



Recommendation ITU-R BT.1907
(01/2012)

**Objective perceptual video quality
measurement techniques
for broadcasting applications
using HDTV in the presence
of a full reference signal**

BT Series
Broadcasting service
(television)



International
Telecommunication
Union
Documents

Document Sharing Hub

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from <http://www.itu.int/ITU-R/go/patents/en> where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at <http://www.itu.int/publ/R-REC/en>)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.

Electronic Publication
Geneva, 2012

© ITU 2012

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

RECOMMENDATION ITU-R BT.1907

**Objective perceptual video quality measurement techniques
for broadcasting applications using HDTV in the presence
of a full reference signal**

(2012)

Scope

This Recommendation specifies methods for estimating the perceived video quality of broadcasting applications using HDTV when a full reference signal is available.

The ITU Radiocommunication Assembly,

considering

- a) that the ability to automatically measure the quality of broadcast video has long been recognized as a valuable asset to the industry;
- b) that Recommendation ITU-R BT.1683 describes objective methods for measuring the perceived video quality of standard definition digital broadcast television in the presence of a full reference;
- c) that Recommendation ITU-R BT.709 describes parameter values for the HDTV standards for production and international programme exchange and Recommendation ITU-R BT.710 describes subjective assessment methods for image quality in high-definition television;
- d) that HDTV is becoming widely used in broadcasting;
- e) that ITU-T Study Group 9, based on the results of the HDTV report sent by VQEG, has produced Recommendation ITU-T J.341, which specified objective perceptual video quality measurement of HDTV in the presence of a full reference;
- f) that objective measurement of the perceived video quality of HDTV may complement subjective assessment methods,

recommends

that the objective video quality model given in Annexes 1, 2 and 3 should be used for objective measurement of perceived video quality for broadcasting applications using HDTV in the presence of a full reference signal.

Annex 1

1 Introduction

This Recommendation provides a perceptual video quality measurement method for use in high definition television (HDTV) non-interactive applications when the full reference (FR) measurement method can be used. The model was developed to estimate subjective quality scores.

The full reference measurement method can be used when the unimpaired reference video signal is readily available at the measurement point, as may be the case of measurements on individual equipment or a chain in the laboratory or in a closed environment such as a television broadcast station. The estimation method includes both calibration and objective video quality estimations.

The validation test material contained both H.264 and MPEG-2 coding degradations and various transmission error conditions (e.g. bit errors, dropped packets). The model in this Recommendation may be used to monitor the quality of deployed networks to ensure their operational readiness. The visual effects of the degradations may include spatial as well as temporal degradations. The model in this Recommendation can also be used for lab testing of video systems. When used to compare different video systems, it is advisable to use a quantitative method (such as that in ITU-T J.149) to determine the model's accuracy for that particular context.

This Recommendation is deemed appropriate for broadcasting services delivered between 1 Mbit/s and 30 Mbit/s. The following resolutions and frame rates were considered in the validation test:

- 1080/59.94/I
- 1080/25/P
- 1080/50/I
- 1080/29.97/P

The following conditions were allowed in the validation test for each resolution:

Test factors
Video resolution: 1920x1080 interlaced and progressive
Video frame rates 29.97 and 25 frames per second
Video bitrates: 1 to 30 Mbit/s
Temporal frame freezing (pausing with skipping) of maximum 2 seconds
Transmission errors with packet loss
Conversion of the SRC from 1080 to 720/P, compression, transmission, decompression, and then conversion back to 1080.
Coding technologies
H.264/AVC (MPEG-4 Part 10)
MPEG-2

Note that 720/P was considered in the validation test plan as part of the test condition (HRC). Because currently 720/P is commonly up-scaled as part of the display, it was felt that 720/P HRCs would more appropriately address this format.

1.1 Application

The applications for the estimation model described in this Recommendation include, but are not limited to:

- 1) Potentially real-time, in-service quality monitoring at the source;
- 2) Remote destination quality monitoring when a copy of the source is available at the point of measurement;
- 3) Quality measurement for monitoring of a storage or transmission system that utilizes video compression and decompression techniques, either a single pass or a concatenation of such techniques;
- 4) Lab testing of video systems.

1.2 Limitations

The video quality estimation model described in this Recommendation cannot be used to replace subjective testing. Correlation values between two carefully designed and executed subjective tests (i.e. in two different laboratories) normally fall within the range 0.95 to 0.98. If this Recommendation is utilized to make video system comparisons (e.g. comparing two codecs), it is advisable to use a quantitative method to determine the model's accuracy for that particular context.

When frame freezing was present, the test conditions typically had frame freezing durations less than 2 seconds. The model in this Recommendation was not validated for measuring video quality in a re-buffering condition (i.e. video that has a steadily increasing delay or freezing without skipping). The model was not tested on other frame rates than those used in TV systems (i.e. 29.97 frames per second and 25 frames per second, in interlaced or progressive mode).

It should be noted that in case of new coding and transmission technologies producing artefacts which were not included in this evaluation, the objective model may produce erroneous results. Here, a subjective evaluation is required.

2 References

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 Subjective assessment (picture): The determination of the quality or impairment of programme like pictures presented to a panel of human assessors in viewing sessions.

3.1.2 Objective perceptual measurement (picture): The measurement of the performance of a programme chain by the use of programme-like pictures and objective (instrumental) measurement methods to obtain an indication that approximates the rating that would be obtained from a subjective assessment test.

3.1.3 Proponent: An organization or company that proposes a video quality model for validation testing and possible inclusion in an ITU Recommendation.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 Frame rate: The number of unique frames (i.e. total frames – repeated frames) per second.

3.2.2 Simulated transmission errors: Errors imposed upon the digital video bit stream in a highly controlled environment. Examples include simulated packet loss rates and simulated bit errors. Parameters used to control simulated transmission errors are well defined.

3.2.3 Transmission errors: Any error imposed on the video transmission. Example types of errors include simulated transmission errors and live network conditions.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR	Absolute Category Rating (see ITU-R BT.500)
ACR-HR	Absolute Category Rating with Hidden Reference (see ITU-T P.910)
AVI	Audio Video Interleave
DMOS	Difference Mean Opinion Score
FR	Full Reference
FRTV	Full Reference TeleVision
HRC	Hypothetical Reference Circuit
ILG	VQEG's Independent Laboratory Group
MOS	Mean Opinion Score
MOSp	Mean Opinion Score, predicted
NR	No (or Zero) Reference
PSNR	Peak Signal-to-Noise Ratio
PVS	Processed Video Sequence
RMSE	Root Mean Square Error
RR	Reduced Reference
SFR	Source Frame Rate
SRC	Source Reference Channel or Circuit
VQEG	Video Quality Experts Group
YUV	Colour Space and file format

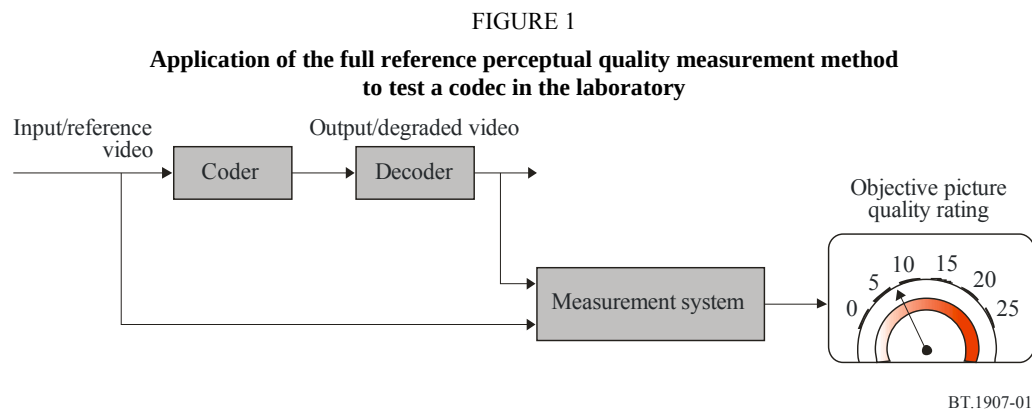
5 Conventions

None.

6 Description of the full reference methodology

The double-ended measurement method with full reference, for objective measurement of perceptual video quality, evaluates the performance of systems by making a comparison between the undistorted input, or reference, video signal at the input of the system, and the degraded signal at the output of the system (Fig. 1).

Figure 1 shows an example of application of the full reference method to test a codec in the laboratory.

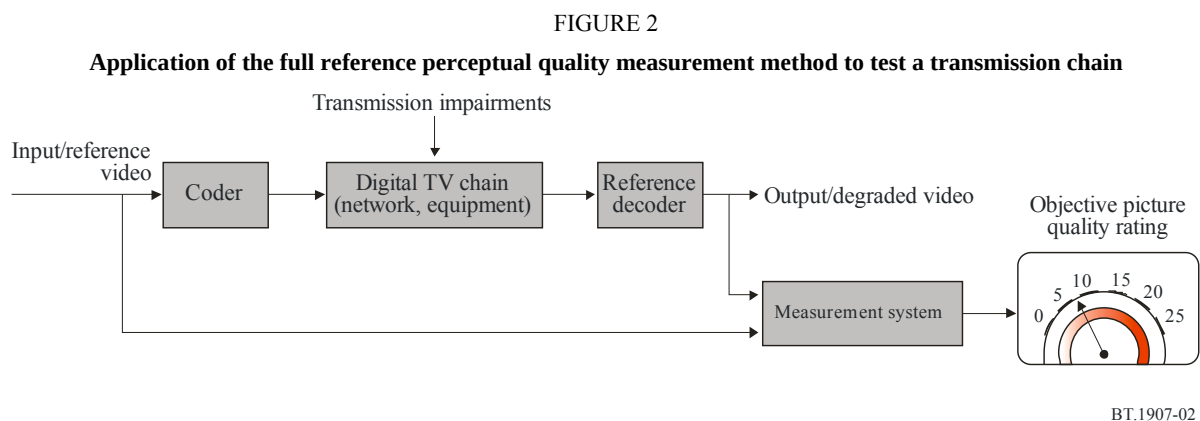


The comparison between input and output signals may require a temporal alignment or a spatial alignment process, the latter to compensate for any vertical or horizontal picture shifts or cropping. It also may require correction for any offsets or gain differences in both the luminance and the chrominance channels. The objective picture quality rating is then calculated, typically by applying a perceptual model of human vision.

Alignment and gain adjustment is known as registration. This process is required because most full reference methods compare reference and processed pictures on what is effectively a pixel-by-pixel basis. The video quality metrics described in Annex 2 include registration methods.

As the video quality metrics are typically based on approximations to human visual responses, rather than on the measurement of specific coding artefacts, they are in principle equally valid for analogue systems and for digital systems. They are also in principle valid for chains where analogue and digital systems are mixed, or where digital compression systems are concatenated.

Figure 2 shows an example of the application of the full reference method to test a transmission chain.



In this case, a reference decoder is fed from various points in the transmission chain, e.g. the decoder can be located at a point in the network, as in Fig. 2, or directly at the output of the encoder, as in Fig. 1. If the digital transmission chain is transparent, the measurement of objective picture quality rating at the source is equal to the measurement at any subsequent point in the chain.

It is generally accepted that the full reference method provides the best accuracy for perceptual picture quality measurements. The method has been proven to have the potential for high correlation with subjective assessments made in conformity with the ACR-HR methods specified in ITU-T P.910.

7 Findings of the Video Quality Experts Group (VQEG)

Studies of perceptual video quality measurements are conducted in an informal group, called the Video Quality Experts Group (VQEG), which reports to ITU-T Study Groups 9 and 12 and ITU-R Study Group 6. The recently completed high definition television phase I test of VQEG assessed the performance of proposed full reference perceptual video quality measurement algorithms.

The following statistics are taken from the final VQEG HDTV report. Note that the body of the VQEG HDTV report includes other metrics including Pearson Correlation and RMSE calculated on individual experiments, confidence intervals, statistical significance testing on individual experiments, analysis on subsets of the data that include specific impairments (e.g. H.264 coding-only), scatter plots, and the fit coefficients.

Primary analysis

The performance of the FR model is summarized in Table 1. The PSNR is calculated according to ITU-T J.340 and included in this analysis for comparison purposes. “Superset RMSE” identifies the primary metric (RMSE) computed on the aggregated superset (i.e. all six experiments mapped onto a single scale). “Top performing group total” identifies the number of experiments (0 to 6) for which this model was either the top performing model or statistically equivalent to the top performing model. “Better than PSNR total” identifies the number of experiments (0 to 6) for which the model was statistically better than PSNR. “Better Than Superset PSNR” lists whether each model is statistically better than PSNR on the aggregated superset. “Superset Correlation” identifies the Pearson Correlation computed on the aggregated superset.

TABLE 1

Metric	PSNR	SwissQual
Superset RMSE	0.71	0.56
Top performing group total	1	5
Better than PSNR total	—	4
Better than superset PSNR	—	Yes
Superset correlation	0.78	0.87