

TECHNICAL REPORT

IEC
TR 62251

First edition
2003-05

**Multimedia systems and equipment –
Quality assessment –
Audio-video communication systems**

IECNORM.COM : Click to view the full PDF of IEC TR 62251:2003



Reference number
IEC/TR 62251:2003(E)

Publication numbering

As from 1 January 1997 all IEC publications are issued with a designation in the 60000 series. For example, IEC 34-1 is now referred to as IEC 60034-1.

Consolidated editions

The IEC is now publishing consolidated versions of its publications. For example, edition numbers 1.0, 1.1 and 1.2 refer, respectively, to the base publication, the base publication incorporating amendment 1 and the base publication incorporating amendments 1 and 2.

Further information on IEC publications

The technical content of IEC publications is kept under constant review by the IEC, thus ensuring that the content reflects current technology. Information relating to this publication, including its validity, is available in the IEC Catalogue of publications (see below) in addition to new editions, amendments and corrigenda. Information on the subjects under consideration and work in progress undertaken by the technical committee which has prepared this publication, as well as the list of publications issued, is also available from the following:

- **IEC Web Site** (www.iec.ch)

- **Catalogue of IEC publications**

The on-line catalogue on the IEC web site (http://www.iec.ch/searchpub/cur_fut.htm) enables you to search by a variety of criteria including text searches, technical committees and date of publication. On-line information is also available on recently issued publications, withdrawn and replaced publications, as well as corrigenda.

- **IEC Just Published**

This summary of recently issued publications (http://www.iec.ch/online_news/justpub/jp_entry.htm) is also available by email. Please contact the Customer Service Centre (see below) for further information.

- **Customer Service Centre**

If you have any questions regarding this publication or need further assistance, please contact the Customer Service Centre:

Email: custserv@iec.ch
Tel: +41 22 919 02 11
Fax: +41 22 919 03 00

TECHNICAL REPORT

IEC
TR 62251

First edition
2003-05

Multimedia systems and equipment – Quality assessment – Audio-video communication systems

IECNORM.COM : Click to view the full PDF of IEC TR 62251:2003

© IEC 2003 — Copyright - all rights reserved

No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

International Electrotechnical Commission, 3, rue de Varembe, PO Box 131, CH-1211 Geneva 20, Switzerland
Telephone: +41 22 919 02 11 Telefax: +41 22 919 03 00 E-mail: inmail@iec.ch Web: www.iec.ch



Commission Electrotechnique Internationale
International Electrotechnical Commission
Международная Электротехническая Комиссия

PRICE CODE

W

For price, see current catalogue

CONTENTS

FOREWORD	4
1 Scope	5
2 References	5
3 Terms and definitions	6
4 Configuration for quality assessment	6
4.1 Input and output channels	6
4.2 Points of input and output terminals	7
5 Video quality	8
5.1 Introduction	8
5.2 End-to-end tone reproduction	9
5.3 End-to-end colour reproduction	11
5.4 End-to-end colour differences	12
5.5 End-to-end peak-signal to noise ratio (PSNR)	16
5.6 End-to-end objective assessment of video quality	20
6 Audio quality	21
6.1 Perceived audio quality with full-reference signals	21
6.2 Sampling rate and quantization resolution	23
6.3 Delay	24
7 Total quality	24
7.1 Synchronization of audio and video (lip sync)	24
7.2 Scalability	25
7.3 Overall quality	26
Annex A (informative) PSNR's defined in three-dimensional spaces applied to hypothetical deterioration over the reference video sources	27
Annex B (informative) End-to-end objective assessment of video quality in spatial frequency domain	31
Annex C (informative) PEAQ objective measurement method outline	36
Bibliography	40
Figure 1 – Model of audio-video communication systems	7
Figure 2 – Schematic diagram for quality assessment	7
Figure 3 – The image of the grey steps defined in IEC 61146-1	9
Figure 4 – An example plot of tone reproduction	10
Figure 5 – The image of the colour reproduction chart defined in IEC 61146-1	11
Figure 6 – Colour differences between reference and streamed video frames at 250 kbps and 30 fps	15
Figure 7 – Examples of PSNR assessment	19
Figure 8 – Basic concept for making objective measurements	22
Figure 9 – Representation of PEAQ model	22
Figure B.1 – Assignment of the block numbers	31
Figure B.2 – Example of wavelet decomposition visualised	31

Figure B.3 – Trends of difference of coefficients of wavelet transform between reference and streamed video frames at 250 kbps and 30 fps	34
Figure C.1 – Basic concept for making objective measurements	36
Figure C.2 – Representation of PEAQ model.....	37
Figure C.3 – FFT based ear model, PEAQ basic version	37
Figure C.4 – Filter bank based ear model, PEAQ advanced version	38
Table 1 – An example of tone reproduction	10
Table 2 – An example of colour reproduction.....	12
Table 3 – Grand averages of colour differences	16
Table 4 – Overall PNSR's averaged over the frames.....	20
Table 5 – Test items and resulting DI and ODG values for the basic version	23
Table 6 – Test items and resulting DI and ODG values for the advanced version	23
Table A.1 – Reference video sources available for objective assessment.....	27
Table A.2 – PSNR's in various colour spaces and the colour difference for SRC13 and SRC14	28
Table A.3 – PSNR's in various colour spaces and the colour difference for SRC15 and SRC16	28
Table A.4 – PSNR's in various colour spaces and the colour difference for SRC17 and SRC18	29
Table A.5 – PSNR's in various colour spaces and the colour difference for SRC19 and SRC20	29
Table A.6 – PSNR's in various colour spaces and the colour difference for SRC21 and SRC22	30
Table B.1 – Summary of difference of coefficients of wavelet coefficients.....	35
Table C.1 – Model output variables, PEAQ basic version	38
Table C.2 – Model output variables, PEAQ advanced version	39

INTERNATIONAL ELECTROTECHNICAL COMMISSION

MULTIMEDIA SYSTEMS AND EQUIPMENT – QUALITY ASSESSMENT – AUDIO-VIDEO COMMUNICATION SYSTEMS

FOREWORD

- 1) The IEC (International Electrotechnical Commission) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of the IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, the IEC publishes International Standards. Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. The IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of the IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested National Committees.
- 3) The documents produced have the form of recommendations for international use and are published in the form of standards, technical specifications, technical reports or guides and they are accepted by the National Committees in that sense.
- 4) In order to promote international unification, IEC National Committees undertake to apply IEC International Standards transparently to the maximum extent possible in their national and regional standards. Any divergence between the IEC Standard and the corresponding national or regional standard shall be clearly indicated in the latter.
- 5) The IEC provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with one of its standards.
- 6) Attention is drawn to the possibility that some of the elements of this technical report may be the subject of patent rights. The IEC shall not be held responsible for identifying any or all such patent rights.

The main task of IEC technical committees is to prepare International Standards. However, a technical committee may propose the publication of a technical report when it has collected data of a different kind from that which is normally published as an International Standard, for example "state of the art".

Technical reports do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful by the maintenance team.

IEC 62251, which is a Technical Report, has been prepared by IEC technical committee 100: Audio, Video and Multimedia Systems and Equipment.

The text of this technical report is based on the following documents:

Enquiry draft	Report on voting
100/561/DTR	100/662/RVC

Full information on the voting for the approval of this technical report can be found in the report on voting indicated in the above table.

This publication has been drafted in accordance with the ISO/IEC Directives, Part 2.

MULTIMEDIA SYSTEMS AND EQUIPMENT – QUALITY ASSESSMENT – AUDIO-VIDEO COMMUNICATION SYSTEMS

1 Scope

This Technical Report specifies items to be measured by objective methods, methods of measurement together with measuring conditions, processing of the measured data and presentation of acquired information for objective assessment of end-to-end quality of audio-video communication systems over digital networks. The measurements are supposed to be conducted in a double-ended and a full reference. The systems are assumed to have electrical interface channels at the input and at the output of audio-video signals for objective assessment.

The extension for systems that do not have such channels is left for further study.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies

IEC 60268-4, *Sound system equipment – Part 4: Microphones*

IEC 60268-5, *Sound system equipment – Part 5: Loudspeakers*

IEC 61146-1:1994, *Video cameras (PAL/SECAM/NTSC) – Methods of measurement – Part 1: Non-broadcast single-sensor cameras*

IEC 61146-2:1997, *Video cameras (PAL/SECAM/NTSC) – Methods of measurement – Part 2: Two- and three-sensor professional cameras*

IEC 61966-2-1:1999, *Multimedia systems and equipment – Colour measurement and management – Part 1: Colour management – Default RGB colour space – sRGB*

Amendment 1 (2003)

IEC 61966-2-1, *Multimedia systems and equipment – Colour measurement and management – Part 2-1: Colour management – Default RGB colour space – sRGB*

IEC 61966-3:2000, *Multimedia systems and equipment – Colour measurement and management – Part 3: Equipment using cathode ray tubes*

IEC 61966-4:2000, *Multimedia systems and equipment – Colour measurement and management – Part 4: Equipment using liquid crystal display panels*

IEC 61966-5:2000, *Multimedia systems and equipment – Colour measurement and management – Part 5: Equipment using plasma display panels*

IEC 61966-9:2000, *Multimedia systems and equipment – Colour measurement and management – Part 9: Digital cameras*

CIE 15.2:1986, *Colorimetry*

ITU-R BS.1387-1 :2001, *Method for objective measurements of perceived audio quality*

ITU-R BT.601-5 :1995, *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*

ITU-T J.144 :2001, *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*

ITU-T P.931 :1998, *Multimedia communications delay, synchronization and frame rate measurement*

3 Terms and definitions

In order to understand this Technical Report, the following terms and definitions apply.

3.1

audio-video communication system

system that handles audio, video and optionally other data streams in a synchronized way within users' perception in order to transmit and/or exchange information, which is assumed to operate over a local- or wide-area digital network

3.2

DMOS

difference between the source and processed Mean Opinion Scores (MOS) resulting from the subjective testing experiment conducted by the Video Quality Expert Group (VQEG)

3.3

PEAQ

perceived evaluation of audio quality defined by ITU-R BS.1387-1

3.4

PSNR

objective video quality metric defined by peak-signal to noise ratio, the noise being calculated from the source and processed video frames

3.5

VQR

objective video quality rating reduced from any objective metric by being optimally correlated with the DMOS

4 Configuration for quality assessment

4.1 Input and output channels

Audio signal and video signal in audio-video streams shall be captured at the input and at the output channel, respectively, of the audio-video communication system as shown in Figure 1.

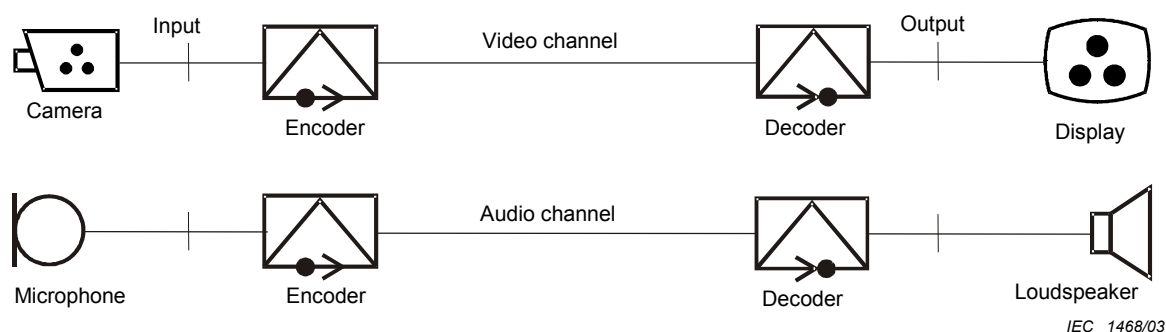
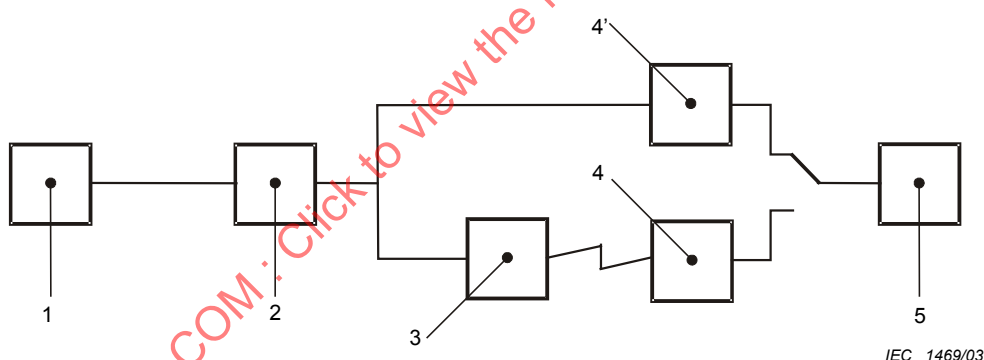


Figure 1 – Model of audio-video communication systems

4.2 Points of input and output terminals

In the spirit of the end-to-end quality assessment of audio-video communication systems, the points for acquisition of raw data should be as far as ultimate end points as possible. However, since the methods of measurement and characterization for equipment which incorporates input transducers such as video cameras and microphones have already been standardized, such as in IEC 61146-1, IEC 61146-2, IEC 61966-9 and IEC 60268-4, and the methods of measurement and characterization of equipment which incorporates output transducers such as video signal displays and loudspeakers, as in IEC 61966-3, IEC 61966-4, IEC 61966-5 and IEC 60268-5, they can be outwith the scope of the range of the end-to-end.

Figure 2 shows a schematic diagram for quality assessment under double-ended and full reference conditions.



Key

- 1 Original audio or video reference.
- 2 Pre-conditioner: reduced dynamic range, frequency range for audio; reduced frame size and frame rate for video to fit to the quality assessment of the audio-video communication systems, if necessary.
- 3 Encoder for network streaming with a specified bit-rate in order to fit to the bandwidth of end-to-end network connection.
- 4 Decoder and rendering for the received data to make them audible and visible.
- 4' Rendering for the preconditioned data to make them audible and visible, optional.
- 5 Data acquisition and calculation for quality assessment to provide information specified in this report.

Figure 2 – Schematic diagram for quality assessment

5 Video quality

5.1 Introduction

For the purpose of end-to-end objective assessment of video quality, two aspects have been covered in this Technical Report; one is static characteristics such as tone reproduction and colour reproduction described in 5.2 and 5.3, the other dynamic characteristics based on streaming of video frames to networks described in 5.4, 5.5 and 5.6.

It is recommended to make use of a set of commonly available video sources as reference such as the test sequences in the Canadian Research Centre (CRC) as the original video reference for the item 1 in Figure 2. Because of its high bit-rate and large frame size, the reference source should be reduced in frame size and bit-rate for use as the item 2 in Figure 2, if necessary, for actual encoding as streaming video to a network with limited bandwidth.

For the dynamic characteristics, reference video sequences currently available are listed in Table A.1. All reference video sources in Table A.1 have been adopted in this Technical Report with the permission of the owner, the Canadian Research Council (CRC), which were used by the Video Quality Expert Group (VQEG) for subject video quality tests to obtain the Difference of Mean Opinion Score (DMOS) and also object Video Quality Metric (VQR) as reported in ITU-R 10-11Q/56-E.

The format of each of the reference video sources is composed of 10 frames (for leader) + video frames for 8 s + 10 frames (for trailer). There are two video formats 525/60Hz and 625/50Hz, but only the 525/60Hz format shown in Table A.1 is adopted in this Technical Report for evaluation.

Each line is in pixel multiplexed 4:2:2 component video format as Cb Y Cr Y ... and so on, encoded in line with ITU-R BT.601-5, where 720 bytes/line for Y, 360 B/line for Cb and 360 B/line for Cr. The lines are concatenated into frames and frames are concatenated to form the sequence files.

The format contains 720 pixels (1 440 bytes) per horizontal line and has 486 active lines per frame. The frame sizes are $1\,440 \times 486 = 699\,840$ B/frame and the sequence sizes are 240 frames file size for 8 s + 20 frames. Thus, file size is $699\,840$ bytes/frame $\times$ 260 frames = $181\,958\,400$ bytes. 30 frame/s will result a bit-rate of $699\,840$ bytes/frame $\times$ 30 frame/s $\times$ 8 bits = $167\,961\,600$ bit/s. Since it is a too high bit-rate to be handled by ordinary personal computers and to be streamed to the Internet, the original test sequences have been reduced in frame size to be 320×240 pixels, and in format to be RGB (instead of YCC) 24-bit/pixel to fit to a typical video format (AVI) where IEC 61966-2-1 is taken into account.

NOTE 1 Pixel-by-pixel error assessment requires a very high degree of normalisation to be used with confidence. The normalisation requires both spatial and temporal alignment as well as corrections for gain and offset. For this purpose, Clause A2 of ITU-R 6Q/39-E should be referred to.

NOTE 2 Since the values of objective quality metrics largely depend on video contents, varieties of commonly available video sources should be used as far as possible.

NOTE 3 Video quality metrics obtained by objective assessment in Clause 5 should be converted to be VQR by optimum correlation with DMOS, which is under consideration within ITU-R WP 6Q.

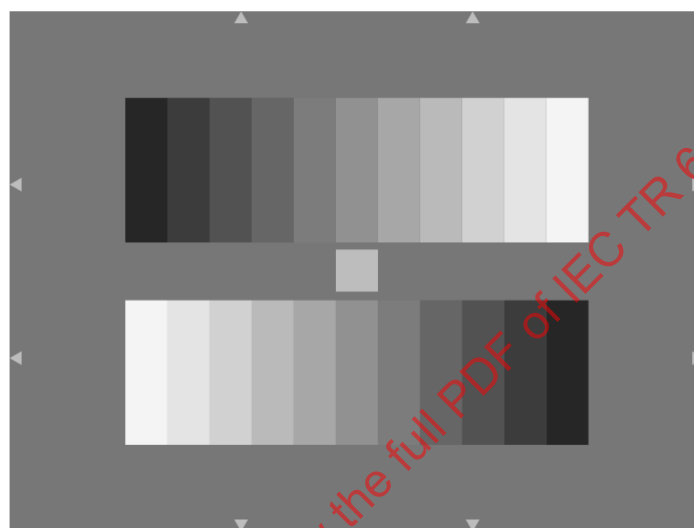
5.2 End-to-end tone reproduction

5.2.1 Items to be assessed

End-to-end non-linearity in term of tone reproduction.

5.2.2 Method of assessment

An image of the grey steps chart defined in IEC 61146-1, as shown in Figure 3, should be used as the reference source for item 1 of Figure 2. The still neutral image should be prepared as a file for item 2 of Figure 2 and repeatedly encoded to be a streaming video transmitted to a network.



IEC 1470/03

Figure 3 – The image of the grey steps defined in IEC 61146-1

The received streaming video should be decoded and rendered by a viewer for the incoming streaming videos. The image data to be displayed should be captured at an output terminal.

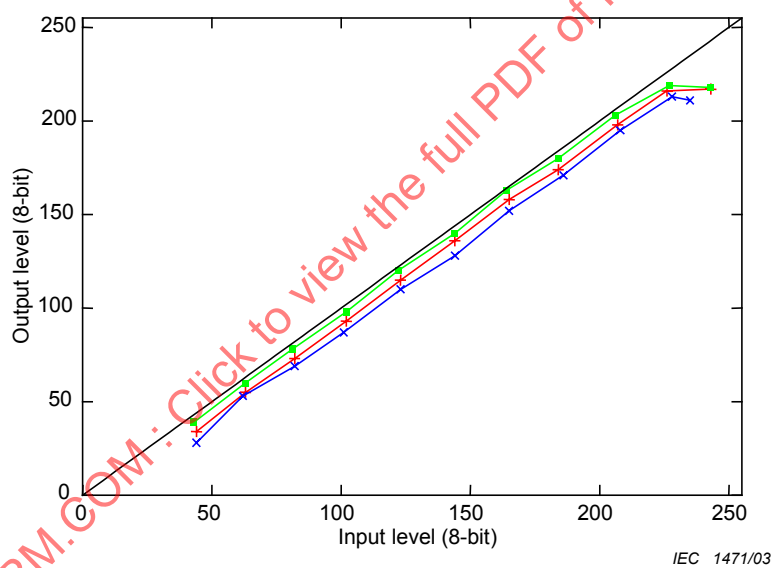
The image data should be compared in terms of three component data, R, G, and B, averaged in each of the corresponding areas.

5.2.3 Presentation of assessment result

The data for display versus the input image data should be reported as a table and a plot as shown in Table 1 and Figure 4, respectively, as examples, together with the audio-video communication system under assessment and the specification of the input-output point.

Table 1 – An example of tone reproduction

	Specification			Input			Output		
	R %	G %	B %	R	G	B	R	G	B
0	2,0	2,0	2,0	44	43	44	34	39	28
1	4,5	4,5	4,5	63	63	62	55	60	53
2	8,1	8,1	8,1	82	81	82	73	78	69
3	13,0	13,0	13,0	102	102	101	93	98	87
4	19,8	19,8	19,8	123	122	123	115	120	110
5	27,9	27,9	27,9	144	144	144	136	140	128
6	37,8	37,8	37,8	165	164	165	158	163	152
7	48,6	48,6	48,6	184	184	186	174	180	171
8	63,0	63,0	63,0	207	206	208	198	203	195
9	77,3	77,3	77,3	226	227	228	216	219	213
10	89,9	89,9	89,9	243	243	235	217	218	211



IEC 1471/03

Figure 4 – An example plot of tone reproduction

5.3 End-to-end colour reproduction

5.3.1 Item to be assessed

End-to-end colour shifts in the CIELAB colour space for a still colour image.

5.3.2 Method of assessment

An image of the colour reproduction chart defined in IEC 61146-1, as shown in Figure 5, should be used as the reference source for item 1 of Figure 2. The still colour image should be prepared as a file for item 2 of Figure 2 and repeatedly encoded to be a streaming video transmitted to a network.



IEC 1472/03

Figure 5 – The image of the colour reproduction chart defined in IEC 61146-1

The received streaming video should be decoded and rendered by a viewer for streaming videos. The colour image data to be displayed should be captured at an output terminal.

The image data should be acquired in terms of three component data, R, G and B, averaged in each of the corresponding areas.

5.3.3 Presentation of assessment result

Input colours and output colours in R, G and B data should be regarded to be in sRGB defined in IEC 61966-2-1. They should be converted to CIE 1976 $L^*a^*b^*$ uniform colour space. Colour differences ΔE_{ab}^* between the reference data and the received data should be calculated and reported as shown in Table 2.

Table 2 – An example of colour reproduction

	Specification			Input (8-bit x 3)			Output (8-bit x 3)			Colour difference ΔE_{ab}^*
	R %	G %	B %	R	G	B	R	G	B	
0	87,053	80,546	87,216	222	205	222	221	211	215	3,5
1	48,904	24,181	23,419	184	134	132	186	135	129	1,4
2	37,405	27,352	12,466	163	141	99	164	144	91	4,2
3	25,874	32,782	5,646	138	154	69	139	156	66	2,0
4	12,176	34,717	19,279	98	158	121	96	158	123	0,9
5	15,414	34,081	41,443	109	156	171	109	158	166	2,2
6	17,982	29,222	61,449	117	146	204	119	145	196	1,9
7	36,893	24,007	52,231	164	130	190	163	137	187	3,9
8	51,332	22,896	45,507	188	130	178	187	132	162	5,3
9	43,311	3,062	4,885	174	52	65	172	56	54	8,1
10	83,988	56,759	4,964	236	197	65	219	201	62	4,7
11	2,426	25,943	13,965	47	138	104	45	140	105	1,2
12	3,259	7,178	18,424	54	77	118	50	77	113	2,7
13	82,033	49,052	37,190	233	184	163	219	186	157	3,7
14	10,356	12,908	4,612	91	101	63	89	100	53	5,2

$$\overline{\Delta E_{ab}^*} = 3,396$$

5.4 End-to-end colour differences

5.4.1 Item to be assessed

The average of colour differences in the psychophysically uniform colour space defined in CIE 15.2 between the reference video frame and the corresponding deteriorated video frame.

5.4.2 Method of assessment

Reference videos in Table A.1 are used as item 1 of Figure 2. Frame size reduced videos in uncompressed AVI-format should be prepared for item 2 of Figure 2. It is necessary to embed frame numbers at this point so that they can be used to identify received frames corresponding to the transmitted frames.

Encoded and transmitted streaming videos shall be continuously captured. Pixel-by-pixel calculations should be conducted.

The average colour difference $\overline{\Delta E_{ab_k}^*}$ in the psychophysically uniform colour space between the reference and the deteriorated frames k is defined as in equation (1).

$$\overline{\Delta E_{ab_k}^*} = \frac{1}{K} \sum_{m=M_1}^{M_2} \sum_{n=N_1}^{N_2} \Delta E_{ab_k}^* \quad (1)$$

where

$$K = \frac{1}{(M_2 - M_1 + 1)(N_2 - N_1 + 1)};$$

$L_{o_k}^*$, $a_{o_k}^*$ and $b_{o_k}^*$

are the triplets in the CIELAB colour space corresponding to each pixel of the reference video frame k ;

$L_{d_k}^*$, $a_{d_k}^*$ and $b_{d_k}^*$

are the triplet in the CIELAB colour space corresponding to each pixel of the deteriorated video frame k ;

$$\Delta E_{ab_k}^* = \sqrt{(L_{d_k}^* - L_{o_k}^*)^2 + (a_{d_k}^* - a_{o_k}^*)^2 + (b_{d_k}^* - b_{o_k}^*)^2}$$

is the CIELAB colour difference between the pixels.

The triplets in the CIELAB colour space should be deduced from the pixel values R , G and B of the reference and the deteriorated video frames in the default RGB colour space (sRGB) defined in IEC 61966-2-1. Each pixel is positioned at row m and column n in a video frame.

5.4.3 Presentation of assessment results

The colour difference between each of the corresponding frames should be plotted versus frame numbers as shown in Figure 6 together with identifications of reference video sources. The conditions of measurement such as frame size in pixels, frame rate, streaming bit-rate should also be reported.

IECNORM.COM : Click to view the full PDF of IEC TR 62251:2003

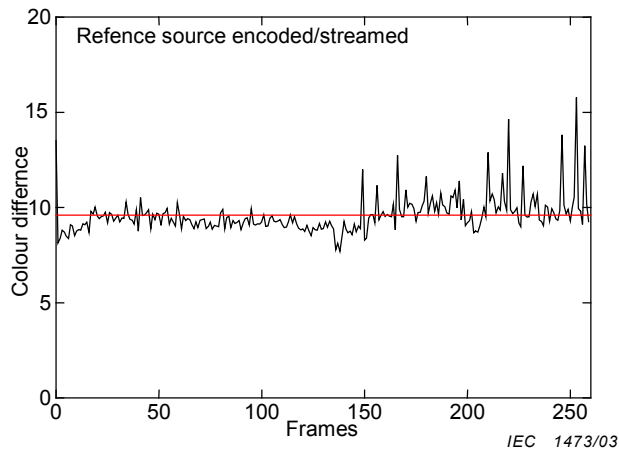


Figure 6a – Example for SRC13_REF__525

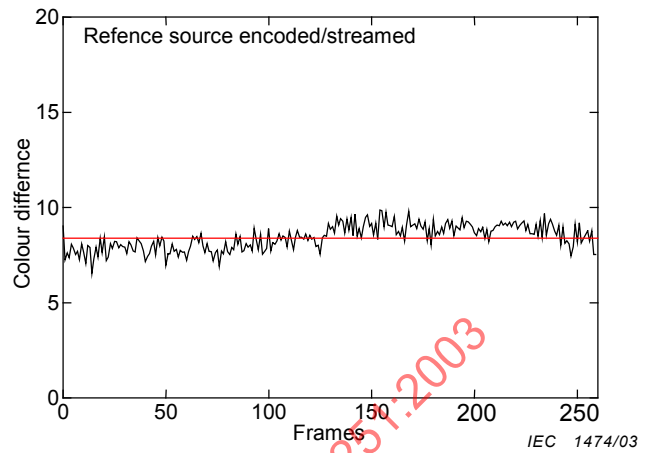


Figure 6b – Example for SRC14_REF__525

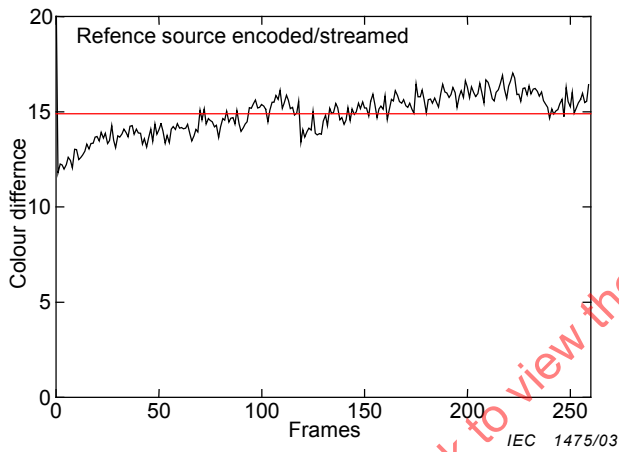


Figure 6c – Example for SRC15_REF__525

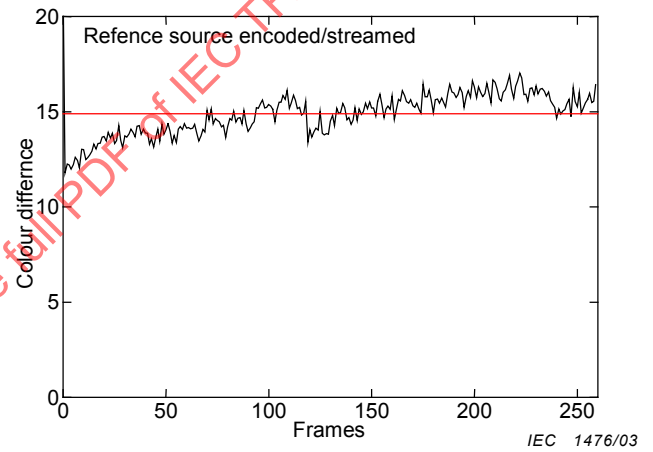


Figure 6d – Example for SRC16_REF__525

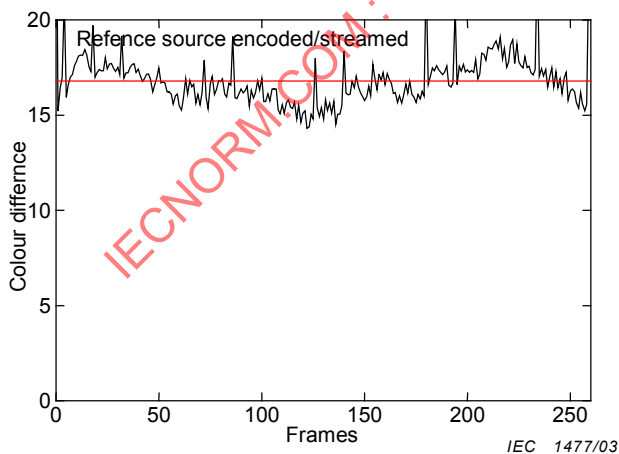


Figure 6e – Example for SRC17_REF__525

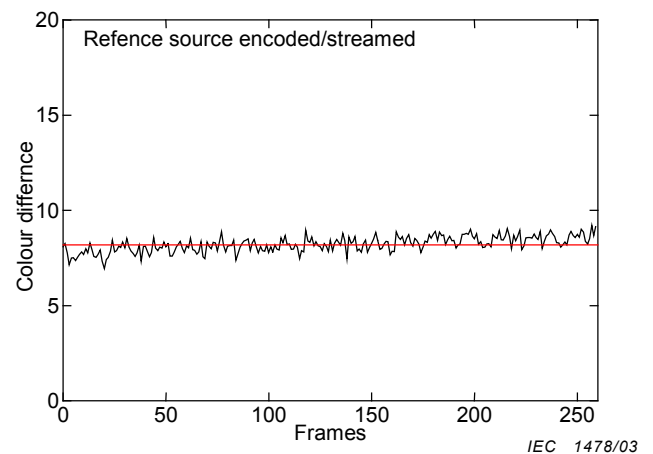


Figure 6f – Example for SRC18_REF__525

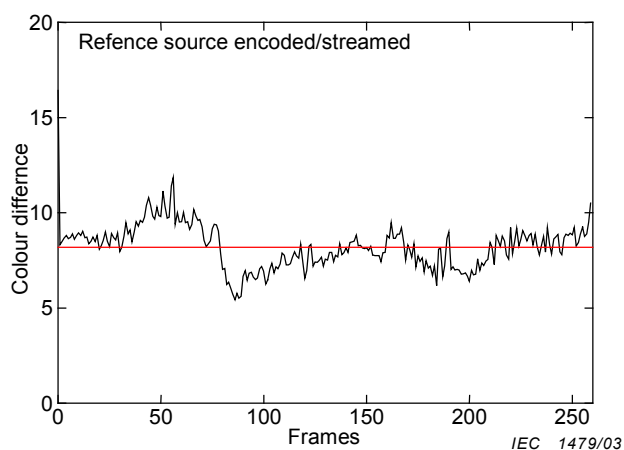


Figure 6g – Example for SRC19_REF__525

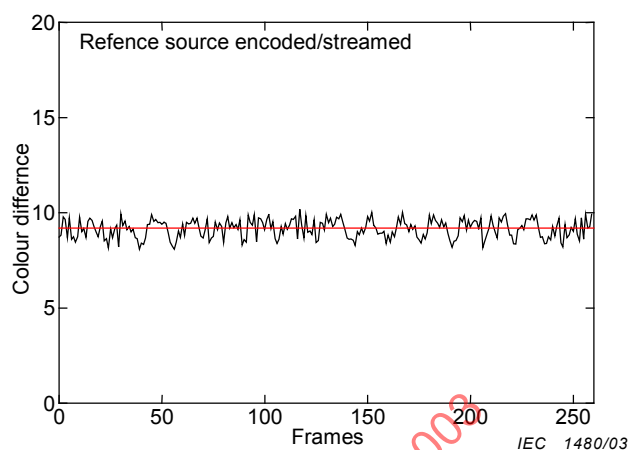


Figure 6h – Example for SRC20_REF__525

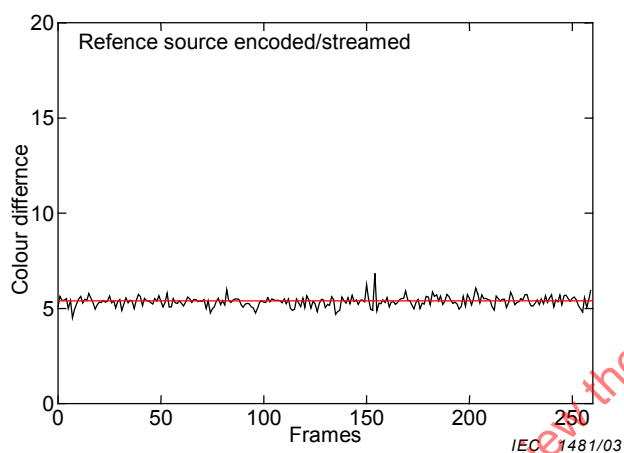


Figure 6i – Example for SRC21_REF__525

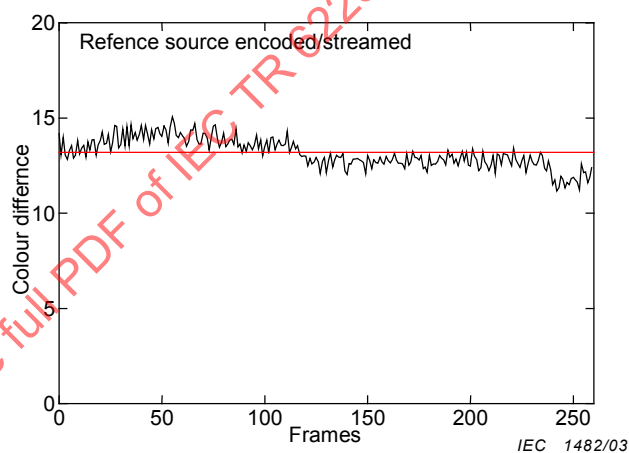


Figure 6j – Example for SRC22_REF__525

Condition of assessment:

video frame size: 320 pixels x 240 pixels

frame rate: 30 s⁻¹

streaming bit-rate: 250 kb s⁻¹

network bandwidth: more than 250 kb s⁻¹

reproduction: Microsoft Media Player® version 7.1

Figure 6 – Colour differences between reference and streamed video frames at 250 kbs⁻¹ and 30 s⁻¹

As a summary of the assessment, acquired data should also be averaged over frames as in equation (2) so as to provide the single metric for objective assessment as the grand average. It should be reported as in Table 3.

$$\overline{\Delta E_{ab}^*} = \frac{1}{(K_2 - K_1 + 1)} \sum_{k=K_1}^{K_2} \overline{\Delta E_{ab_k}^*} \quad (2)$$

Table 3 – Grand averages of colour differences

Reference video source	Grand average of colour difference
SRC13_REF__525	9,6
SRC14_REF__525	8,4
SRC15_REF__525	14,9
SRC16_REF__525	8,3
SRC17_REF__525	16,8
SRC18_REF__525	8,2
SRC19_REF__525	8,2
SRC20_REF__525	9,2
SRC21_REF__525	5,4
SRC22_REF__525	13,2

5.5 End-to-end peak-signal to noise ratio (PSNR)

5.5.1 Item to be assessed

Peak-signal to noise power ratios, PSNRs, in three-dimensional coordinate systems.

5.5.2 Method of assessment

Reference videos in Table A.1 are used as item 1 of Figure 2. Frame size reduced videos in uncompressed AVI-format should be prepared for item 2 of Figure 2. It is necessary to embed frame numbers at this point so that they can be used to identify received frames corresponding to the transmitted frames.

Encoded and transmitted streaming videos shall be continuously captured. Pixel-by-pixel calculation should be conducted.

The peak-signal to noise ratio (PSNR) between a full reference image and a reproduced image recommended in ITU-T J.144 should be used. It defined the PSNR by the following equation (3).

$$PSNR = 10 \lg \left(\frac{S_{\max}^2}{MSE} \right) \quad (3)$$

$$MSE = \frac{1}{K} \sum_{p=P_1}^{P_2} \sum_{m=M_1}^{M_2} \sum_{n=N_1}^{N_2} (d(p, m, n) - o(p, m, n))^2$$

where

$$K = \frac{1}{(P_2 - P_1 + 1)(M_2 - M_1 + 1)(N_2 - N_1 + 1)};$$

$d(p, m, n)$ and $o(p, m, n)$

represent, respectively, degraded and original pixel vectors at frame p , row m and column n ;

$S_{\max}$

is the maximum possible value of the pixel vectors.

For colour images, each picture element is normally composed of three dimensional values, red (R), green (G) and blue (B). Thus, the definition in equation (4) applies for the mean-square errors.

$$MSE_{RGB} = \frac{1}{K} \sum_{p=P_1}^{P_2} \sum_{m=M_1}^{M_2} \sum_{n=N_1}^{N_2} \left((R_d - R_o)^2 + (G_d - G_o)^2 + (B_d - B_o)^2 \right) \quad (4)$$

where $S_{\max(RGB)} = 3 \times 2^{2(N-1)}$ for the values in N -bit encoding.

It is recommended to evaluate the PSNR in the more uniform colour space, CIE 1976 LAB, as in equation (5).

$$MSE_{Lab} = \frac{1}{K} \sum_{p=P_1}^{P_2} \sum_{m=M_1}^{M_2} \sum_{n=N_1}^{N_2} (\Delta E_{ab}^*)^2 \quad (5)$$

where $S_{\max(Lab)} = \sqrt{(L_{\max}^*)^2 + (a_{\max}^*)^2 + (b_{\max}^*)^2}$, actual value of which depends on a colour gamut of original RGB colour space.

It is recommended to use the default RGB colour space defined by IEC 61966-2-1, in which $S_{\max(Lab)} = 148,254$.

NOTE It should be noted that the terms for summation in equation (5) are the square of the colour differences in the psychophysically uniform colour space described in 5.4.

Additionally, luminance signal Y and two colour difference signals C_b and C_r denoted as Ycc will also be calculated for comparison as in equation (6).

$$MSE_{Ycc} = \frac{1}{K} \sum_{p=P_1}^{P_2} \sum_{m=M_1}^{M_2} \sum_{n=N_1}^{N_2} \left((Y_d - Y_o)^2 + (C_{b_d} - C_{b_o})^2 + (C_{r_d} - C_{r_o})^2 \right) \quad (6)$$

where $S_{\max(Ycc)} = 1,01659$ in YCbCr system defined in IEC 61966-2-1, Amendment 1.

5.5.3 Presentation of assessment results

The PSNRs in three-dimensional spaces Lab, Ycc and RGB together with one-dimensional PSNRs in L^* and Y should be reported as shown in Figure 7.

The conditions of measurement such as frame size in pixels, frame rate, streaming bit-rate should also be reported.

NOTE In order to demonstrate software developed by Chiba University in collaboration with Mitsubishi Electric Corp. for various quality metrics regarding the known hypothetical deterioration used in the Video Quality Expert Group (VQEG) in terms of three-dimensional PSNRs and one-dimensional PSNRs together with the average colour difference are attached in Annex A for information.

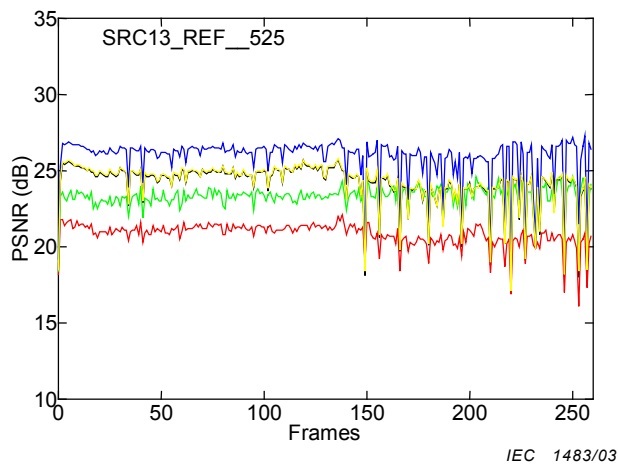


Figure 7a – SRC13_REF__525

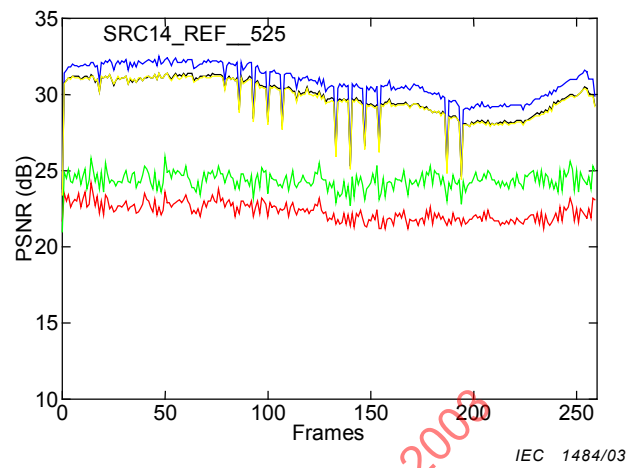


Figure 7b – SRC14_REF__525

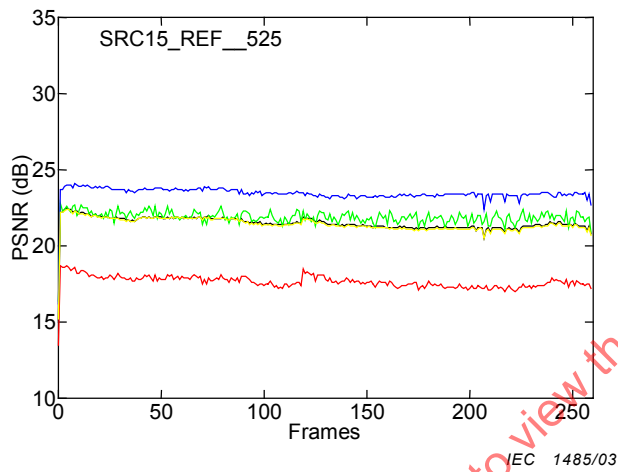


Figure 7c – SRC15_REF__525

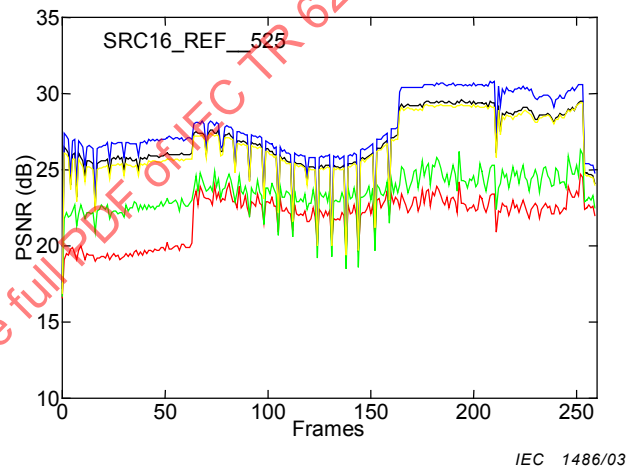


Figure 7d – SRC16_REF__525

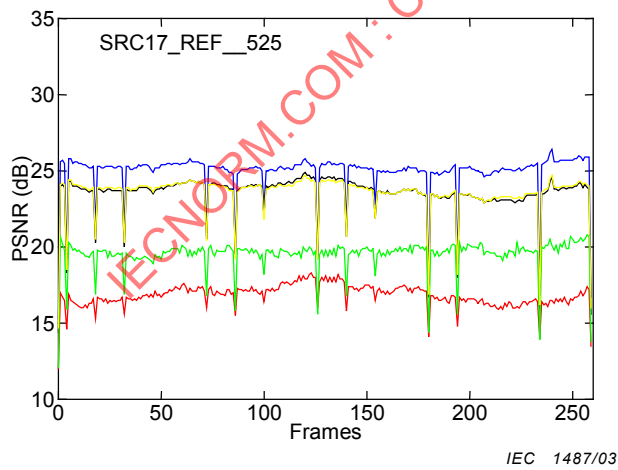


Figure 7e – SRC17_REF__525

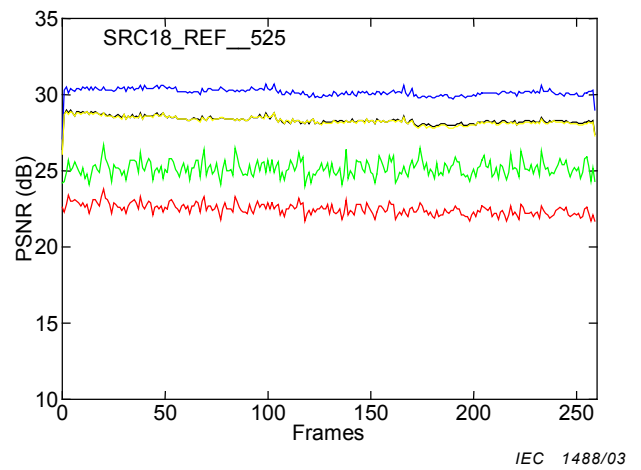


Figure 7f – SRC18_REF__525

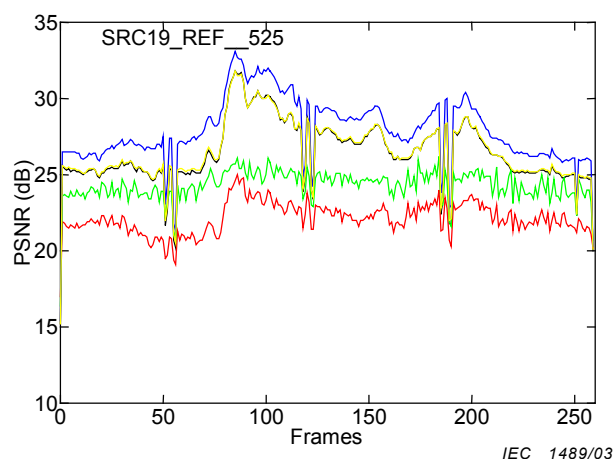


Figure 7g – SRC19_REF__525

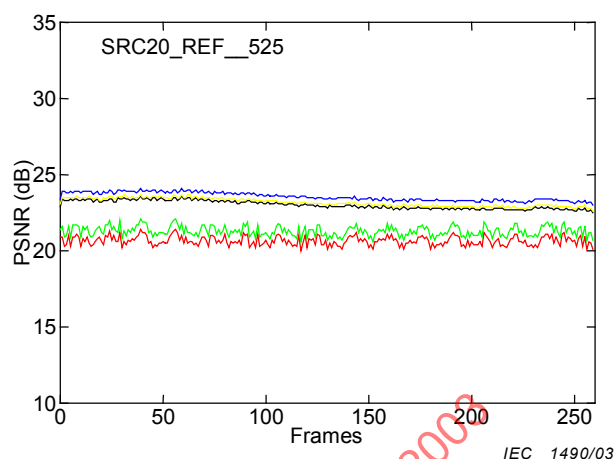


Figure 7h – SRC20_REF__525

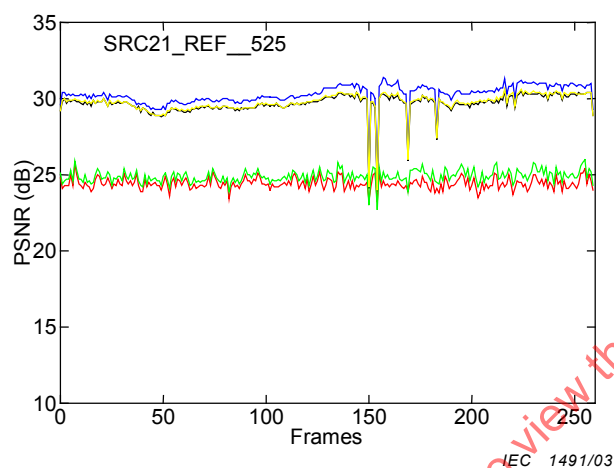


Figure 7i – SRC21_REF__525

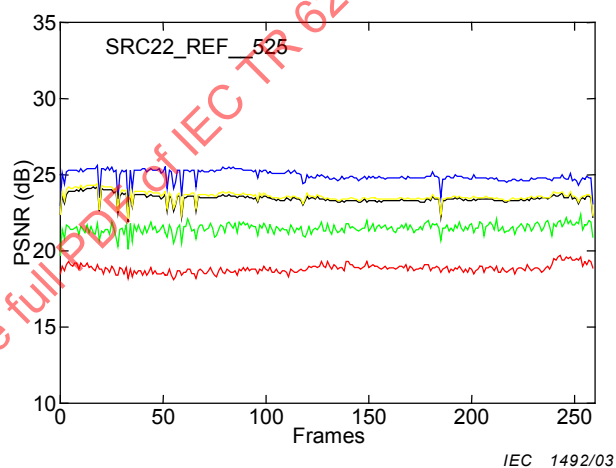


Figure 7j – SRC22_REF__525

Condition of assessment:

video frame size: 320 pixels x 240 pixels

frame rate: 30 s⁻¹

streaming bit-rate: 250 kbs⁻¹

network bandwidth: more than 250 kbs⁻¹

reproduction: Microsoft Media Player® version 7.1

Figure 7 – Examples of PSNR assessment

As a summary of the assessment, the PSNRs should also be averaged over frames so as to provide the overall metrics for objective assessment as in equation (7). It should be reported as in Table 4.

$$\overline{PSNR} = \frac{1}{(K_2 - K_1 + 1)} \sum_{k=K_1}^{K_2} PSNR_k \quad (7)$$

Table 4 – Overall PNSRs averaged over the frames

Reference identification	RSNR in CIELAB	RSNR in YCC	RSNR in RGB	RSNR in L*	RSNR in Y
SRC13_REF_525	20,9	24,4	24,4	23,3	26,1
SRC14_REF_525	22,3	29,9	30,0	24,4	30,9
SRC15_REF_525	17,7	21,5	21,5	21,9	23,5
SRC16_REF_525	22,1	27,0	27,2	23,7	28,2
SRC17_REF_525	16,9	23,7	23,7	19,6	25,1
SRC18_REF_525	22,5	28,3	28,3	25,2	30,2
SRC19_REF_525	22,3	27,0	27,0	24,4	28,4
SRC20_REF_525	20,7	23,2	23,0	21,3	23,6
SRC21_REF_525	24,4	29,8	29,7	24,9	30,3
SRC22_REF_525	18,8	23,6	23,5	21,5	24,9

5.6 End-to-end objective assessment of video quality

5.6.1 Item to be assessed

Estimation of subjective Difference Mean Opinion Scores (DMOS) using a model emulating human visual and perceptual characteristics for digital videos.

5.6.2 Method of assessment

For this purpose, the VQEG in its phase 1 test studied the proposed models from ten proponents (actually 9 out of 10 were considered effective) as reported in the document ITU-R 10-11Q/56-E. They are summarized as follows:

- image evaluation based on segmentation that provides quality prediction over a set of predefined scenes;
- visual discrimination model that simulates the responses of human spatiotemporal visual mechanisms;
- a model to emulate human-visual characteristics using spatiotemporal 3-dimensional filters;
- Mean Square Error (MSE) weighted by human visual filters such as pixel-based, block-based and sequence-based filters;
- perceptual distortion metric based on a spatiotemporal model of the human visual system;
- a model composed of a perceptual model and a feature extractor specifically tuned to certain type of distortions;
- digital video quality incorporating many aspects of human visual sensitivity in a simple image processing;
- perceptual video quality measure using the same approach in measuring video quality as the perceptual speech quality measure in measuring speech quality;
- a model using reduced bandwidth features extracted from spatial-temporal regions and linear combination of parameters to estimate the subjective quality rating.

Performance of all the models was tested in terms of feature extraction capability against the conventional peaksignal-to-noise ratio method.

The VQEG is under the phase 2 test for full reference television among new proponents. A feasible method of assessment (model) is still under consideration.

NOTE "A new method" was submitted by the Republic of Korea that incorporates spatiotemporal wavelet transform as described in ITU-R 6Q/42-E. In this Technical Report, it was examined in the sRGB domain as shown in Annex B.

5.6.3 Presentation of assessment results

Video quality rating as an estimation of difference mean opinion scores should be reported together with the model used and the conditions.

NOTE An example of the presentation is under consideration.

6 Audio quality

6.1 Perceived audio quality with full reference signals

6.1.1 Item to be assessed

Objective Difference Grade (ODG) values, measured by PEAQ (Perceived Evaluation of Audio Quality) method recommended in ITU-R BS.1387-1.

6.1.2 Justification

Perceived audio quality (PEAQ) is one of the key factors when designing digital audio-video communication systems. Formal listening tests have been the relevant method for judging audio quality. However, subjective quality assessments are both time consuming and expensive. It was desirable to develop an objective measurement method in order to produce an estimate of the audio quality. Traditional objective measurement methods, like signal-to-noise-ratio (SNR) or total-harmonic-distortion (THD) have never really been shown to relate reliably to the perceived audio quality. The problems become even more evident when the methods are applied modern codecs, which are both non-linear and non-stationary. After thorough verification, ITU-R recommends an objective measurement method, known as PEAQ (Perceived Evaluation of Audio Quality), to estimate the perceived audio quality of equipment under test, for example a low bit-rate codec. This method is specified in ITU-R BS.1387-1 and described briefly in Annex B.

The output variable from the PEAQ objective measurement method is the objective difference grade (ODG) and distortion index (DI). The ODG corresponds to the subjective difference grade (SDG) in the subjective domain. The resolution of the ODG is limited to one decimal. One should however be cautious and not generally expect that a difference between any pair of ODGs of a tenth of a grade is significant. The DI has the same meaning as the ODG. However, DI and ODG can only be compared quantitatively but not qualitatively. As a general rule, the ODG should be used as the quality measure for ODG values greater than approximately $-3,6$. The ODG correlates very well with subjective assessment in this range. When ODG value is less than $-3,6$, the DI should be used. Therefore, both ODG or DI variables shall be measured.

6.1.3 Method of assessment and algorithm of PEAQ

The basic concept for the PEAQ objective measurement method is illustrated in Figure 8. It consists of two inputs, one for the (unprocessed) reference audio signal, corresponding to item 2 of Figure 2, and the other for the audio signal under the test. The latter may, for example, be the output signal of digital audio-video communication systems, corresponding to the output of item 4 of Figure 2, that is stimulated by the reference signal.

This measurement method is applicable to most types of audio signal processing equipment, both digital and analogue. It is, however, applied by focusing on digital audio communication channels in this document. The block “device under test” corresponds to items 2 and 3 of Figure 2.

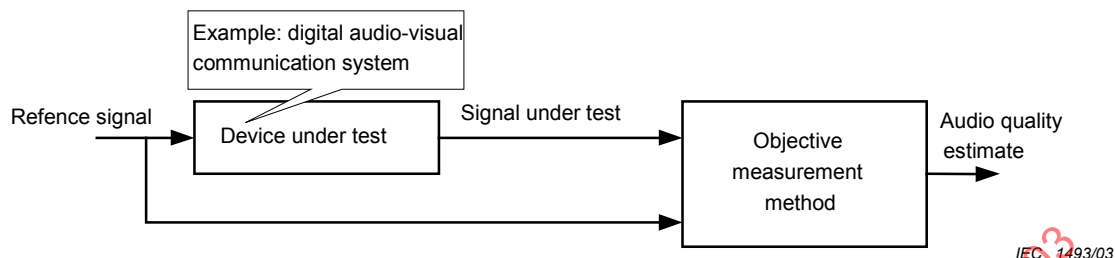


Figure 8 – Basic concept for making objective measurements

A representation of the PEAQ model is shown in Figure 9. The PEAQ method is based on generally accepted psychoacoustic principles. In general, it compares a signal that has been processed in some way with the corresponding time-aligned reference signal. In the first step of signal processing, the peripheral ear is modelled as known as “perceptual model”, or “ear model.” Concurrent frames of the reference and the processed signals are each transformed to the outputs of ear models. In a consecutive step, algorithm models the audible distortion present in the signal under test by comparing the outputs of the ear models. The information obtained by these processes results in several values, so called MOVs (model output variables) and may be useful for detailed analysis of the signal.

The final goal is to drive a quality metric, consisting of a single number that indicates the audibility of the distortions present in the signal under test. In order to archive this, some further processing of the MOVs is required which simulates the cognitive part of the human auditory system. Therefore, the PEAQ algorithm incorporates an artificial neural network.

There are two versions of the PEAQ; a “Basic” version, featuring a low complexity approach, and an “Advanced” version for higher accuracy at the trade off of higher complexity. The structure of both versions is very similar, and fits exactly into the PEAQ model shown in figure 9. The major difference between the basic and the advanced version is hidden in the respective ear models and the set of MOVs used. Annex C provides more information about PEAQ, which helps readers to understand the measurement results.

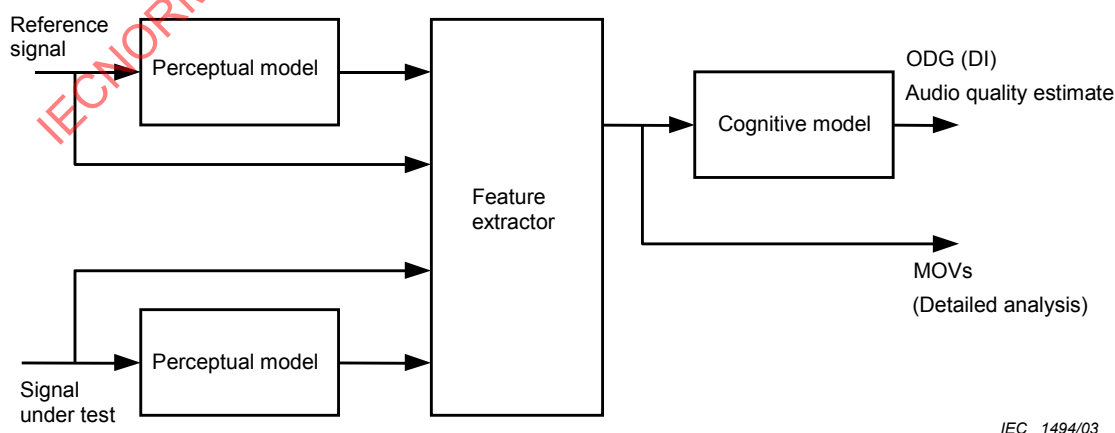


Figure 9 – Representation of the PEAQ model

It is recommended to use the reference items available from the ITU as WAV-files (Microsoft RIFF format) on a CD-ROM. All reference items have been sampled at 48 kHz in 16-bit PCM. The reference and test signals provided by the ITU have already been time and level adapted to each other, so that no additional gain or delay compensation is required.

The measurement algorithm must be adjusted to a listening level of 92 dB SPL.

6.1.4 Presentation of assessment results

PEAQ measurement results should be reported in terms of the names of the reference and signal under test items², and the resulting DI and ODG values in a table.

Table 5 is related to the basic version, and Table 6 contains the values for the advanced version.

Table 5 – Test items and resulting DI and ODG values for the basic version

Item	DI	ODG	Item	DI	ODG	Item	DI	ODG
Acodsna.wav	1,304	–0,676	Fcodtr2.wav	–0,045	–1,927	lcodhrp.wav	1,041	–0,876
Bcodtri.wav	1,949	–0,304	fcodtr3.wav	–0,715	–2,601	lcodpip.wav	1,973	–0,293
Ccodsax.wav	0,048	–1,829	gcodcla.wav	1,781	–0,386	mcodcla.wav	–0,436	–2,331
Dcodryc.wav	1,648	–0,458	hcodryc.wav	2,291	–0,166	ncodsfe.wav	3,135	0,045
Ecodsmg.wav	1,731	–0,412	Hcodstr.wav	2,403	–0,128	scodclv.wav	1,689	–0,435
Fcodsb1.wav	0,677	–1,195	lcodsna.wav	–3,029	–3,786			
Fcodtr1.wav	1,419	–0,598	kcodsme.wav	3,093	0,038			

Table 6 – Test items and resulting DI and ODG values for the advanced version

Item	DI	ODG	Item	DI	ODG	Item	DI	ODG
Acodsna.wav	1,632	–0,467	Fcodtr2.wav	0,162	–1,711	Lcodhrp.wav	1,538	–0,523
Bcodtri.wav	2,000	–0,281	fcodtr3.wav	–0,783	–2,662	lcodpip.wav	2,149	–0,219
Ccodsax.wav	0,567	–1,300	Gcodcla.wav	1,457	–0,573	Mcodcla.wav	0,430	–1,435
Dcodryc.wav	1,725	–0,415	Hcodryc.wav	2,410	–0,126	Ncodsfe.wav	3,163	0,050
Ecodsmg.wav	1,594	–0,489	Hcodstr.wav	2,232	–0,187	Scodclv.wav	1,972	–0,293
Fcodsb1.wav	1,039	–0,877	lcodsna.wav	–2,510	–3,664			
Fcodtr1.wav	1,555	–0,512	Kcodsme.wav	2,765	–0,029			

6.2 Sampling rate and quantization resolution

6.2.1 Item to be assessed

Sampling rate and bandwidth of the reference and the processed audio signals.

6.2.2 Method of assessment

Sampling rate is relevant to the bandwidth of audio signals. For high-quality audio signals, the sampling rate 48 kHz is used. The sampling rate and bandwidth of the reference and processed audio signals should be extracted.

² The names of the corresponding reference items are derived by replacing the substring “cod” in the names of the test items by “ref”, e.g. the reference item for “bcodtri.wav” is “breftri.wav”.

Resolution of quantization relates to the dynamic range of audio signals or quantization noise. For high-quality audio signals, the linear (or uniform) quantization method, which have 16-bit quantization resolution, are used. The resolution and quantization method should be identified.

6.2.3 Presentation of assessment results

Extracted and identified values should be reported.

6.3 Delay

6.3.1 Item to be assessed

Delay time in seconds from audio inputs to an encoder and received digital audio signals.

6.3.2 Method of assessment

Pulsed audio signals should be used as input in terms of item 2 of Figure 2. Lap time between the input of item 3 and the output of item 4 of Figure 2 should be measured in seconds.

NOTE In most audio communication systems over the digital networks, a buffering scheme is incorporated. Therefore, buffering time is also taken into account in the measurement.

6.3.3 Presentation of assessment result

The measured delay time should be reported in seconds.

7 Total quality

7.1 Synchronization of audio and video (lip sync)

7.1.1 Item to be assessed

Temporal synchronization between audio and video channels.

7.1.2 Method of assessment

The raison d'être of true multimedia systems, in contrast to a mere collection of unrelated media channels, is the ability to keep temporal synchronization among different channels. It is therefore vitally important to include a temporal synchronization quality measurement in the quality assessment items of audio-video communication systems.

The framework for measurement of temporal synchronization among media channels is given in ITU-T Recommendation P.931. Its basic assumption is that media signal can be captured at such interfaces as the camera output and the display input for the visual channel and the microphone output and the loudspeaker input for the audio channel. This assumption is shown in Figure 1.

Media signals at those interfaces are digitized, if necessary, broken into fixed sized frames, and given timestamps. For the details of this procedure, refer to ITU-T Recommendation P.931.

The audio and the video media streams being considered, digitized frames of these media streams are given sequence numbers as follows:

- $A(m)$ and $V(n)$ are the input audio and the video frames, respectively (m and n are the sequence numbers for each stream). It is assumed that they are associated in the sense that they correspond to the same event.

- $A'(p)$ and $V'(q)$ are the output audio and the video frames, respectively.
- $T_A(m)$ and $T'_A(n)$ are the timestamps for $A(m)$ and $A'(n)$, respectively. Timestamps for other frames are defined in the same way.

For each input frame, however not all the input frames are to be used as described in ITU-T P.931, the corresponding output frame is to be found. Since the media stream data is modified, distorted, dropped and reshaped, the matching process is non-trivial. For video frames, the method which utilizes the metrics of PSNRs assessed in Clause 5 will be used. For audio frames, a two-stage process, which employs the audio envelopes for a coarse stage and the power spectral densities for a fine stage, will be used. For details, refer to ITU-T P.931.

It is assumed here that the matching relationships between $A(m)$ and $A'(p)$, $A'(p)$, and $V'(n)$ and $V'(q)$ are established. Under this assumption, the time skew between the audio and video frames is given by the following equation (8):

$$S_{AV}(p, q) = O'_{AV}(p, q) - O_{AV}(m, n) \quad (8)$$

where $Q_{AV}(m, n) = T_A(m) - T_V(n)$ and $Q'_{AV}(p, q) = T'_A(p) - T'_V(q)$.

NOTE 1 It is important to choose appropriate input audio signals to obtain a valid and meaningful assessment result. If the video signal contains still or almost-still scenes, the process for matching the input and the output frames will become difficult or even impossible. A similar caution is to be applied to the assessment of the audio channel.

NOTE 2 Modern video compression schemes give fluctuating compression, transmission (when variable bit-rate encoding is employed) and decompression times, depending on the input properties. Therefore, the appropriate input signals suited for the assumed application should be used for assessment.

NOTE 3 For systems with a low video frame rate, it is sometimes natural and desirable to have a larger temporal skew between the video and the audio streams, because the video delay fluctuates while the audio data generally flows isochronously.

Selection of standard audio-video input streams suited for common usage is left for future study.

7.1.3 Presentation of assessment results

The report from the measurement should be expressed so that any variation between individual measurement is clearly illustrated. Classical summary statistics (for example, minimum, maximum, average and standard deviation) may also be reported.

7.2 Scalability

7.2.1 Item to be assessed

Autonomous function to tune frame rate dynamically depending on available bandwidth between the sender and the receiver.

7.2.2 Method of assessment

The method for measurement of scalability is left for future study.

7.2.3 Presentation of assessment results

Under consideration.

7.3 Overall quality

7.3.1 Item to be assessed

Overall quality factor in terms of audio and video interaction.

7.3.2 Method of assessment

Overall quality of audio-video communication systems OQ_{AV} should be defined as in equation (9).

$$OQ_{AV} = aQ_V + bQ_A + cQ_{V\&A} \quad (9)$$

where Q_V is the objective quality metric assessed in clause 5, Q_A is the objective quality metric assessed in clause 6, $Q_{V\&A}$ is the objective quality metric assessed in this clause; the coefficients a , b and c are the weighting factors, which depend on actual applications of the audio-video communication system.

7.3.3 Presentation of assessment results

The overall quality factor should be reported with sufficient information on the audio-video communication system under assessment.

IECNORM.COM : Click to view the full PDF of IEC TR 62251:2003

Annex A (informative)

PSNRs defined in three-dimensional spaces applied to hypothetical deterioration over the reference video sources

A.1 Introduction

This annex is intended to demonstrate the definitions PSNRs in three-dimensional vector space for each pixel that consists of a frame of videos. The definition of PSNR in the CIELAB is given in equation (5), PSNR in sYCC in equation (6), PSNR in sRGB in equation (4). The average colour difference defined in equation (1) is also included in this annex for comparison together with one-dimensional PSNRs in L^* and Y .

The values of the objective quality measures will be easily compared with other possible future measures and the results of subjective assessment of video quality.

A.2 Test sources and hypothetical deterioration

In this annex, 16 known different hypothetical deteriorations over the digital video files prepared in ITU-R BT.601-5 format and used in the Video Quality Expert Group (VQEG) are indicated. The source videos are labelled from SRC13_REF__525.yuv to SRC22_REF__525.yuv as shown in table A.1. They are made use of by the permission of the VQEG.

Software for varieties of objective measures have been developed in Chiba University, Japan, in collaboration with Mitsubishi Electric Corp. The values have been obtained for reduced frame size of 320 x 240 pixels per frame over 260 frames. In other words $P1=1$, $P2=260$, $M1=1$, $M2=240$ and $N1=1$, $N2=320$ in equation (B.4). Numerical results are shown in Tables A.2 to A.6.

Table A.1 – Reference video sources available for objective assessment

Designation	Name	Contents
SRC13_REF__525	Balloon-pops	Film, saturated colour, movement
SRC14_REF__525	New York 2	Masking effect, movement
SRC15_REF__525	Mobile & Calendar	Colour, movement
SRC16_REF__525	Betes_pas_betes	Colour, synthetic, movement, scene cut
SRC17_REF__525	Le_point	Colour, transparency, movement in all the directions
SRC18_REF__525	Autumn leaves	Colour, landscape, zooming, water fall movement
SRC19_REF__525	Football	Colour, movement
SRC20_REF__525	Sailboat	Almost still
SRC21_REF__525	Susie	Skin colour
SRC22_REF__525	Tempete	Colour, movement

Table A.2 – PSNRs in various colour spaces and the colour difference for SRC13 and SRC14

	Lab	sYCC	sRGB	L*	Y	ΔE		Lab	SYCC	sRGB	L*	Y	ΔE
hrc1/src13	20,5	23,2	23,6	26,3	26,3	8,3	hrc1/src14	22,4	25,8	25,9	26,6	28,1	7,5
hrc2/src13	23,6	23,5	23,2	25,9	25,0	5,4	hrc2/src14	25,7	24,3	24,5	25,4	24,3	4,9
hrc3/src13	22,2	22,7	22,3	25,6	24,6	5,8	hrc3/src14	24,9	23,8	24,0	25,1	24,1	4,7
hrc4/src13	21,4	22,1	21,7	25,6	24,7	7,4	hrc4/src14	24,6	23,9	24,0	25,4	24,3	5,5
hrc5/src13	20,4	19,3	19,0	21,2	20,3	8,0	hrc5/src14	22,5	19,7	20,0	20,7	19,6	5,9
hrc6/src13	22,2	22,6	22,1	25,9	24,9	6,2	hrc6/src14	24,5	23,6	23,8	25,3	24,1	5,0
hrc7/src13	22,2	21,1	20,7	23,0	22,1	5,9	hrc7/src14	24,5	21,5	21,7	22,5	21,4	4,1
hrc8/src13	21,9	22,3	21,9	25,3	24,5	6,7	hrc8/src14	24,3	23,5	23,7	24,9	24,0	5,3
hrc9/src13	21,6	20,6	20,3	22,8	21,8	6,9	hrc9/src14	24,3	21,4	21,7	22,4	21,4	4,5
hrc10/src13	22,1	20,9	20,6	23,0	22,0	6,3	Hrc10/src14	24,3	21,4	21,6	22,5	21,4	4,4
hrc11/src13	21,7	22,8	22,5	24,5	25,3	6,9	Hrc11/src14	25,5	26,0	26,1	24,6	26,3	4,1
hrc12/src13	22,4	23,6	23,3	24,8	26,0	5,9	Hrc12/src14	26,0	26,2	26,4	24,8	26,4	3,7
hrc13/src13	21,3	20,7	20,6	23,4	22,2	6,8	Hrc13/src14	21,5	20,8	21,7	23,2	21,7	5,4
hrc14/src13	21,2	20,3	20,0	22,7	21,6	7,9	Hrc14/src14	23,9	21,3	21,6	22,4	21,3	5,3
hrc15/src13	21,9	22,1	21,7	25,3	24,4	7,6	Hrc15/src14	25,8	25,8	26,0	27,2	26,3	5,6
hrc16/src13	22,1	22,8	22,3	25,8	25,2	7,0	Hrc16/src14	26,0	26,0	26,2	27,4	26,5	5,3

NOTE 1 hrc16/src14 and so on correspond to hypothetically degraded video (hrc16) from the reference source video (src14), respectively.

NOTE 2 All videos are in the size of 320 pixels x 240 pixels, each of which has a 24-bit colour depth.

Table A.3 – PSNRs in various colour spaces and the colour difference for SRC15 and SRC16

	Lab	sYCC	sRGB	L*	Y	ΔE		Lab	SYCC	sRGB	L*	Y	ΔE
hrc1/src15	11,8	13,6	13,1	20,7	19,5	24,8	hrc1/src16	20,3	21,6	21,8	23,8	25,7	9,5
hrc2/src15	17,0	18,5	18,4	24,2	23,1	10,8	hrc2/src16	27,1	28,1	28,0	31,1	32,0	4,4
hrc3/src15	15,1	16,7	16,5	23,1	21,7	13,2	hrc3/src16	29,2	29,0	28,9	31,0	31,9	2,4
hrc4/src15	13,6	15,0	14,5	23,0	21,2	18,7	hrc4/src16	22,9	23,7	23,6	28,3	28,3	6,0
hrc5/src15	14,6	15,4	15,2	19,8	18,9	15,9	hrc5/src16	21,7	22,0	21,9	24,8	25,5	6,0
hrc6/src15	14,0	15,4	15,0	23,0	21,3	17,3	hrc6/src16	23,5	24,2	24,0	28,7	28,7	5,1
hrc7/src15	16,1	17,0	16,9	21,1	20,3	12,1	hrc7/src16	22,8	22,9	22,8	25,7	26,4	4,4
hrc8/src15	14,0	15,3	15,0	22,6	20,9	17,6	hrc8/src16	23,4	24,2	24,0	28,4	28,5	5,2
hrc9/src15	15,9	16,6	16,5	20,7	19,7	13,1	hrc9/src16	22,8	22,7	22,7	25,5	26,1	4,6
hrc10/src15	16,4	17,3	17,2	22,1	21,0	11,9	hrc10/src16	24,9	25,8	25,4	28,9	30,3	3,8
hrc11/src15	15,8	17,2	17,0	22,9	22,0	12,8	hrc11/src16	25,4	27,5	27,3	27,8	31,6	3,8
hrc12/src15	16,0	17,5	17,3	23,3	22,7	12,0	hrc12/src16	25,7	27,9	27,6	28,0	32,2	3,5
hrc13/src15	15,4	16,2	16,1	21,5	19,7	14,6	hrc13/src16	23,3	23,5	23,6	29,1	29,5	4,3
hrc14/src15	15,6	16,1	16,0	20,6	19,2	14,3	hrc14/src16	22,9	22,6	22,6	25,2	25,4	5,2
hrc15/src15	15,7	16,1	16,0	21,1	19,3	15,4	hrc15/src16	23,7	23,3	23,5	26,0	26,2	5,8
hrc16/src15	15,7	16,3	16,2	21,6	19,9	14,9	hrc16/src16	23,9	23,5	23,7	26,2	26,5	5,6

NOTE 1 hrc16/src14 and so on correspond to hypothetically degraded video (hrc16) from the reference source video (src14), respectively.

NOTE 2 All videos are in the size of 320 pixels x 240 pixels, each of which has a 24-bit colour depth.

Table A.4 – PSNRs in various colour spaces and the colour difference for SRC17 and SRC18

	Lab	sYCC	sRGB	L*	Y	ΔE		Lab	sYCC	sRGB	L*	Y	ΔE
hrc1/src17	15,8	19,2	19,2	20,8	23,3	16,7	hrc1/src18	18,3	21,0	20,7	23,2	25,6	10,2
hrc2/src17	20,2	23,2	23,6	26,6	26,9	9,2	hrc2/src18	22,8	24,8	24,5	28,0	28,7	6,0
hrc3/src17	20,2	23,2	23,3	26,2	27,1	8,3	hrc3/src18	22,4	24,2	23,8	27,7	28,0	6,5
hrc4/src17	18,6	21,2	21,6	25,2	25,0	11,1	hrc4/src18	18,1	20,4	19,7	26,6	26,9	9,9
hrc5/src17	18,0	20,1	20,5	22,7	23,0	11,8	hrc5/src18	18,9	20,1	20,0	21,7	22,6	9,0
hrc6/src17	18,5	20,8	21,1	24,9	24,7	10,0	hrc6/src18	19,3	21,6	21,0	27,2	27,4	8,4
hrc7/src17	19,8	21,8	22,1	24,5	24,9	8,6	hrc7/src18	20,3	21,5	21,5	22,8	23,7	7,2
hrc8/src17	18,1	20,5	20,8	24,3	24,2	10,9	hrc8/src18	19,5	21,7	21,2	26,8	27,2	8,4
hrc9/src17	18,6	20,6	20,9	23,4	23,7	10,3	hrc9/src18	20,4	21,5	21,6	22,8	23,6	7,4
hrc10/src17	19,7	21,9	22,2	24,8	25,2	8,9	hrc10/src18	21,6	23,0	22,8	25,3	26,0	6,5
hrc11/src17	18,1	20,4	20,7	23,2	23,9	10,8	hrc11/src18	21,5	24,4	24,0	26,7	29,8	6,4
hrc12/src17	19,0	21,4	21,8	24,1	25,1	9,3	hrc12/src18	21,7	24,5	24,1	26,7	30,1	6,1
hrc13/src17	16,9	18,6	19,0	22,0	22,0	13,2	hrc13/src18	21,9	24,0	23,6	27,7	27,9	6,9
hrc14/src17	18,2	20,3	20,6	23,4	23,4	11,6	hrc14/src18	21,3	22,8	22,6	25,2	25,7	7,2
hrc15/src17	17,8	20,0	20,4	23,0	23,0	13,4	hrc15/src18	21,6	23,7	23,3	29,0	28,5	7,8
hrc16/src17	18,2	20,7	21,2	23,8	24,0	12,5	hrc16/src18	21,7	23,9	23,4	29,4	29,2	7,4

NOTE 1 hrc16/src14 and so on correspond to hypothetically degraded video (hrc16) from the reference source video (src14), respectively.

NOTE 2 All videos are in the size of 320 pixels x 240 pixels, each of which has a 24-bit colour depth.

Table A.5 – PSNRs in various colour spaces and the colour difference for SRC19 and SRC20

	Lab	sYCC	sRGB	L*	Y	ΔE		Lab	sYC C	sRG B	L*	Y	ΔE
hrc1/src19	20,0	22,6	22,6	23,2	25,6	7,8	hrc1/src20	15,8	17,4	17,2	20,1	20,2	12,7
hrc2/src19	23,6	25,1	24,9	27,9	28,6	4,8	hrc2/src20	20,6	20,8	20,7	23,7	22,1	6,9
hrc3/src19	23,1	24,6	24,4	27,7	28,0	5,8	hrc3/src20	18,7	19,3	19,3	22,6	21,3	8,2
hrc4/src19	20,3	22,5	22,1	26,6	27,1	6,9	hrc4/src20	18,7	19,2	19,0	22,6	21,0	8,8
hrc5/src19	19,8	20,8	20,7	22,5	23,3	7,4	hrc5/src20	18,7	16,2	16,0	18,5	16,6	8,3
hrc6/src19	20,6	22,7	22,3	27,0	27,2	6,6	hrc6/src20	18,8	19,4	19,2	23,1	21,4	8,1
hrc7/src19	21,0	21,7	21,7	22,8	23,5	5,9	hrc7/src20	19,4	17,5	17,3	19,6	18,1	7,2
hrc8/src19	20,7	22,6	22,3	26,6	26,9	6,8	hrc8/src20	18,6	19,2	19,0	22,8	21,2	8,4
hrc9/src19	21,2	23,1	22,7	27,2	27,2	6,2	hrc9/src20	20,0	20,3	20,1	23,5	22,0	6,4
hrc10/src19	20,0	21,3	21,1	24,4	24,7	7,8	Hrc10/src20	20,3	18,8	18,6	21,3	19,5	6,5
hrc11/src19	21,4	23,6	23,3	25,6	27,7	6,3	Hrc11/src20	19,9	21,5	21,4	23,3	23,8	6,6
hrc12/src19	22,4	24,7	24,4	25,9	28,8	5,4	Hrc12/src20	20,3	21,9	21,7	23,4	24,1	6,2
hrc13/src19	20,8	21,8	21,7	24,2	24,3	6,9	Hrc13/src20	19,8	18,8	18,7	21,4	19,8	7,8
hrc14/src19	21,1	22,1	22,0	24,8	25,1	7,1	Hrc14/src20	19,6	18,4	18,2	21,1	19,3	7,6
hrc15/src19	23,3	24,6	24,4	28,8	28,5	5,9	Hrc15/src20	19,5	20,3	20,2	23,5	22,2	8,4
hrc16/src19	23,6	25,1	24,8	29,4	29,4	5,4	Hrc16/src20	19,6	20,4	20,4	23,7	22,4	8,3

NOTE 1 hrc16/src14 and so on correspond to hypothetically degraded video (hrc16) from the reference source video (src14), respectively.

NOTE 2 All videos are in the size of 320 pixels x 240 pixels, each of which has a 24-bit colour depth.

Table A.6 – PSNR's in various colour spaces and the colour difference for SRC21 and SRC22

	Lab	sYCC	sRGB	L*	Y	ΔE		Lab	sYCC	sRGB	L*	Y	ΔE
Hrc1/src21	23,1	25,3	25,8	22,8	25,8	5,9	hrc1/src22	14,6	18,0	17,6	22,3	24,1	16,8
Hrc2/src21	29,3	29,1	29,2	28,5	29,6	3,2	hrc2/src22	18,9	21,7	21,0	26,3	26,5	9,3
Hrc3/src21	29,4	28,8	28,8	28,4	29,3	2,9	hrc3/src22	17,0	19,9	19,3	24,8	24,9	11,0
Hrc4/src21	28,4	27,7	27,9	27,3	28,2	3,5	hrc4/src22	17,4	20,1	19,4	25,4	25,6	11,2
Hrc5/src21	25,7	24,0	24,1	22,8	24,0	3,5	hrc5/src22	17,4	18,9	18,0	21,5	21,9	11,2
Hrc6/src21	29,5	28,3	28,5	27,9	28,6	2,8	hrc6/src22	17,2	20,0	19,3	25,7	25,8	10,8
Hrc7/src21	26,0	24,4	24,5	23,1	24,4	3,0	hrc7/src22	18,0	19,9	19,2	22,8	23,3	9,8
Hrc8/src21	29,1	28,1	28,3	27,5	28,4	3,0	hrc8/src22	17,2	19,9	19,2	25,1	25,2	11,1
Hrc9/src21	30,7	29,4	29,5	28,5	29,6	2,0	hrc9/src22	17,9	20,5	19,8	25,4	25,5	9,7
hrc10/src21	28,5	26,9	27,0	25,8	26,9	2,5	hrc10/src22	18,2	20,3	19,5	23,9	24,2	9,7
hrc11/src21	28,8	30,6	30,7	26,7	31,0	2,4	hrc11/src22	18,0	20,8	20,3	24,4	25,6	10,0
hrc12/src21	28,9	30,8	30,9	26,7	31,2	2,2	hrc12/src22	18,3	21,3	20,7	24,9	26,5	9,3
hrc13/src21	27,4	25,8	25,9	25,0	25,9	3,2	hrc13/src22	16,9	18,9	18,5	22,7	22,6	12,2
hrc14/src21	28,2	26,7	26,8	25,7	26,8	2,9	hrc14/src22	17,8	19,7	19,0	23,2	23,2	11,0
hrc15/src21	30,5	30,4	30,5	30,3	31,1	3,2	hrc15/src22	17,8	20,2	19,8	24,4	23,9	12,0
hrc16/src21	30,6	30,5	30,6	30,4	31,2	3,2	hrc16/src22	18,1	20,8	20,3	25,4	25,1	11,3

NOTE 1 hrc16/src14 and so on correspond to hypothetically degraded video (hrc16) from the reference source video (src14), respectively.

NOTE 2 All videos are in the size of 320 pixels x 240 pixels, each of which has a 24-bit colour depth.

Annex B (informative)

End-to-end objective assessment of video quality in the spatial frequency domain

B.1 Item to be assessed

The root mean square errors between corresponding blocks in the wavelet transformed domain corresponding to the reference video and the deteriorated video, which has been proposed in ITU-R 6Q/42-E.

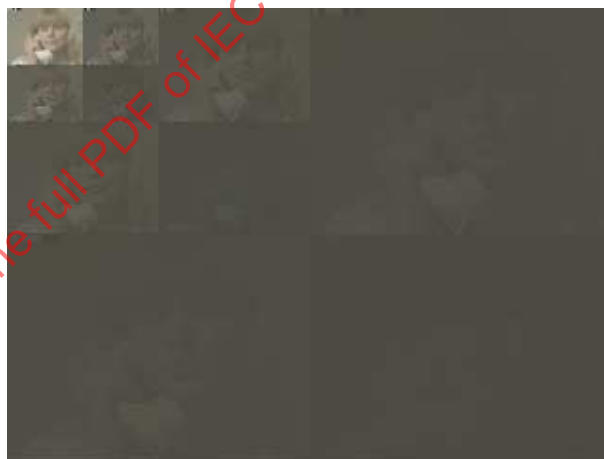
Three-level wavelet transform is assumed. Therefore, there are 10 blocks as shown in Figure B.1 and Figure B.2.

1
2
5
8

3
4

6
7

9
10



IEC 1496/03

IEC 1495/03

Figure B.1 – Assignment of the block numbers

Figure B.2 – Example of wavelet decomposition visualized

B.2 Method of assessment

Reference videos in Table A.1 are used as item 1 of Figure 2. Frame size reduced videos in uncompressed AVI-format should be prepared for item 2 of Figure 2. It is necessary to embed frame numbers at this point so that they can be used to identify received frames corresponding to the transmitted frames.

Encoded and transmitted streaming videos shall be continuously captured. Pixel-by-pixel calculations should be conducted.

The root mean square errors between each of corresponding blocks $p = 1...10$ to the original video frame and the deteriorated video frame k should be acquired as the following.

Let coefficients in the wavelet domain be $c_{R_{o,ijpk}}$, $c_{G_{o,ijpk}}$ and $c_{B_{o,ijpk}}$ for the position (i, j) of the block p of reference red, green, and blue pixel data, respectively; $c_{R_{d,ijpk}}$, $c_{G_{d,ijpk}}$ and $c_{B_{d,ijpk}}$ for the position (i, j) of the block p of deteriorated red, green, and blue pixel data, respectively.

Deterioration d_{pk} at the block p of the frame k in the wavelet domain should be evaluated by the sum square error as in equations (B.1) and (B.2).

$$d_{pk} = \sum_i \sum_j \left(\Delta c_{R_{ijpk}}^2 + \Delta c_{G_{ijpk}}^2 + \Delta c_{B_{ijpk}}^2 \right) \quad (B.1)$$

where

$$\begin{aligned} \Delta c_{R_{ijpk}} &= c_{R_{d,ijpk}} - c_{R_{o,ijpk}} \\ \Delta c_{G_{ijpk}} &= c_{G_{d,ijpk}} - c_{G_{o,ijpk}} \\ \Delta c_{B_{ijpk}} &= c_{B_{d,ijpk}} - c_{B_{o,ijpk}} \end{aligned} \quad (B.2)$$

B.3 Presentation of assessment results

The metric of the sum of square errors between blocks corresponding to the wavelet transformed frames should be plotted versus frame numbers as shown in Figure B.3 together with identifications of reference video sources. The conditions of measurement such as frame size in pixels, frame rate, streaming bit-rate should also be reported.

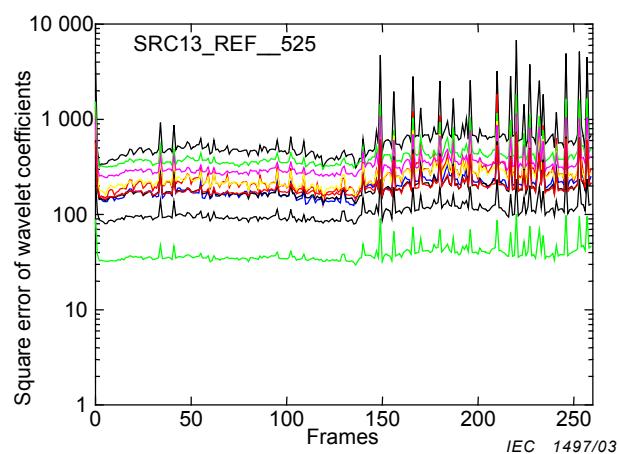


Figure B.3a – Example for SRC13_REF_525

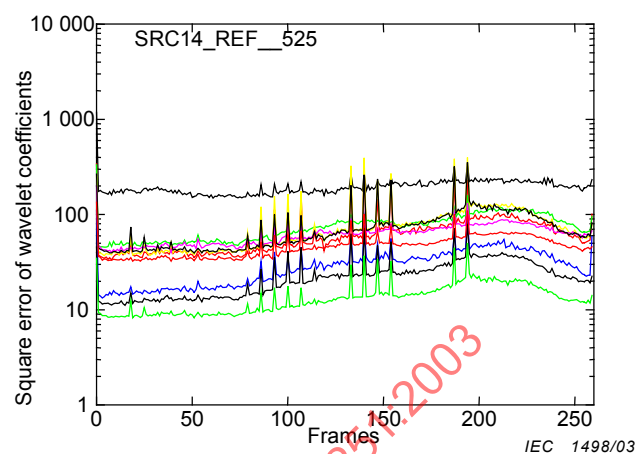


Figure B.3b – Example for SRC14_REF_525

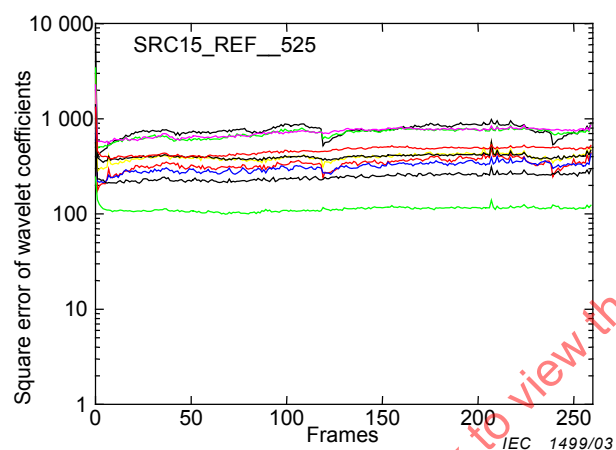


Figure B.3c – Example for SRC15_REF_525

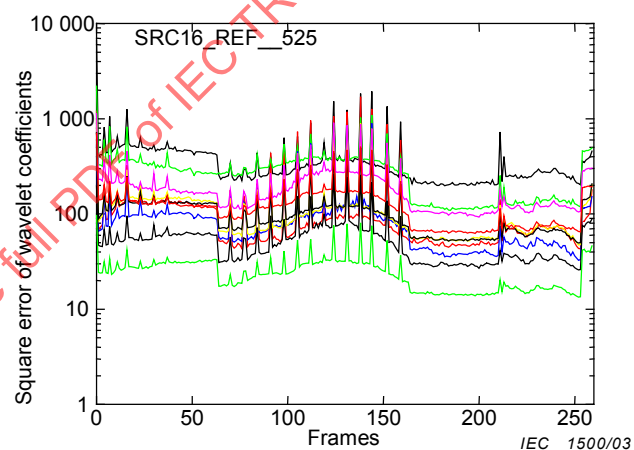


Figure B.3d – Example for SRC16_REF_525

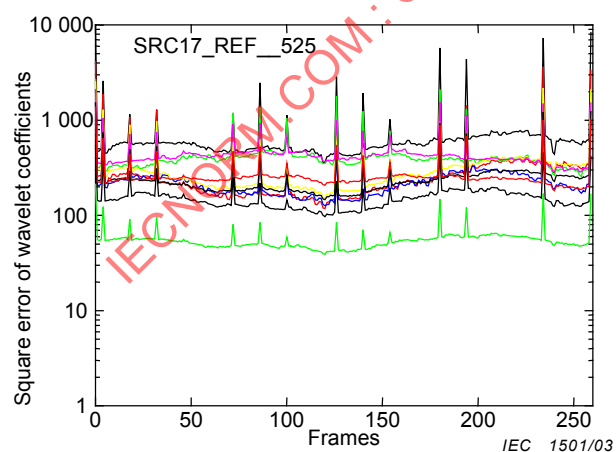


Figure B.3e – Example for SRC17_REF_525

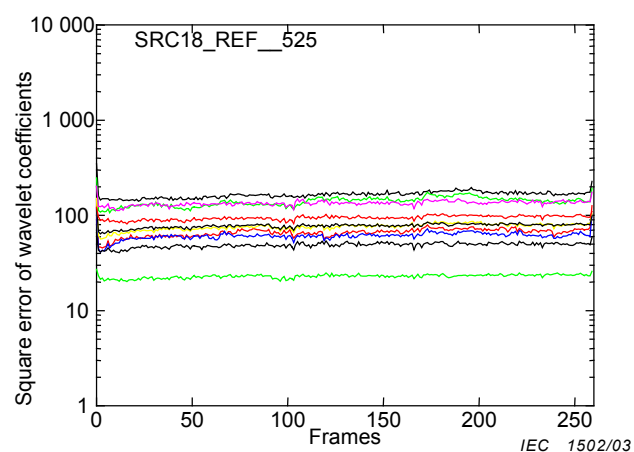


Figure B.3f – Example for SRC18_REF_525