the network have a number of quality evaluation systems, and need to get a common reference of the different quality values and scales used by different algorithms.

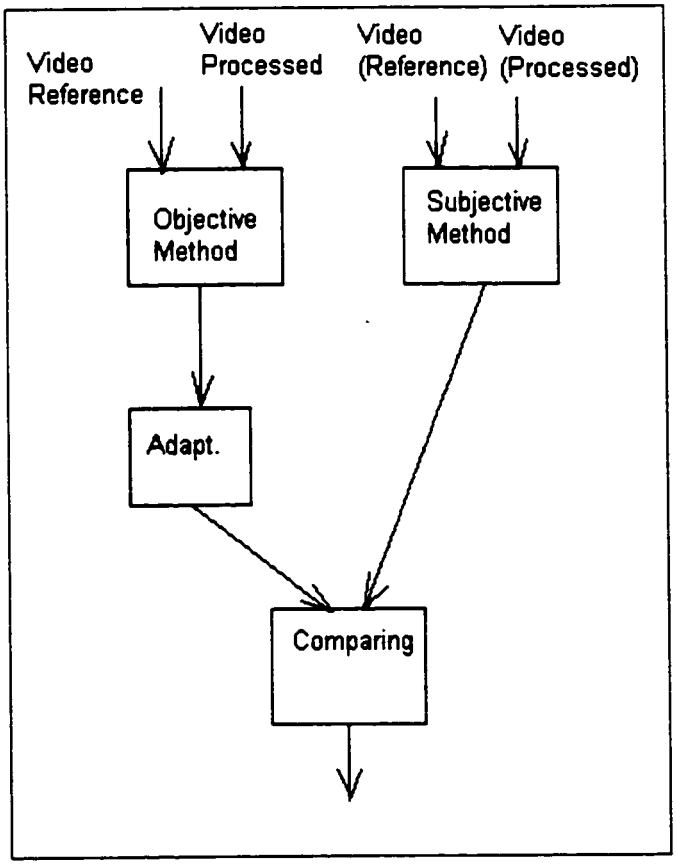


Fig 6.9 Double Ended systems evaluation

3- It can be expanded as well to include both single stimulus and double stimulus systems. In case the network has double stimulus system to do out of service test of the transmission links for example, and another single stimulus system to do the continuous online monitoring of quality.

4- In case there is a change in the condition of the network. That means if the network changes the type of encoders/decoders (CODECs) used for example. There maybe a need to re-design the test again to get the proper characterization of the network. However, in this case, there will be a need to re-create test sequences and artifacts simulating the new network.

5- To better characterize the performance of the network and closely follow similar standards put for analog systems performance. Such performance is usually important and could become included in such quality standards like ISO-9000.

6.5 Practical Experiment

As a demonstration of this simple technique, it would have been extremely expensive to do all the subjective tests to demonstrate this technique. Thus, used the already available subjective results from the VQEG report. I used a subset of the VQEG previous sequences with the permission and help of Communication Research Center (CRC-Canada) to get the results for the objective part of the experiment. CRC was one of the labs assigned with subjective testing during the VQEG tests.

6.5.1 Subjective Evaluation

I used the 60 Hz (525 lines) sequences that were encoded in MPEG2, and with digital type artifacts, and used the already available subjective measurement of those streams [VQEG00]. Thus this would simulate a TV station or network in North America (525 line TV system) with such material and set of artifacts represented in HRCs. The list of video sequences used and the corresponding encoding is shown in table 6.1. As well the set of HRCs used are listed in table 6.2.

Assigned number	Sequence	Characteristics	Source
13	Baloon-pops	Film, saturated color, movement	CCETT
14	NewYork 2	Masking effect, movement)	AT&T/CSELT
15	Mobile&Calendar	Available in both formats, color, movement	CCETT
16	Betes_pas_betes	Color, synthetic, movement, scene cut	CRC/CBC
17	Le_point	Color, transparency, movement in all the directions	CRC/CBC
18	Autumn_leaves	Color, landscape, zooming, water fall movement	CRC/CBC
19	Football	Color, movement	CRC/CBC
20	Sailboat	Almost still	EBU
21	Susie	Skin color	EBU
22	Tempete	Color, movement	EBU

Table 6.1. 525/60 format sequences

The first frame for all video sequences is enclosed in Appendix. A. They consisted a complete range of a typical TV station material. Each stream represented a type of material with certain critical factors to both DCT transform, and transmission. These characteristics included: color, movement, sharp building edges, waterfalls and water movement, skin color, animation and graphics.

ASSIGNED NUMBER	BIT RATE	Resolution	METHOD	COMMENTS
14	2 Mb/s	¾	mp@ml	This is horizontal resolution reduction only
13	2 Mb/s	¾	sp@ml	
12	4.5 Mb/s		mp@ml	With errors TBD
11	3 Mb/s		mp@ml	With errors TBD
10	4.5 Mb/s		mp@ml	
9	3 Mb/s		mp@ml	
5	8 & 4.5 Mb/s		mp@ml	Two codecs concatenated
2	19-19-12 Mb/s		422p@ml	3rd generation

Table 6.2. Test conditions (HRCs)

I have used 10 video streams of 60 Hz (525 line streams), with 8 Hypothetical Reference Circuit (HRCs), which were a subset of the VQEG streams. They were a mix of high quality and low quality from the VQEG definitions. All were MPEG2 compressed, but had different profiles and resolution. Their bit rates ranged from 2 Mbits/sec., and 19 Mbits/sec. As practical use for standard definition streams, this is a reasonable range for a full range of conditions for a TV station. Typical rates for broadcast are between 2 and 6 Mbits/sec. The combination of video stream and HRCs created about 80 streams, giving a decent impression of a station programming variety.

6.6. The test setup

I have used a very simple setup including:

1- A D-1 tape machine to playback the video streams. This is the same tape machine used for the subjective evaluation tests at the CRC.

2- A Sony 19" Monitor connected to the tape machine.

3- ITU-601 interface. This is the 270 Mbits/sec. interface for uncompressed video. It's common interface to most new tape machines with digital outputs, and many hard disk servers. The IFN/R&S system can accept either an ITU-601 uncompressed video or a compressed video in the ASI (Asynchronous Serial Interface) or SPI (Serial Parallel Interface) formats.

4- The IFN/R&S video quality unit (Single Stimulus Model).

5- A Computer Laptop with video quality software for the R&S/IFN system (called Quality Explorer Software).

6- RS-232 interface between the R&S/IFN system and the Laptop. This interface is used to acquire the results recorded from the unit in word and excel formats. The setup is shown in fig 6.10.

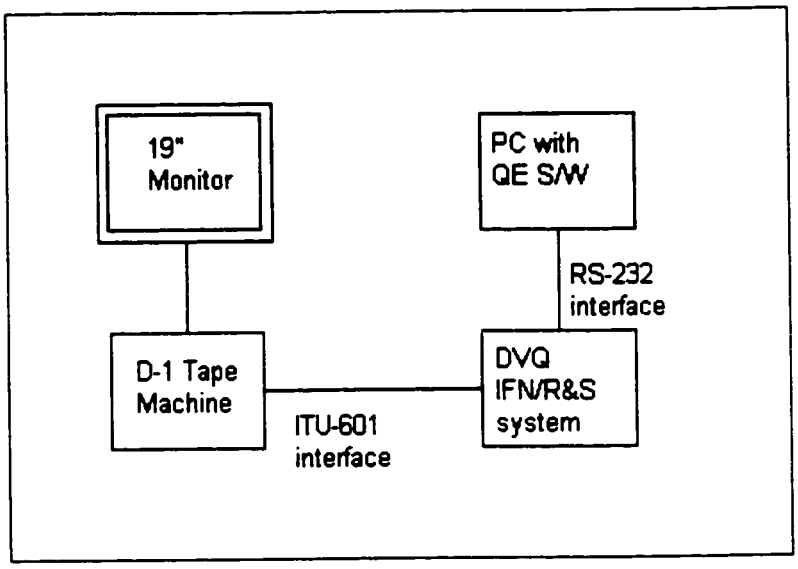


Fig 6.10 Test Setup for the Objective Measurement.

The test procedure was simple:

1- Start the D1 tape from the first frame of each test sequence of the Sequence/HRC combination (80 sequences).

2- Press the start button of the DVQ. This started the algorithm from the first step, which is calculating the picture resolution. After this step, which takes a few seconds, the algorithm is able to calculate the quality of the video sequence on a scale from 0-100.

3- Start the Quality Explorer (R&S) software on the computer to start acquiring the quality results. This is basically a curve of all measured values (twice per second) or once every 400 ms. As well, an excel sheet of all the measured values.

4- Obtain the values from the excel sheets, and compare them with the acquired graph, and obtain the average value per sequence. The results are shown in the Excel sheet of the results. There is one column representing the mentioned objective values. Another one for subjective values as per [VQEG00].

There was some alignment done at the end after obtaining the results. The reason was, as the stream ended, the tape would go to Grey. When that happened, the temporal activity went to 100%. Therefore, in some streams, I deleted values of the last one or 2

frames were the transition occurred for a 100% temporal change. So, the quality values were as best as possible corresponding to the measured video stream.

6.7 The test results and analysis

The results are shown in the enclosed figures. The Pearson linear regression was performed between subjective and objective values. The correlation obtained using linear regression was 76.54%. This correlation can be obtained very easily mathematically. I've tried also using polynomial regression to see the changes of correlation in these cases as per [VQEG-ob98]. The correlation changes a bit to 76.7% for a second order polynomial, 77.14% for a third order polynomial, and 77.48% for a 4^{th} order polynomial.

However, the linear regression is widely acceptable [VQEG00]. The result is reasonable, and correlates to the values obtained by Woerner and Lauterjung [Woe01] of 82.79%.

Furthermore, it's not exactly comparing apples to apples when considering the VQEG results, as they included analog & H-263 streams. HRCs included artifacts for analog artifacts, and others. However, looking at results in a more general way, correlations for those streams and artifacts ranged roughly from 70s to 80s.

Thus for MPEG2 streams and a sample of algorithms used in a typical TV station, this algorithm performs well, taking in consideration it's a single stimulus system.

There were other factors affecting the results, that should be considered in other Networks designing their test procedures, and eventually including in a standard:

1- The video streams were only 8 seconds long. Since this is a single stimulus system, it took a few seconds to calculate the resolution, and start calculating the quality. Therefore, the average number of values measured per stream were around 9 or 10 values. Almost the first half of the stream was not measured. Therefore, for future testing specifically designed for single stimulus systems, the streams should be longer than

that, as the first 3-4 seconds will be discarded. If the length of the stream is much higher, that will give a better confidence in the measurement values.

2- This testing was done only on a subset of the test sequences i.e. 60 Hz streams. The discarding of 50Hz streams may have changed the overall results a bit. However, in ordinary situations, a network would only be broadcasting one of those standards.

3- The test was performed in a manual fashion by 2 people. One person started the D1 tape machine. The other started the DVQ unit and software. The delay of start of stream may induce a small error factor as well. Again this error can be reduced when the streams are longer, and thus these milliseconds in the beginning will not affect the result. The other solution is in new software development, when one start from the software would initialize the unit, and stop on the condition of temporal activity reaching 100%.

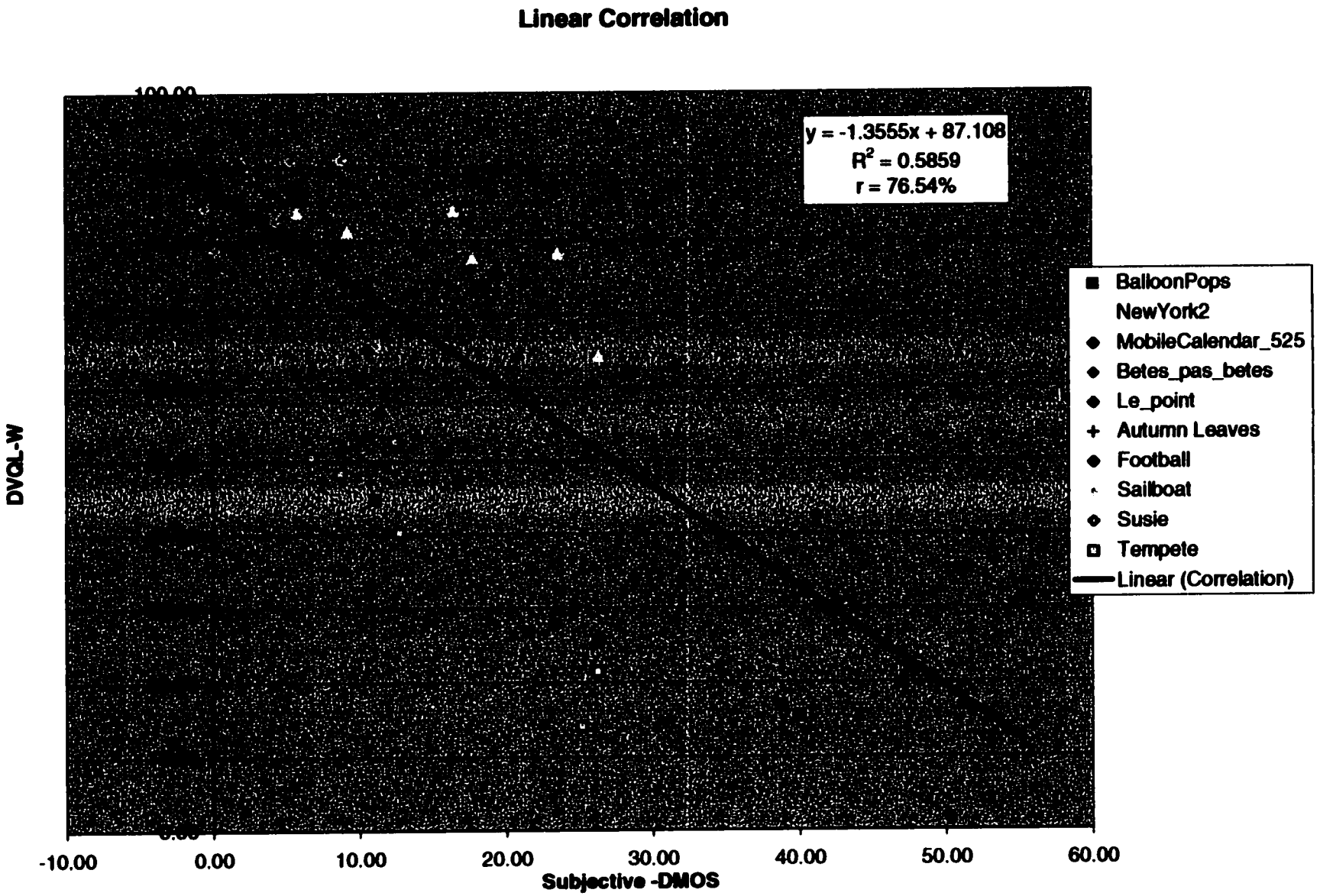
4- I had to disregard some values because the unit took almost the whole 8 seconds to calculate the resolution, so the measure of the last second or so didn't represent the whole stream quality value.

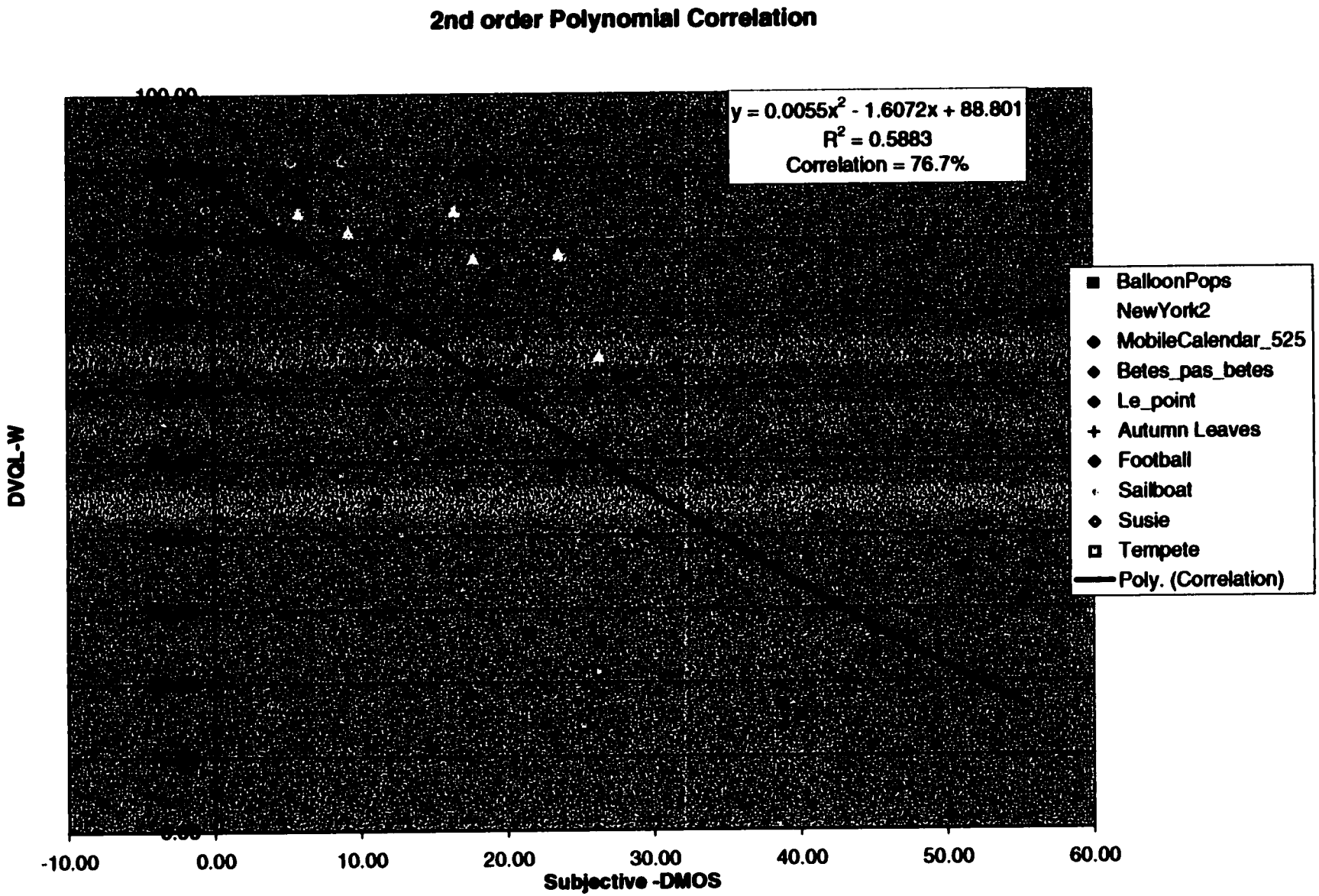
Correlation of Subective to Objective Measurement of Video Quality

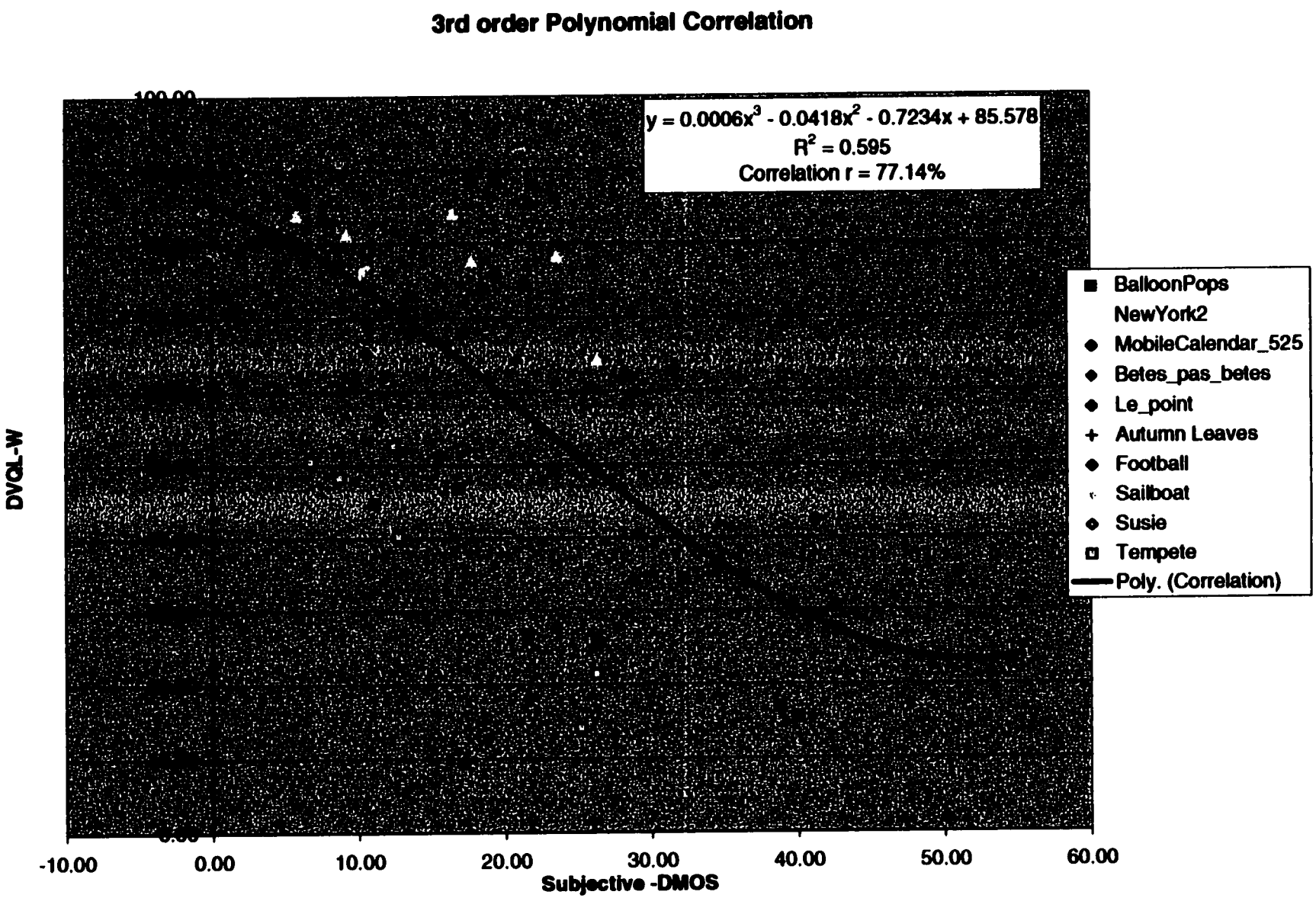
Sequence/HRC	source/HRC no.	Objective DVQL-W	DMOS Subjective
HRC2_BalloonPops	Src13/Hrc2	79.50	5.69
HRC5_BalloonPops	Src13/Hrc5	44.70	11.06
HRC9_BalloonPops	Src13/Hrc9	25.56	26.31
HRC10_BalloonPops	Src13/Hrc10	46.80	16.80
HRC11_BalloonPops	Src13/Hrc11	16.90	38.73
HRC12_BalloonPops	Src13/Hrc12	26.82	21.56
HRC13_BalloonPops	Src13/Hrc13	34.20	32.19
HRC14_BalloonPops	Src13/Hrc14	15.60	40.01
HRC2_NewYork2	Src14/Hrc2	90.20	2.16
HRC5_NewYork2	Src14/Hrc5	82.30	7.97
HRC9_NewYork2	Src14/Hrc9	81.30	9.25
HRC10_NewYork2	Src14/Hrc10	83.90	5.82
HRC11_NewYork2	Src14/Hrc11	77.60	17.76
HRC12_NewYork2	Src14/Hrc12	84.10	16.47
HRC13_NewYork2	Src14/Hrc13	64.20	26.37
HRC14_NewYork2	Src14/Hrc14	78.22	23.60
HRC2_525_MobileCalendar	Src15/Hrc2	88.90	17.76
HRC5_525_MobileCalendar	Src15/Hrc5	74.60	21.36
HRC9_525_MobileCalendar	Src15/Hrc9	60.55	25.95
HRC10_525_MobileCalendar	Src15/Hrc10	73.78	18.57
HRC11_525_MobileCalendar	Src15/Hrc11	61.90	28.59
HRC12_525_MobileCalendar	Src15/Hrc12	75.78	19.39
HRC13_525_MobileCalendar	Src15/Hrc13	31.44	55.29
HRC14_525_MobileCalendar	Src15/Hrc14	63.00	31.64
HRC2_Betes_pas_betes	Src16/Hrc2	92.90	2.34
HRC5_Betes_pas_betes	Src16/Hrc5	86.56	3.90
HRC9_Betes_pas_betes	Src16/Hrc9	86.89	5.57
HRC10_Betes_pas_betes	Src16/Hrc10	89.20	19.47
HRC11_Betes_pas_betes	Src16/Hrc11	80.40	14.04
HRC12_Betes_pas_betes	Src16/Hrc12	85.60	6.19
HRC13_Betes_pas_betes	Src16/Hrc13	73.00	13.74
HRC14_Betes_pas_betes	Src16/Hrc14	83.44	7.70
HRC2_Le_point	Src17/Hrc2	90.10	1.99
HRC5_Le_point	Src17/Hrc5	69.30	10.71
HRC9_Le_point	Src17/Hrc9	51.33	15.58
HRC10_Le_point	Src17/Hrc10	67.90	13.63
HRC11_Le_point	Src17/Hrc11	49.20	23.38
HRC12_Le_point	Src17/Hrc12	65.22	10.61
HRC13_Le_point	Src17/Hrc13	1.80	50.16
HRC14_Le_point	Src17/Hrc14	34.10	28.80

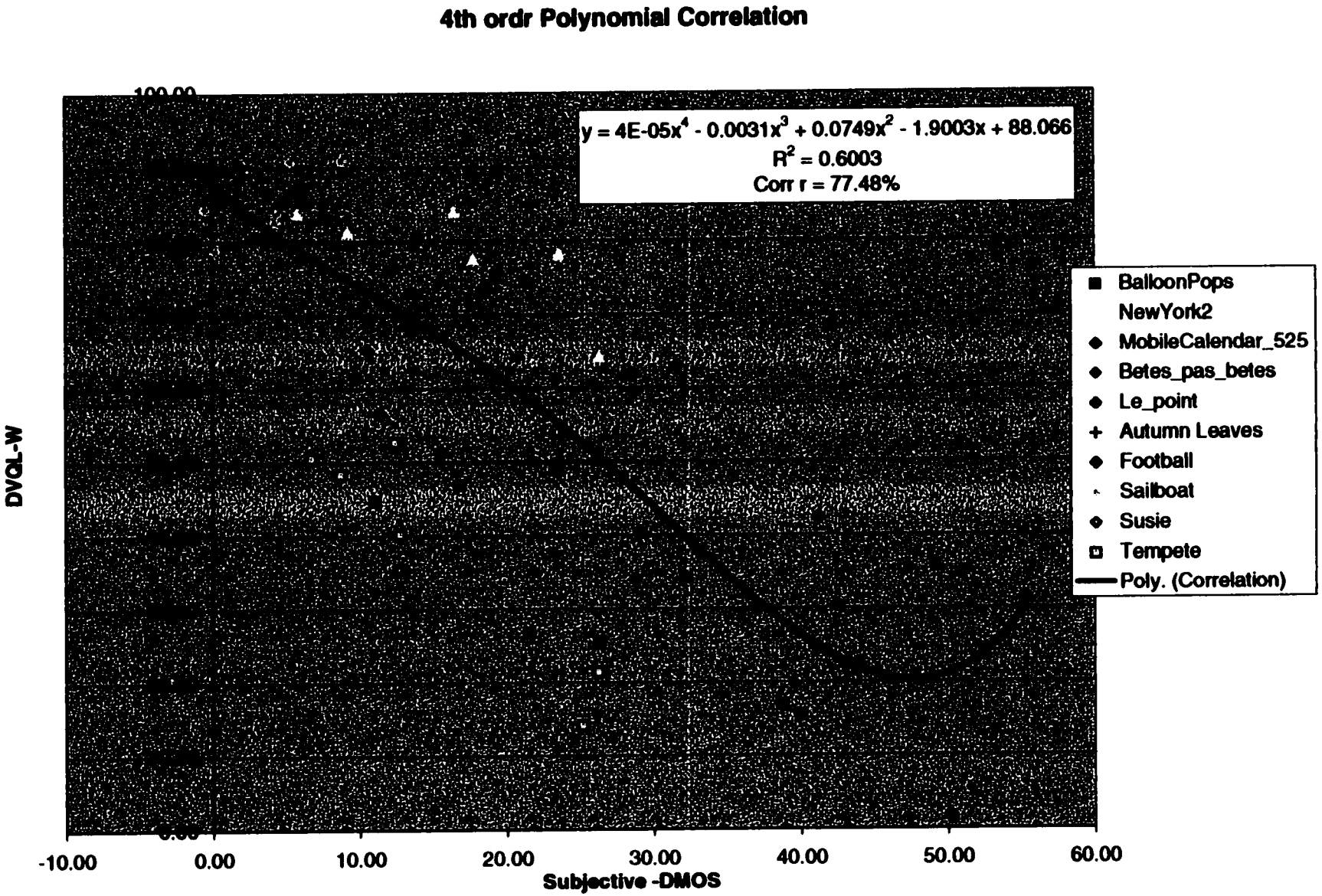
Sequence/HRC	source/HRC no.	Objective DVQL-W	DMOS Subjective
HRC2_AutumnLeaves	Src18/Hrc2	86.67	5.96
HRC5_AutumnLeaves	Src18/Hrc5	81.80	12.58
HRC9_AutumnLeaves	Src18/Hrc9	79.34	8.54
HRC10_AutumnLeaves	Src18/Hrc10	85.33	7.64
HRC11_AutumnLeaves	Src18/Hrc11	80.89	3.54
HRC12_AutumnLeaves	Src18/Hrc12	85.30	6.25
HRC13_AutumnLeaves	Src18/Hrc13	64.18	20.80
HRC14_AutumnLeaves	Src18/Hrc14	74.70	15.54
HRC2_Football	Src19/Hrc2	80.00	4.38
HRC5_Football	Src19/Hrc5	56.55	11.40
HRC9_Football	Src19/Hrc9	40.50	29.11
HRC10_Football	Src19/Hrc10	61.40	9.80
HRC11_Football	Src19/Hrc11	34.90	50.94
HRC12_Football	Src19/Hrc12	59.45	28.64
HRC13_Football	Src19/Hrc13	42.33	41.21
HRC14_Football	Src19/Hrc14	30.42	42.48
HRC2_Sailboat	Src20/Hrc2	84.44	-0.50
HRC5_Sailboat	Src20/Hrc5	82.70	4.29
HRC9_Sailboat	Src20/Hrc9	78.70	0.22
HRC10_Sailboat	Src20/Hrc10	90.80	5.36
HRC11_Sailboat	Src20/Hrc11	83.80	4.38
HRC12_Sailboat	Src20/Hrc12	91.00	8.80
HRC13_Sailboat	Src20/Hrc13	65.80	11.17
HRC2_Susie	Src21/Hrc2	89.60	6.41
HRC5_Susie	Src21/Hrc5	88.22	-0.66
HRC9_Susie	Src21/Hrc9	80.30	2.74
HRC10_Susie	Src21/Hrc10	88.80	-1.06
HRC11_Susie	Src21/Hrc11	76.10	12.23
HRC12_Susie	Src21/Hrc12	82.00	8.06
HRC13_Susie	Src21/Hrc13	75.40	3.30
HRC2_Tempete	Src22/Hrc2	72.30	4.31
HRC5_Tempete	Src22/Hrc5	48.30	8.74
HRC9_Tempete	Src22/Hrc9	50.50	6.75
HRC11_Tempete	Src22/Hrc11	40.20	12.76
HRC12_Tempete	Src22/Hrc12	52.67	12.44
HRC13_Tempete	Src22/Hrc13	14.11	25.19
HRC14_Tempete	Src22/Hrc14	21.50	26.25

-0.76545


Linear Correlation
y = -1.3555x + 87.108
R2 = 0.5859
r = 76.54%
DVQL-W
Subjective -DMOS
-10.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
BalloonPops
NewYork2
MobileCalendar_525
Betes_pas_betes
Le_point
Autumn Leaves
Football
Sailboat
Susie
Tempete
Linear (Correlation)


2nd order Polynomial Correlation
y = 0.0055x² - 1.6072x + 88.801
R² = 0.5883
Correlation = 76.7%
DVQL-W
Subjective -DMOS
-10.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
BalloonPops
NewYork2
MobileCalendar_525
Betes_pas_betes
Le_point
Autumn Leaves
Football
Sailboat
Susie
Tempete
Poly. (Correlation)


3rd order Polynomial Correlation
y = 0.0006x3 - 0.0418x2 - 0.7234x + 85.578
R2 = 0.595
Correlation r = 77.14%
DVQL-W
Subjective -DMOS
-10.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
BalloonPops
NewYork2
MobileCalendar_525
Betes_pas_betes
Le_point
Autumn Leaves
Football
Sailboat
Susie
Tempete
Poly. (Correlation)


4th ordr Polynomial Correlation
y = 4E-05x^4 - 0.0031x^3 + 0.0749x^2 - 1.9003x + 88.066
R^2 = 0.6003
Corr r = 77.48%
DVQL-W
100.00
-10.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
Subjective -DMOS
BalloonPops
NewYork2
MobileCalendar_525
Betes_pas_betes
Le_point
Autumn Leaves
Football
Sailboat
Susie
Tempete
Poly. (Correlation)

Chapter 7
Conclusion and Future Work

The objective of this Thesis was to study a generalization method of the ITU recommendations and VQEG work to develop a simple application for In-field calibration procedure for TV networks to validate the video quality subjective/objective correlation for an objective video quality monitoring system based on their individualized set of Network characterization.

A simulation of a typical TV network variety of material and set of different data rates and inserted artifacts allocated. This was a subset of 60Hz low and high quality MPEG2 compressed video programming material in different MPEG profiles. And made a decent simulation from the VQEG source material and HRCs used in their report [VQEG00], and also by Woerner and Lauterjung [Woer00]. The real subjective testing for networks would have to follow a different set of procedures described in chapter 5 to ensure the test follows a standard like the ITU-500 recommendation in a level corresponding to viewer or operational level as described in the recommendations.

The objective test used a simple setup as per fig 6.10.

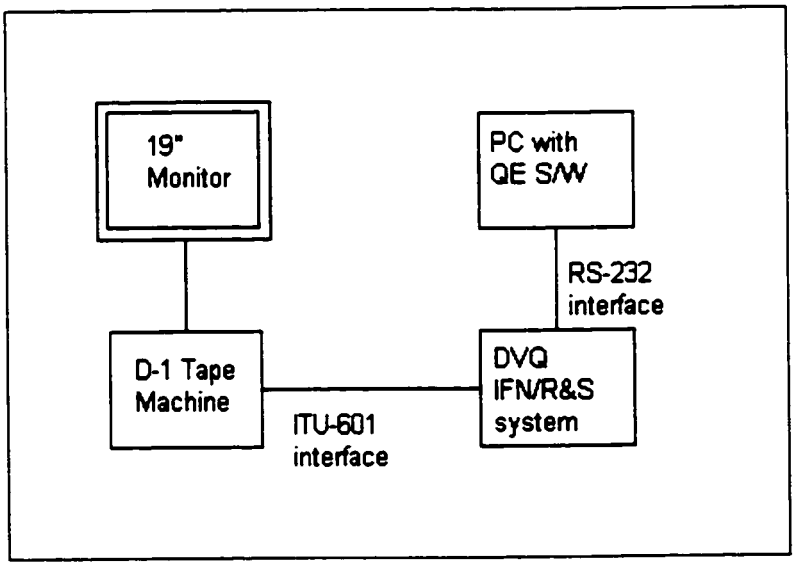


Fig 6.10 Test Setup for the Objective Measurement

Using the Pearson linear regression between subjective DMOS values and obtained objective predicted values would give the correlation relationship.

This technique will characterize the Network correlation based on it's specific characterization such as type of material, the encoders and decoders used, transmission links and set of expected artifacts in this specific case. The technique could be further expanded to characterize other quality evaluation systems used in the Network, to create an integrated quality system with known performance.

The wide spread of digital TV and MPEG2 compressed video as per the ATSC terrestrial standard in North America and the DVB standard in Europe and the rest of the world for terrestrial, cable and satellite, it becomes very important to start developing new standards for quality measurements, and create that "Echelon of Standards" for the whole industry. As the subjective measurement is the main reference of how we perceive video quality, all standards have to refer to that. Creating a standard that puts in consideration the individual circumstances and differential factors for each network is important to create a reference of quality of MPEG2 compressed video from the viewer perspective, which eventually is the real customer for Broadcasting.

There is a lot of work in the area of video quality and its' standards. A lot of work is done in the metrics for video quality and development of new metrics that perform better than earlier models. Current models are also enhancing their original models. Thus being able to calibrate any enhanced model as described, would be critical for the users (TV Networks) to see the difference in performance in a measured way.

An area like nonlinear approach to combine different features in a video stream to come up with video quality value is a particularly interesting field. Lin [Lin95] has used a back-propagation neural network for that. It's my personal view that using non-linear methods like Neural Network

and maybe Fuzzy Logic could open the door to customizing the quality evaluations "weighting" process even more to follow subjective evaluation even more closely. This is an application that Neural Network has good advantage in because of its' training process capability. Some TV networks for example need to calibrate that reading to their Golden eye evaluations, not regular viewers as per ITU-500 recommendation.

As per other research groups like VQEG and T1A1 in the USA, the VQEG will do similar round of video quality evaluations for reduced reference and no reference algorithms, since there has been a few developed in the past couple of years. Their work as mentioned earlier will help the ITU, SMPTE and other international bodies create more recommendations and standards in the field.

There is further work in the development of encoders to utilize low bandwidth data rate and get better quality.

The other area that is opening a huge field is the spread of HDTV broadcast, and the need to develop quality models and metrics for those formats. At the moment, trying to process uncompressed decoded Standard Definition video in the data rates of 270 Mbits/sec. has been difficult. Thus going the next step of uncompressed video streams of 1.55 Gbits/sec. is still quite challenging. However, there are efforts in this area at the moment.

The other area that needs to be addressed more, and there are efforts in, is the analysis of audio quality embedded in MPEG2 streams. Especially going to Dolby-AC3 in the ATSC, there are 6 audio channels (referred to as 5.1 channels), and getting into such analysis will be important.

8- References:

1- Ara01 — Arabatti, A.; Myler, H.; "An Objective Quality Measure Based on Subjective Measure for JPEG Compression"; Teranex Inc, Orlando, FL, USA, VQEG meeting, Orlando, FL, July/2001

2- Bai01 — Baina, J.; Bretillon, P.; Goudezeune, G.; "Reference Model for Quality Meter Description"; TDF, France, VQEG meeting, FL, USA, July/2001

3- - Bass96 — Basso, A.; Dalgic, I.; Tobagi, F.; and Van den Brander Lambrecht, C: "Study of MPEG-2 coding performance based on perceptual quality metric"; in PCS'96 International Picture Coding Symposium, vol. 1, pp. 263-268, Mar. 1996

4- Bee94 — Beerends, J.G.; "Modelling a cognitive effects that play a role in the perception of speech quality, in Speech Quality Assessment, Workshop papers, Bochum, pp1-9, November, 1994

5- Bee97 — Beerends, J. G., "Objective Measurement of Video Quality", KPN research, Netherlands, ITU study group 12-contribution 7, Feb./1997

6- Bha97 — Bhaskaran, V. & Konstantinides, K.: "Image and Video Compression Standards" – Algorithms and Architectures – Second Edition, Kluwer Academic Publishers, Norwell, Mass., USA, 1997

7- Bru99 — Bruin, R.; Smits, J., "Digital Video Broadcasting; Technology, Standards and Regulations", Artech House, Norwood, MA, USA, 1999

8- Burt83 — Burt, P.J.; Adelson, E.; "The Laplacian pyramid as a compact image code"; IEEE Transactions on Communications, COM-31, 532-540, 1983

9- Car96 — Carney, T.; Klein, S.; and Hu, Q; "Visual masking near spatiotemporal edges", in Human Vision and Electronic Imaging (B.E. Rogowitz and J. Allebach, eds.) vol. 2657, pp393-402, SPIE, 1996

10- Cas96 — Castleman, K. R., "Digital Image Processing", Prentice Hall Inc., New Jersey, USA, 1996

11- Cav01 — Caviedes, J.; Jung, J.;"No Reference Metric for a Video Quality Control Loop"; Philips Research, NY,USA/ Philips France, VQEG meeting, Orlando, FL, July/2001

12- Coom72 — Coombs, C. F.; "Basic Electronic Instrument handbook";

Mcgraw-Hill Book Co., USA, 1972

13- Daly93 — Daly, S; "The visible differences predictor, an algorithm for the assessment of image fidelity"; in Digital Image and Human Vision (A. Watson, ed.), ch 7, pp 179-208, MIT Press, Cambridge, Massachusetts, 1993

14- Eff98 — Effelsberg, W.; Steinmetz, R., "Video Compression Techniques", dpunkt-verlag, Heidelberg, Germany 1998

15- Eva95 — Evans, B., "Understanding Digital TV; The route to HDTV", IEEE press, NY, USA, 1995

16- Fen98 — Fenimore, C.; Libert, J.; Wolf, S.; "Perceptual Effects of Noise in Digital Video Compression"; NTIA, CO, USA, SMPTE meeting, Oct 28-31/1998

17- Fib87 — Fibush, D. K., " A Guide to Digital Television Systems and Measurements", Tektronix Inc., USA, 1987.

18- Fis95 — Fisher, Y., "Fractal Image Compression", Springer-Verlag, NY, USA, 1995

19- Fra98 — Fanzen, N. et al, " Verification of Normalization Processing for VQEG Video Test Data", Tektronix Measurement Div., 18, May/1998

20- Ham00 — Hamada, T., "Video Codec Evaluation Scheme and Implementation Based on Characteristics of Human Visual Perception", KDD R&D laboratories, Japan, 2000

21- Ham00_2 — Hamada, T., "Proposal for an Object Based Model Considering Object Complexity", KDD Labs/Pixelmetrix, Japan, VQEG meeting, Ottawa, March/2000

22- Ham01 — Hamada, T.; "Digital Video/Audio NR+RR Monitoring System Based on Motion Compensated Interframe/ Intraframe Objective Parameters"; KDD Corp., Japan, VQEG meeting, Orlando, FL, USA, July/2001

23- Ham97 — Hamada, H; Namba, S; "Objective Picture Quality Scale for Digital Compressed Picture for Broadcast"; NHK Research Labs.; VQEG meeting, Oct./1997

24- Has97 — Haskell, B. G., "Digital Video: an introduction to MPEG-2", Chapman and Hall, USA, 1997

25- ITU-500 — ITU-R Recommendation BT.500-10: "Methodology for the

Subjective Assessment of the Quality of Television Pictures", ITU, geneva, Switzerland, March/2000

26- ITU-813 ITU-R Recommendation BT.813: "Methods for Objective picture quality assessment in relation to impairments from digital coding of Television Signals", ITU, Geneva, Switzerland, 1992

27 - Jay93 Jayant, N.; Johnston, J.; and Safranek, R.; "Signal compression based on models of human perception"; Proceedings of the IEEE, vol. 81, pp. 1385-1422, Oct. 1993

28 - Katt91 Katto, J.; Onda, K.; and Yasuda, Y.; "Variable bit-rate coding based on human visual system"; Signal Processing: Image Communication, vol. 3, pp. 321-331, Sept. 1991

29- Kaw01 Kawada, R; "A method for estimating PSNR of coded pictures by using of embedded invisible markers"; KDDI Labs Inc., Japan, VQEG meeting, Boulder, CO, May/2001

30- Klein92 Klein, S.; Silverstein, A; Carney, T.; "Relevance of human visionto JPEG-DCT compression"; in Human Vision, Visual Processing, and Digital Display III (B.E. Rogowitz, ed.), vol 1666, pp. 200-215, SPIE, 1992

31- Kne01 Knee, M, "A Single-Ended Picture Quality Measure for MPEG-2", Snell & Wilcox, UK, VQEG meeting, Orlando, FL., USA, July 2001

32- Kuh00 Kuhn, M.; Antkowiak, J.;" Statistical Multiplex - what does it mean for DVB-T", FKT Fachzeitschrift fur Fernsehen, Film und elektronische Medien, Germany- April/2000

33- Lamb96 Van den Branden Lambrecht, C; " A working spatio-temporal model of the human visual system for image restoration and quality assessment applications"; in International Conference on Acoustics, Speech and Signal Processing ICASSP' 96, 1996

34- Lau98 Lauterjung, J.; "Picture Quality Measurement"; Rohde & Schwarz GmbH, Munich, Germany. Proceedings of the International Broadcasting Convention IBC 1998, Amsterdam, Netherlands, 1998

35- Lau99 Lauterjung, J. 1999; "First Results of digital video quality measurement in DVB networks. Proceeding of the Radio Show, Montreux, Switzerland, 1999

36- Lin95 Lin, F.; Mersereau, R; "A constant subjective quality MPEG encoder"; in International Conference on Acoustics, Speech and Signal Processing ICASSP'95, pp. 2177-2180, 1993

37- Lub00 Lubin, J.; "The Use of Psychophysical Data and Models in the Analysis of Display System Performance"; Sarnoff Corp., USA, VQEG meeting, Munich, Germany, June/2000

38- Lut99 Luther, C.; Inglis, A., "Video Engineering", McGraw Hill Inc., 3rd edition, USA 1999.

39- Mau92 Mausl, R., "Refresher topics – Television technology", Rohde & Schwarz GmbH, Germany 1992

40- Meer97 Van der Meer, P., "Variable Bit Rate Compressed Video", Delft Univ. Press, Delft, Netherlands, 1997

41- Miya92 Miyahara, M.; Kotani, K; & Algazi, V.; "Objective Picture Quality Scale (PQS) for image coding" SID digest, pp859-862, 1992

42- Nach74 Nachmias, J.; Sansbury, R.; "Grating contrast: Discrimination may be better than detection"; Vision Res. 14, 1039-1042, 1974

43- Nish00 ITU, study group 9;"Objective Picture Quality Evaluation For the Digital Broadcasting By Introduction of Criticality"; Geneva, May 15-19, 2000

44- Nish97 Nishida, Yukihiro; "Objective Picture Quality Evaluation For the Digital Broadcasting By Introduction of Criticality"; NHK Research Labs, Japan, VQEG meeting 14-16 Oct./1997

45- Olz86 Olzak, L.; Thomas, J.; "Seeing spatial pattern", in Handbook of perception and Human Performance (K. Boff, L. Kaufman, and J. Thomas eds.), ch.7, Wiley, New York, 1986.

46- Pan99 Panasonic Technology Inc., "The Video Compression Book", 1999

47- Pess97 Pessoa, A; ITU-T Study Group 12, Contribution12-39, "Video Quality Assessment using objective parameters based on image segmentation"; December 1997

48- Pete91 Peterson, H.; Peng, H.; Morgan, J. and Watson, A.;" Quantization of color image components in the DCT domain"; in Human Vision, Visual Processing and Digital Display II (B.E. Rogowitz, M. Brill, and J. Allebach, eds.), vol. 1453, pp. 210-222, SPIE, 1991

49- Pete92	Peterson, H.; "DCT basis functions visibility in RGB space"; in Society for Information Display Digest of technical Papers (J. Morreale, ed.), Society for Information Display, 1992
50- Poy96	Poynton, C.; "A Technical Introduction of Digital Video", John Wiley & Sons Inc., USA 1996.
51- Rob98	Robin, M.; Poulin, M., "Digital Television Fundamentals: Design and installation of video and audio systems", McGraw Hill, USA 1998.
52- Rob99	Robin, M., "Testing MPEG compressed Video", December/1999 issue, page 80-88, Broadcast Engineering, USA, Dec./1999.
53- Sag89	Saghri, J; Cheatham, P; Habibi, A; "Image Quality measure based on a human visual system model"; Optical Engineering, vol 28, pp. 813-818, July 1989.
54- Sar97	Sarnoff Corp., "Sarnoff JND Vision Model Algorithm Description and Testing", Princeton, NJ, USA, August 4, 1997
55- Saw99	Saw, Y., "Rate-Quality Optimized Video Coding", Kluwer Academic Publishers, Mass. USA, 1999
56- Sug01	Sugimoto, O. et al; "Automatic Objective Picture Quality Measurement Method using invisible Markers Signals as Reduced References", KDDI R&D Labs Inc., Saitama, Japan, VQEG meeting, FI, USA, July/2001
57- Sym98	Symes, P., "Video Compression", McGraw Hill, USA 1998
58- Tan01	Tan, K.T.; Ghanbari, M; "Blockiness Measurement for MPEG Video";Singaore Polytechnic/ Univ. of Essex, VQEG meeting, Orlando, FL, July/2001
59- Tek97	Tektronix Inc., "A Guide to Picture Quality Measurements for Modern Systems", 1997.
60- Tek98	Tektronix Inc., "A Guide to MPEG Fundamentals and Protocol Analysis (Including DVB and ATSC)", 1998.
61- Vass73	Vassilev, A; "Contrast Sensitivity near borders: Significance of test stimulus form, size and duration", Vision Research vol. 13, pp. 719-730, 1973
62- Vor92	Voran, S.; "The Effect of Multiple Scenes on Objective Video

Quality Assessment", NTIA, CO, USA, T1A1.5 meeting, Document# T1A1.5/92-136, July 13/1992

63- VQE00 VQEG, ITU, "Final Report from the Video Quality Expert Group on the Validation of Objective Models of Video Quality Assessment", March 2000

64- VQEG- ob98 ITU study group 12; "Evaluation of new methods for objective testing of video quality: objective test plan"; 1998

65- Wat99 Watkinson, J., "MPEG-2" , focal press, UK, 1999

66- Wes95 Westen, S.; Lagendijk, R.; Beimond J.; "Perceptual Image Quality based on a multiple channel HVS model"; in International Conference on Acoustics, Speech and Signal Processing ICASSP' 95, vol 4, pp 2351-2354, May 1995

67- Wes97 Westwater, R. & Furht, B., "Real-Time Video Compression – Techniques & Algorithms", Kluwer Academic Publishers, Norwell, Mass., USA, 1997

68- Wink99 Winkler, S.; "A Perceptual Distortion Metric for Digital Color Video", Human Vision and Elctronic Imaging IV, Proceedings Volume 3644, B.E. Rogowitz and T.N. Pappas eds., pages 175-184, SPIE, Bellingham, WA, 1999

69- Woe00 Woerner, A., "DVQ- An objective picture quality measurement not requiring a reference (NR)", Rohde & Schwarz GmbH, T1A1.1/2000-044 committee meeting, Charleston, Jan 29th, 2001

70- Woe01 Woerner, A.; Lauterjung, J; "A real time single ended algorithm for objective quality monitoring of compressed video signals", Rohde & Schwarz GmbH, Munich, Germany, VQEG meeting, Orlando, FL, USA, July, 2001

71- Wolf92 Wolf, S.; "An Automated Technique for Measuring Transmitted Frame Rate (TFR) and Average Frame Rate (AFR)", NTIA, CO, USA, T1A1.5 meeting, July, 13/1992

72- Wolf97 Wolf, S.; et al, "Objective and Subjective Measure of MPEG Video Quality"; NTIA, CO, USA, SMPTE meeting, Nov 21-24/1997

73- Wolf97 Wolf, S.; Webster, A.; "Subjective and Objective Measures of Scene Criticality", NTIA/ITS, Boulder, CO, USA, ITU meeting, Turin Italy Oct./1997

74- Wolf98 Wolf, S.; Pinson, M.; "In-Service Performance Metrics for

MPEG-2 Video Systems"; NTIA, CO, USA, IAB, Nov. 12-13/1998

75- Wolf99 Wolf, S.; Pinson, M.; "Spatial-Temporal distortion metrics for in-service quality monitoring of any digital video system"; NTIA, Boulder, CO., USA, Spie meeting, Sept, 11/1999

9- Acronyms and Glossary:

16 VSB	Vestigial sideband modulation with 16 discrete amplitude levels.
8 VSB	Vestigial sideband modulation with 8 discrete amplitude levels. VSB is an analog modulation technique used to reduce the amount of spectrum needed to transmit information through cable TV, or over-the-air broadcasts used in the NTSC (analog) standard. 8-VSB is the U.S. ATSC digital television transmission standard.
AES	Audio Engineering Society.
Aliasing	Defects or distortion in a television picture or audio. Defects are typically seen as jagged edges on diagonal lines and twinkling or brightening. In digital video, aliasing is caused by insufficient sampling or poor filtering of the digital video.
Anchor frame	A video frame that is used for prediction. I-frames and P-frames are generally used as anchor frames, but B-frames are never anchor frames.
ANOVA	ANalysis Of Variance
ANSI	American National Standards Institute.
Artifacts	Undesirable elements or defects in a video picture. Most common in digital are macroblocks, which resemble pixelation of the video image, and pops and clicks in audio.
ASI	Asynchronous serial interface
Aspect Ratio 16:9	Aspect ratio of widescreen DTV formats for all HDTV and some SDTV (Standard Definition) video. "16" unit width corresponds to "9" unit height, proportionally, regardless of the actual size of the screen.

Aspect Ratio 4:3	Aspect ratio of the NTSC TV screen, with "4" unit width corresponding to "3" unit height, proportionally, regardless of the actual size of the screen.
ATM	Asynchronous Transfer Mode. A digital signal protocol for efficient transport of both constant-rate and bursty information in broadband digital networks. The ATM digital stream consists of fixed-length packets called "cells," each containing 53 8-bit bytes—a 5-byte header and a 48-byte information payload.
ATSC	Advanced Television Standards Committee
Bit Rate	The rate at which the compressed bit stream is delivered from the channel to the input of a decoder.
Block	A block is an 8-by-8 array of pixel element values or DCT coefficients representing luminance or chrominance information.
B-pictures	Bidirectional picture. Pictures that use both future and past pictures as a reference. This technique is termed bidirectional prediction. B-pictures provide the most compression. B-pictures do not propagate coding errors as they are never used as a reference.
Byte-aligned	A bit in a coded bit stream is byte-aligned if its position is a multiple of 8-bits from the first bit in the stream.
CAT	Conditional access table
CBR	Constant Bit Rate, Operation where the bit rate is constant from start to finish of the compressed bit stream.
CCIR	Comite Consultatif International des Radiocommunications
cd/m2	Candles per square meter. A unit for measuring luminance.
CDTV	Conventional Definition TV. This term is used to signify the

	analog NTSC television system as defined in ITU-R Recommendation 470. See also standard definition television and ITU-R Recommendation 1125.
Channel	A digital medium that stores or transports a digital television stream.
CIF	Common Intermediate Format: 288 lines X 352 pixles/ line (Luminance), 144 lines X 176 pixels/line (Chrominance)
Codec	Coder-decoder. A device that converts analog video and audio signals into a digital format for transmission. Also converts received digital signals back into analog format.
COFDM	Coded orthogonal frequency division multiplexing. COFDM can transmit many streams of data simultaneously, each one occupying only a small portion of the total available bandwidth. The DTV standard used in Europe.
Compression	Reduction in the number of bits used to represent an item of data.
CRC	Communications Research Center (Canada)
DCT	Discrete Cosine Ttransform; A mathematical transform that can be perfectly undone and which is useful in image compression.
DDR	Digital Disk Recorder. A video recording device that uses a hard disk drive or optical disk drive mechanism. Disk recorders offer nearly instantaneous access to recorded material.
Decoded Stream	The decoded reconstruction of a compressed bit stream.
Decoder	An embodiment of a decoding process.
Decoding (process)	The process defined in the Digital Television Standard that reads an input coded bit stream and outputs decoded pictures or audio samples.
D-frame	Frame coded according to an MPEG-1 mode which uses

	DC coefficients only.
Digital Betacam	A development of the original analog Betacam which records digitally on a Betacam-style cassette by Sony.
DMOS	Differential Mean Opinion Score
DMOSp	Differential Mean Opinion Score – predicted
Dolby AC-3	Digital Dolby. The approved 5.1 channel (surround-sound) audio standard for ATSC digital television, using approximately 13:1 compression. Six discreet audio channels are used: Left, Center, Right, Left Rear (or side), Right Rear (or side), and a subwoofer -- LFE, "low frequency effects" -- (considered the ".1" as it is limited in bandwidth).
DSCQS	Double Stimulus Continuous Quality Scale
DTS	Decoding time stamp
DVB-C	Digital Video Broadcasting-Cable
DVB-S	Digital Video Broadcasting-Satellite
DVB-T	Digital Video Broadcasting-Terrestrial
DVCR	Digital video cassette recorder
EIT	Event information table
Elementary stream (ES)	A generic term for one of the coded video, coded audio or other coded bit streams. One elementary stream is carried in a sequence of PES packets with one and only one stream_id.
Encoder	An embodiment of an encoding process.
Encoding (process)	A process that reads a stream of input pictures or audio samples and produces a valid coded bit stream as defined in the Digital Television Standard.
Entropy coding	Variable length lossless coding of the digital representation of a signal

	to reduce redundancy.
Event	An event is defined as a collection of elementary streams with a common time base, an associated start time, and an associated end time.
Field	For an interlaced video signal, a "field" is the assembly of alternate lines of a frame. Therefore, an interlaced frame is composed of two fields, a top field and a bottom field.
FR	Full Reference
Frame	A frame contains lines of spatial information of a video signal. For progressive video, these lines contain samples starting from one time instant and continuing through successive lines to the bottom of the frame. For interlaced video a frame consists of two fields, a top field and a bottom field. One of these fields will commence one field later than the other.
GOP	Group of pictures; A group of pictures consists of a sequence of I, B and P frames
HDTV	High definition television has a resolution of approximately twice that of conventional television in both the horizontal (H) and vertical (V) dimensions and a picture aspect ratio (HxV) of 16:9. ITU-R Recommendation 1125 further defines "HDTV quality" as the delivery of a television picture which is subjectively identical with the interlaced HDTV studio standard.
High Level	A range of allowed picture parameters defined by the MPEG-2 video coding specification which corresponds to high definition television.
HRC	Hypothetical Reference Circuit
Huffman coding	A type of source coding that uses codes of different lengths to represent symbols which have unequal likelihood of occurrence.
I Frames	Pictures that are coded using information present only in the picture itself

	and not depending on information from other pictures. I-pictures provide a mechanism for random access into the compressed video data. I-pictures employ transform coding of the pel blocks and provide only moderate compression.
IEC	International Electrotechnical Commission.
IRT	Institut Rundfunk Technische (Germany)
ISO	International Organization for Standardization.
ITU	International Telecommunications Union
JEC	Joint Engineering Committee of EIA and NCTA.
JND	Just noticeable difference. Quality system by Sarnoff labs.
Level	A range of allowed picture parameters and combinations of picture parameters.
Macroblock	The four 8 by 8 blocks of luminance data and the two (for 4:2:0 chroma format), four (for 4:2:2 chroma format) or eight (for 4:4:4 chroma format) corresponding 8 by 8 blocks of chrominance data coming from a 16 by 16 section of the luminance component of the picture. Macroblock is sometimes used to refer to the pel (Pixel Elements) data and sometimes to the coded representation of the pel values and other data elements defined in the macroblock header.
Mbps	1,000,000 bits per second.
MOS	Mean Opinion Score
MOSp	Mean Opinion Score, predicted
Motion vector	A pair of numbers which represent the vertical and horizontal displacement of a region of a reference picture for prediction.
MP@HL	Main profile at high level.
MP@ML	Main profile at main level.
MPEG-1	Refers to ISO/IEC standards 11172-1 (Systems), 11172-2 (Video), 11172-3 (Audio), 11172-4 (Compliance Testing), and 11172-5 (Technical Report).

MPEG-2	Refers to ISO/IEC standards 13818-1 (Systems), 13818-2 (Video), 13818-3 (Audio), 13818-4 (Compliance).
MSE	Mean squared error
NIT	Network Information Table
NR	No (or Zero) Reference)
NTSC	National Television Standard Code. 525 line color system.
Packet	A packet consists of a header followed by a number of contiguous bytes from an elementary data stream. It is a layer in the system coding syntax.
Padding	A method to adjust the average length of an audio frame in time to the duration of the corresponding PCM samples, by continuously adding a slot to the audio frame
PAL	Phase Alternating Line color TV system.
PAT	Program associated table
Payload	Payload refers to the bytes which follow the header byte in a packet. For example, the payload of a transport stream packet includes the PES_packet_header and its PES_packet_data_bytes or pointer_field and PSI sections, or private data. A PES_packet_payload, however, consists only of PES_packet_data_bytes. The transport stream packet header and adaptation fields are not payload.
PCR	Program Clock Reference; A time stamp in the transport stream from which decoder timing is derived.
PCR	Program clock reference
PEM	Pooling of error means
PES	Packetized elementary stream.
PES Stream	A PES stream consists of PES packets, all of whose payloads consist of data from a single elementary stream, and all of which have the

	same stream_id.
Picture	Source, coded or reconstructed image data. A source or reconstructed picture consists of three rectangular matrices representing the luminance and two chrominance signals.
PID	Packet identifier; A unique integer value used to associate elementary Streams of a program in a single or multi-program transport stream.
Pixel	"Picture element" or "pel." A pixel is a digital sample of the color intensity values of a picture at a single point.
P-pictures	Pictures that are coded with respect to the nearest previous I or P-picture. This technique is termed forward prediction. P-pictures provide more compression than I-pictures and serve as a reference for future P-pictures or B-pictures. P-pictures can propagate coding errors when P-pictures (or B-pictures) are predicte from prior P-pictures where the prediction is flawed.
PQR	Picture quality rating
Profile	Predicted Pictures. A defined subset of the syntax specified in the MPEG-2 video coding specification
Program	A program is a collection of program elements. Program elements may be elementary streams. Program elements need not have any defined time base; those that do have a common time base and are intended for synchronized presentation.
PS	Program Segment
PSI	Program Specific Information; PSI consists of normative data which is necessary for the demultiplexing of transport streams and the successful regeneration of programs.
PSIP	Pronounced "P-SIP" - "Program and system information protocol."

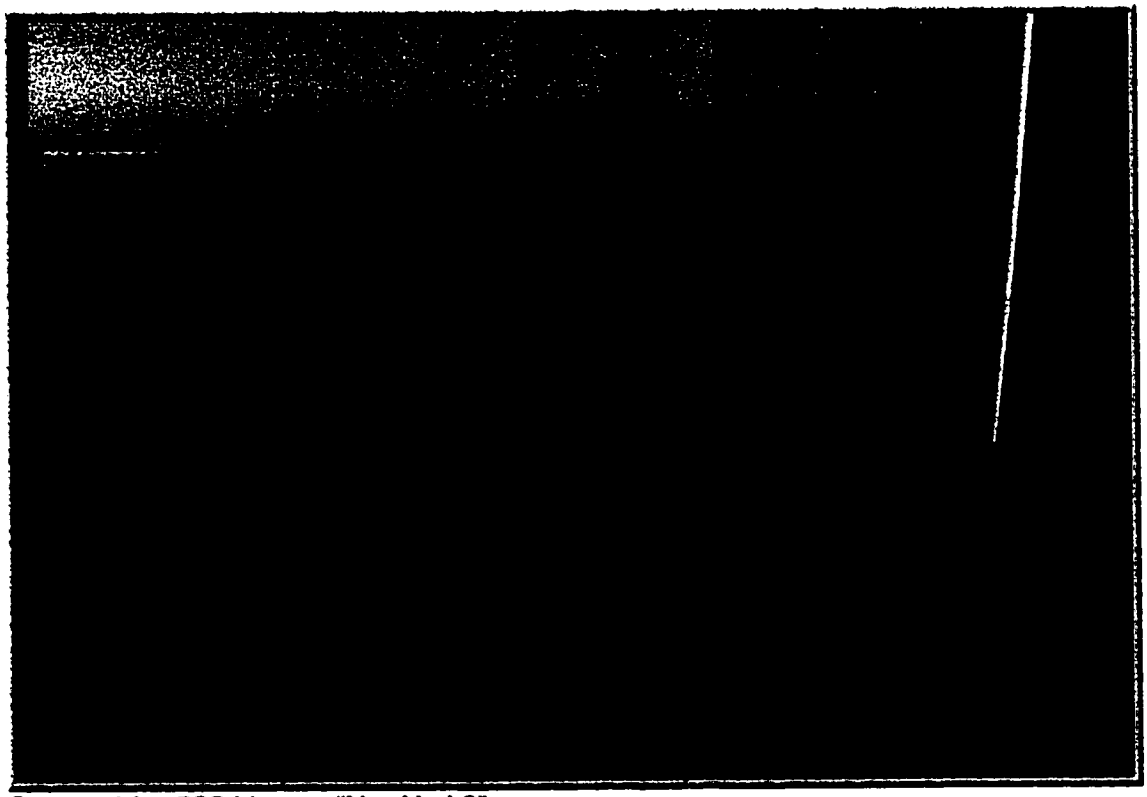
	A part of the ATSC digital television specification that enables a DTV receiver to identify program information contributed by content providers and use it to create sophisticated electronic program guides.
PSNR	Peak signal to noise ratio
PTS	Presentation time-stamp; A field that may be present in a PES packet header that indicates the time that a presentation unit is presented in the system target decoder.
PTS	Presentation time stamp
QAM	Quadrature Amplitude Modulation
QCIF	Quarter Common Intermediate Format. Has half of the CIF resolution.
QPSK	Quadrature Phase Shift Keying
Quantizer	A processing step which intentionally reduces the precision of DCT coefficients
RR	Reduced Reference
RST	Running status table
SCR	System Clock Reference. A time stamp in the program stream from which decoder timing is derived.
Scrambling	The alteration of the characteristics of a video, audio or coded data stream in order to prevent unauthorized reception of the information in a clear form. This alteration is a specified process under the control of a conditional access system.

SDTV	Standard Definition Television. Digital formats that do not achieve the video quality of HDTV, but are at least equal, or superior to, NTSC pictures. SDTV may have either 4:3 or 16:9 aspect ratios, and it includes surround sound. Variations of fps (frames per second), lines of resolution, and other factors of 480p and 480i make up the 12 SDTV formats in the ATSC standard. This equivalent quality may be achieved from pictures sourced at the 4:2:2 level of ITU-R Recommendation 601 and subjected to processing as part of the bit rate compression. The results should be such that when judged across a representative sample of program material, subjective equivalence with NTSC is achieved. Also called standard digital television. See also conventional definition television and ITU-R Recommendation 1125.
SECAM	Systeme Electronique Coloeur avec Memoire
Slice	A series of consecutive macroblocks.
SMPTE	Society of Motion Picture and Television Engineers.
SPI	Synchronous parallel interface
SRC	Source Reference Channel or Circuit
SSCQE	Single Stimulus Continuous Quality Evaluation
TDT	Time and data table
TOT	Time offset table
VBR	Variable Bit Rate Operation where the bit rate varies with time during the decoding of a compressed bit stream.

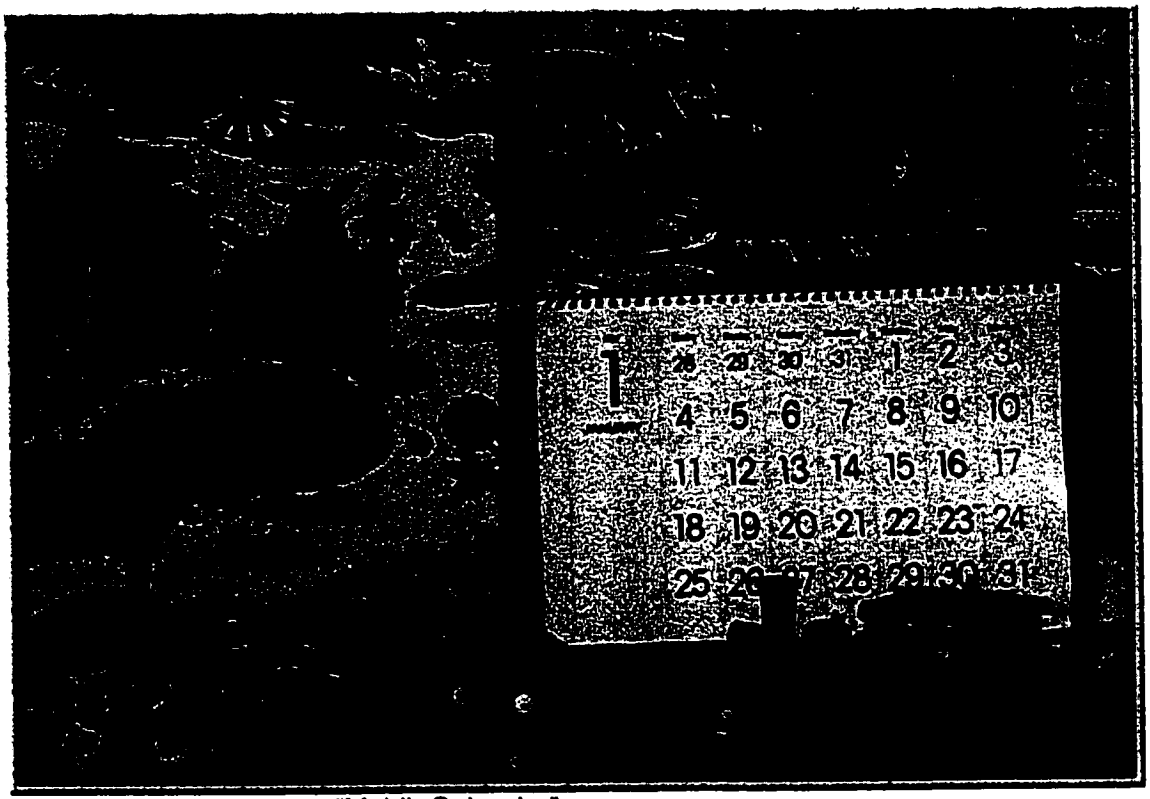
Video Sequence	A video sequence is represented by a sequence header, one or more groups of pictures, and an end_of_sequence code in the data stream.
VQEG	Video Quality Experts Group
VTR	Video Tape Recorder

Appendix A

Source 13 – 525 Lines – "Baloon Pops"

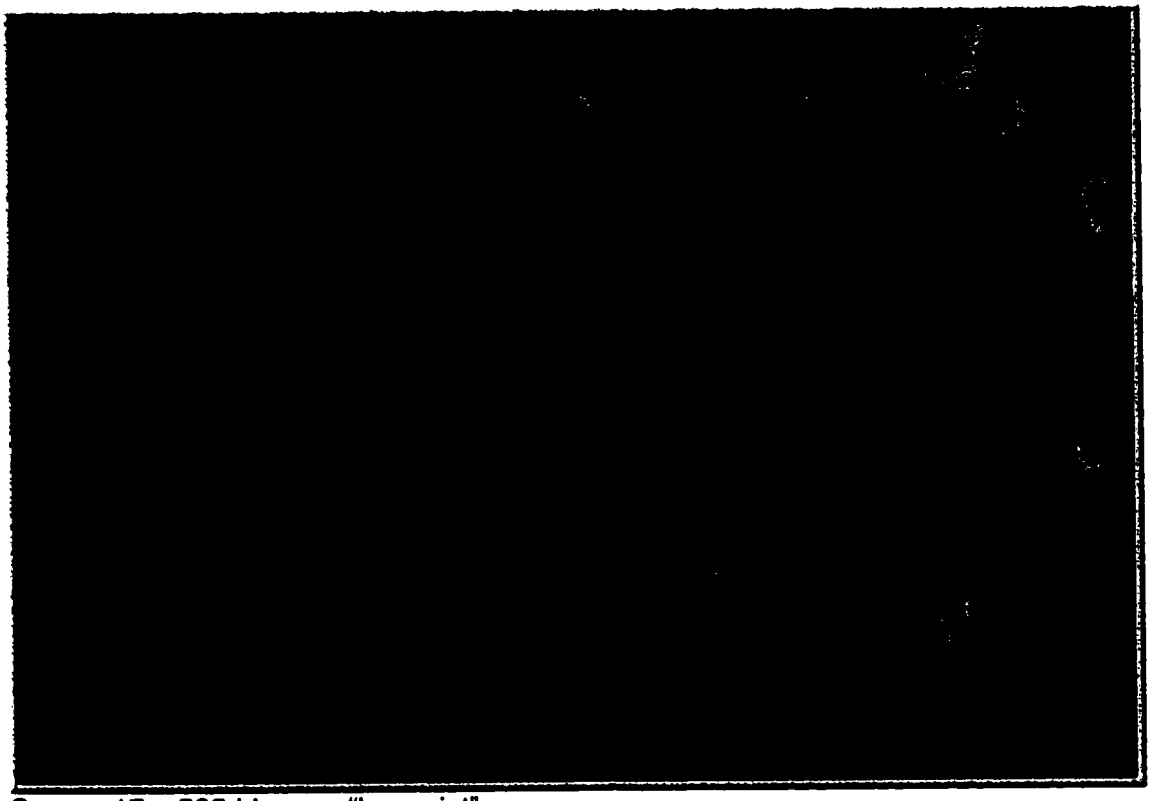


Source 14 – 525 Lines – "NewYork2"

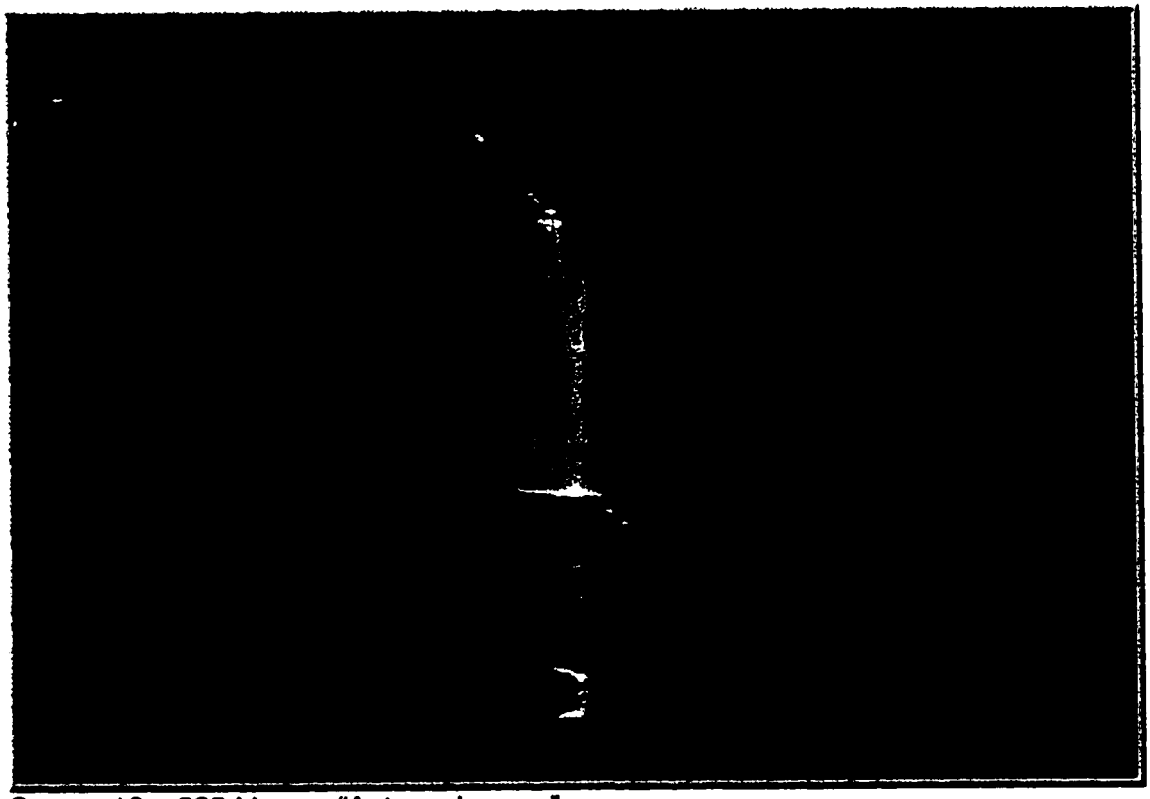


Source 15 – 525 Lines – "MobileCalendar"

Source 16 – 525 Lines – "Betes_pas_betes"



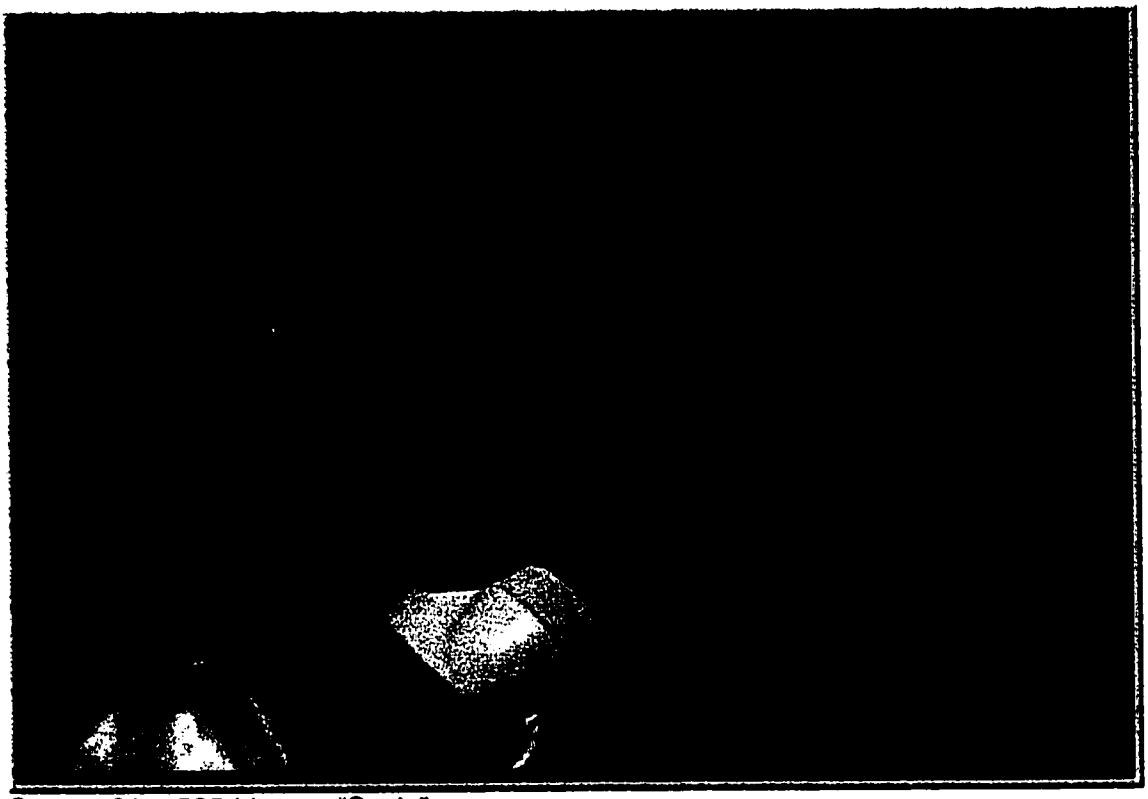
Source 17 – 525 Lines - : "Le_point"



Source 18 – 525 Lines – "Autumn Leaves"

Source 19 – 525 Lines – "Football"

Source 20 - 525 Lines – "Sailboat"



Source 21 – 525 Lines – "Susie"

Source 22 – 525 Lines – "Tempete"