REPORT ITU-R BT.2020-1

# OBJECTIVE QUALITY ASSESSMENT TECHNOLOGY IN A DIGITAL ENVIRONMENT

(1999-2000)


## Summary

This is a revision of the Report ITU-R BT.2020 on the status of the technology for objective quality assessment of audio and video.

The Radiocommunication Joint Working Party (JWP) 10-11Q inherited from Working Parties (WP) 11E and 10C and Task Group (TG) 10-4 the task to define Recommendations on quality assessment. This Report addresses more particularly Question ITU-R 64-4/11 – Objective picture quality parameters and associated measurement and monitoring methods for television images. This Question reflects the actual interest of the broadcasting community in techniques for objective quality assessment and monitoring of broadcast audio and video. Digital television and radio are now in operation in several countries hence the demand for Recommendations is increasing.

Considerable progress has been achieved by the completion of Recommendation ITU-R BS.1387 – Method for objective measurements of perceived audio quality – on objective quality assessment of digital audio. Additional work is planned on monitoring methods for digital audio. The Video Quality Experts Group (VQEG) has undertaken an extensive set of validation tests for full reference double-ended objective picture quality measurement methods. But there is still a lot of work to be done on all methodological approaches to objective picture quality assessment of video.

This Report is a first step towards resolving the remaining open issues. The Report is structured as follows:

– § 1: Evolution of measurement techniques from analogue to compressed digital.

– § 2: Review of Recommendations.

– § 3: Review of on-going activities and developments.

– § 4: JWP 10-11Q approach to the definition of future Recommendations.

– § 5: Current conclusions on targets and priorities for future Recommendations.

JWP 10-11Q intends to collaborate with other Study Groups (SGs) and WPs in order to better appraise the overall situation in digital measurement and possibly avoid the dissemination of similar but different solutions. It is believed that the identification of applications and requirements is the most reasonable way to achieve that goal.

This Report will be maintained to take into account new requirements and to keep track of the evolution in digital objective quality evaluation.


## 1 Evolution of measurement techniques from analogue to compressed digital

This paragraph briefly explains the evolution of measurement from the use of indirect signal analysis to direct analysis of the content.

The well known logistic functions (e.g. the vertical blanking interval (VBI) test lines) that have allowed the design and monitoring of analogue TV are no longer valid for the following reasons:

– The signal structure for broadcast transmission has changed. It is now based on the use of digital transport streams for which protocol analysers have been developed.

– Digital delivery requires compression to be effective using complex non-linear encoding techniques. The use of such non-linear techniques impedes the use of traditional test signal analysis.

– Quality is now strongly content dependent and therefore time varying, which adds another level of complexity.

For these reasons there is low correlation between classic indirect objective measurements and the related video and audio quality.

Possible solutions are a combination of digital stream and picture content analysis. The first one is relatively easy to handle as the system behaviour and features are perfectly defined in specifications. As a consequence new objective picture quality assessment models have been developed. Digital objective quality evaluation now relies on feature extraction and perceptual model processing or some combination of both (thereby taking simultaneously into account the encoding processes and the characteristics of human perception).

The following is a preliminary list of measurement applications addressed by this Report:

– Codec and statistical multiplexers development, evaluation and installation.

– In-service and out-of-service network monitoring.

– Quality assessment of compressed production material.

– Monitoring of generic input material.

– Real time continuous monitoring.

It is therefore envisaged to recommend specific models on which measurement equipment would be developed for quality assessment and monitoring. It is currently admitted that different models could be adopted for different application specific domains.

# 2 Review of Recommendations

## 2.1 Existing Recommendations and Reports

*Audio*: Recommendation ITU-R BS.1387 – Method for objective measurements of perceived audio quality.

*Video*: ANSI [1996].

ITU-T Recommendation J.143, User requirements for objective perceptual video quality measurements in digital cable television.

Report ITU-R BT.2020, Objective quality assessment technology in a digital environment.

## 2.2 Planned Recommendations

*Video*

Radiocommunication JWP 10-11Q – Objective assessment of video quality; in cooperation with the Video Quality Experts Group (VQEG).

Telecommunication Standardization SG 9 preliminary draft new ITU-T Recommendation J.OVQ – Fullref, Perceptual video quality measurement techniques in the presence of a full reference.

Telecommunication Standardization SG 9 preliminary draft new Recommendation J.OVQ – Redref, Perceptual video quality measurement techniques in the presence of a reduced reference.

Telecommunication Standardization ITU-T SG 9 preliminary draft new Recommendation J.OVQ – Noref, Perceptual video quality measurement techniques in the absence of a reference.

Telecommunication Standardization SG 12 draft new ITU-T Recommendation P.OVQ – Objective assessment of video quality (full reference); in cooperation with the Video Quality Experts Group (VQEG) this study item relates to video quality assessment at bit rates of 768 kbit/s and higher.

Telecommunication Standardization SG 12 draft new ITU-T Recommendation P.RSQ – Reduced source bandwidth double-ended objective video quality assessment; this class of measurement is needed when the source and compressed video are not available at the same location.

Telecommunication Standardization SG 12 draft new ITU-T Recommendation P.LBQ – Objective video quality assessment at low bit rates (~16 kbit/s to 1.5 Mbit/s); this study item will cover low bit rate videoconferencing and multimedia applications.

Telecommunication Standardization SG 12 draft new ITU-T Recommendation P.TRQ – Objective video quality assessment with transmission impairments on packet, mobile and other networks.

# 3 Review of on-going activities and developments

This Report includes a review of the state of the art concerning quality assessment in the digital environment and identification of the main digital methodological approaches. The different approaches are defined based on the definitions developed by ITU-T SG 9 in Recommendation ITU-T J.143.

## 3.1 Identification of the main digital methodological approaches

### 3.1.1 Double-ended systems

A generic double-ended system is designed to operate, with two inputs, one the reference material and one the material under test. Usually these systems are not necessarily required to operate in real time and may work only with a limited library. The aim of these systems is basically the assessment (or the ranking) of the performance of digital codecs; nevertheless they can be used to assess the quality provided by a complete digital delivery chain that includes coding, transmission and decoding. The quality indication of these kinds of systems is usually expected to be the most accurate.

### 3.1.2 Double-ended systems using reduced reference

These systems are tailored to provide monitoring of the performance of a digital transmission network. The main feature of these systems is the ability to assess the quality in real time and in service without the use of dedicated reference signal. The quality information is collected at the entrance of the network and delivered to any nodal point together with the signal. At the nodal point where quality is to be assessed, the quality information is recalculated locally and compared to the received information to perform the quality check. The quality indicators provided by these systems may be not as accurate as in the case of the double-ended systems (with complete reference). These systems provide an indication of the “availability” of the service guaranteed by the “transparency” of the transmission process.

### 3.1.3 Single-ended systems

This family of systems is based on the analysis of existing material “as it is”. The origin of the impairment is not known and it is difficult to go beyond some limitations. Basically the single-ended systems look for some particular *a priori* impairments possibly originated by a generic digital coder or due to some discontinuities on a digital transmission link. Also for these reasons the quality indicators provided by these systems are limited in performance and at the present time do not cover all the possible impairments. These systems can also be used to provide an indication of the “availability” of the service.

## 3.2 Status of the systems currently available or proposed

The VQEG is evaluating some of the existing double-ended objective picture quality measurement methods. The VQEG is an informal organization encouraged by Radiocommunications SG 11, JWP 10-11Q, Telecommunication Standardization SGs 9 and 12.

Table 1 summarizes the current known situation. All have been classified according to their family (D = double-ended; S = single-ended; RRD = reduced reference double-ended).

Studies are planned inside the Institute of Electrical and Electronics Engineers (IEEE) to provide a pool of test scenes degraded in a controlled way. Each scene will have a corresponding perceptual scale associated with it, that is, calibrated in successive steps of just-noticeable-differences of impairment. These scenes will hopefully represent a good pool of reference material to test the forthcoming systems.

TABLE 1

| Company or Laboratory | Partner | Country | Audio | Video | Methodology | Real time | In service | Commercial product | VQEG test |
|---|---|---|---|---|---|---|---|---|---|
| CCETT | | France | | X | S | | | X | |
| CCETT | | France | X | | D/S | | | X | |
| CRC | | Canada | X | | D | | | X | |
| CRC[1] | | Canada | X | | D | | | X | |
| FHG[1] | Opticom | Germany | X | | D | X | | X | |
| KDD | | Japan | | X | D | | | X | X |
| KPN[1] | <br>Opticom | Netherlands<br>Germany | | X | D | | | X | |
| Mitsubishi | NHK | Japan | | X | D | | | X | X |
| Opticom | | Germany | X | | D | | | X | |
| Rohde & Schwarz | IFN | Germany | | X | S | | X | X | X |
| Snell & Wilcox | | UK | | X | S | | | X | |
| TDF | | France | X | X | RRD | | | X | |
| Tektronix | Sarnoff | USA | | X | D | | | X | X |
| Tektronix | | USA | | X | S | X | X | X | |
| | | | | | | | | | |
| ECI Telecom | | Israel | | X | D | | | | |
| CPqD | | Brazil | | X | D | | | | X |
| EPFL | | Switzerland | | X | D | | | | X |
| KPN | Swiss Telecom | Netherlands<br>Switzerland | | X | D | | | | X |
| NASA | | USA | | X | D | | | | X |
| NTIA | | USA | | X | RRD | | | | X |
| Tapestries | EC ACTS | European Consortium | | X | D | | | | X |

[1] These products were produced and sold prior to completing the PEAQ standard (Recommendation ITU-R BS.1387 – Method for objective measurements of perceived audio quality). Some of these products are still commercially available.

## 3.3 Objective video quality: Current VQEG status

The draft VQEG Report (Document 10-11Q/56, 21 January 2000) describes the results of the evaluation process of objective video quality models as submitted. Each of ten proponents submitted one model to be used in the calculation of objective scores for comparison with subjective evaluation over a broad range of video systems and source sequences. Over 26 000 subjective opinion scores were generated based on 20 different source sequences processed by 16 different video systems and evaluated at eight independent laboratories worldwide. The subjective tests were organized into four quadrants: 50 Hz high quality, 50 Hz low quality, 60 Hz high quality and 60 Hz low quality. High quality in this context refers to production quality video and low quality refers to distribution quality. The high quality quadrants included video at bit rates between 3 Mbit/s and 50 Mbit/s. The low quality quadrants included video at bit rates between 768 kbit/s and 4.5 Mbit/s. Strict adherence to Recommendation ITU-R BT.500 procedures for the double stimulus continuous quality scale (DSCQS) method was followed in the subjective evaluation. The subjective and objective test plans included procedures for validation analysis of the subjective scores and four metrics for comparing the objective data to the subjective results.

Depending on the metric that is used, there are seven or eight models (out of a total of nine) whose performance is statistically equivalent. The performance of these models is also statistically equivalent to that of power signal-to-noise ratio (PSNR). PSNR is a measure that was not originally included in the test plans but it was agreed later to include it as a reference objective model. It was also discussed and determined that three of the models did not generate proper values due to software or other technical problems.

In addition to analysis based on the total data set, subsets based on the four subjective test quadrants and the total data with exclusion of certain video processing systems were analysed to determine sensitivity of results to various application-dependent parameters.

Based on this analysis, the VQEG is not presently prepared to propose one or more models for inclusion in ITU Recommendations on objective picture quality measurement. Although the VQEG is not in a position to validate any models, the test was a great success. One of the most important achievements of the VQEG effort is the collection of an important new data set. Up until now, model developers have had a very limited set of subjectively rated video data with which to work. Once the current VQEG data set is released, future work is expected to dramatically improve the state of the art of objective measures of video quality.

### 3.4 Proposed reference model for in-service video quality monitoring

As stated above, three methodologies representing different measurement strategies for the assessment of the quality of video have been defined:

– methodology using the complete video reference (double-ended);

– methodology using reduced reference information (double-ended);

– methodology using no reference signal (single-ended).

JWP 10-11Q believes the design and the development of a video quality monitor should consider a general structure of the measurement procedure for reduced reference and single-ended methodologies (Document 10-11Q/57, 26 January 2000). The reference model is composed of the following four layers:

– *Measurement methodology* defines the class or the strategy relative to the application requirement;

– *Measurement method* is composed of a set of modules, algorithmic and associated ones, implemented to process inputs such as original signals or processed reference data, and provide output results such as processed reference data, level of impairment or final quality notation;

– *Algorithmic module(s)* is the basic block of signal processing functions composing the method. It composes the core of the method from which the final objective qualification is delivered;

– *Associated module(s)* is an additional function that aids the algorithmic module(s) in its operation by addressing such issues as dating, synchronization, presentation of data, etc.

Details of this approach are fully described in Annex 1. JWP 10-11Q recommends this reference model for use by Telecommunication Standardization SGs 9 and 12 and the VQEG. Future adaptations of this generic model may be appropriate to accommodate other applications. The application of this model to audio applications is under consideration.

## 4 JWP 10-11Q approach to the definition of future Recommendations

### 4.1 Review of requirements for application specific areas

JWP 10-11Q has prepared and distributed a questionnaire to seek the requirements and needs from the broadcasting community in objective quality assessment technology.

In the analysis of the data collected from the questionnaire, it is considered to be important to take into account the nature (or the role) of the company providing the answers. For this reason the resulting data will be separately considered taking into account the following preliminary categories:

– Broadcasters.

– Network providers.

– Regulation bodies.

The questionnaire produced a lot of interest among the broadcasters, with more than 20 answering. The responses indicate that quality of service (QoS) is one of their highest priorities and needs to be urgently addressed by the relevant standards bodies.

Upon examination of the responses, the following considerations stand out:

– All the answers except one came from European companies or organizations; suggesting that another request be submitted to companies and organizations of Region 2.

– From the content of the responses the priorities of the broadcasters can be determined.

– Generally the responses contained detailed comments on issues not explicitly covered by the questionnaire items. These details must be carefully analysed and reported in a proper fashion.

– Only one response addressed the requirement to check the quality of a mux or codec.

– What emerges as the highest priority of the broadcasters is monitoring of the whole chain from production to distribution. The following points have to be considered MANDATORY points for monitoring:

  – the output of the production chain (distribution ready);

  – all nodes of the transmission network;

  – end-user.

The above considerations may suggest requirements for “reduced-reference” and “single-ended” systems more than “double-ended” ones. It can be concluded based on this preliminary analysis that broadcasters are extremely interested in obtaining standards and related products designed to address the above issues.

It is expected that more information on user requirements will be made available for consideration at the JWP 10-11Q September 2000 meeting.

### 4.2 Coordination with other SGs and WPs including Telecommunication Standardization SGs 9 and 12

The proposed reference model provides a solid foundation for coordinating technical approaches and future recommendations to implement the reduced reference and single-ended methodologies. Liaison statements to Telecommunication Standardization SGs 9 and 12 and the VQEG will request that they all adopt this approach to develop common technical specifications for ITU Recommendations for objective measurement of picture quality. JWP 10-11Q intends to aggressively seek contributions from other SGs, WPs and Sector Members to provide proposed technical specifications for all or part(s) of the reference model. Contributions are encouraged to be submitted as soon as possible for consideration at the September 2000 meeting of JWP 10-11Q and no later than the following meeting in spring 2001.

## 5 Current JWP 10-11Q position

JWP 10-11Q recognizes the excellent work accomplished by the VQEG towards evaluation of the full reference double-ended methodology. It is now recognized that there is an urgent need for reduced reference and single-ended methodologies to be used for operational monitoring of broadcast television and associated multimedia services.

For the future work of the VQEG, JWP 10-11Q supports the idea to evaluate single-ended, double-ended and reduced reference objective video quality measurement models in parallel but recommends that the investigation and validation of reduced reference and single-ended systems be given the highest priority.

Since monitoring of the quality of video signals e.g. over satellite, cable terrestrial and broadband digital network distribution systems requires compliant methods of objective assessment, all standardizing bodies should agree on one common method for each methodological approach and application specific area.

Therefore the reference model described in Annex 1 should be used as a framework for the development and evaluation of proposed measurement methods covering the three methodologies. This recommended reference model provides a solid foundation for coordinating technical approaches and future recommendations to implement the reduced reference and single-ended methodologies. A liaison statement to Telecommunication Standardization SGs 9 and 12 and the VQEG requests that they all adopt this approach to develop common technical specifications for ITU Recommendations for objective measurement of picture quality. JWP 10-11Q encourages collaboration to be set up within the VQEG for the definition of one or several common solutions.

JWP 10-11Q intends to aggressively seek contributions from SGs, WPs and Sector Members to provide proposed technical specifications for all or part(s) of the reference model. Contributions are encouraged to be submitted as soon as possible for consideration at the September 2000 meeting of JWP 10-11Q and no later than the following planned meeting in spring 2001.

However, haste must not overtake the proper evaluation of the objective methods. Considering the importance of this matter, JWP 10-11Q plans to allocate significant resources to the preparation of the decision process that will follow the delivery of future VQEG results.

REFERENCES

ANSI [1996] ANSI T1.801.03, Digital transport of one-way video signals – Parameters for objective performance assessment. American National Standards Institute. United States of America.

ANNEX 1

## Method for in-service video quality monitoring over digital television (DTV) broadcasting networks

### 1 Introduction

The need for Recommendations on objective methods for video quality evaluation has been recognized both by ITU-R and ITU-T and they both have set up Questions of studies related to this topic (Q. ITU-R 64/11, ITU-T Q.11/12 and ITU-T Q.22/9).

The management of QoS over DTV networks imposes the choice of the adequate technique for video quality measurement. Some requirements have already been identified as crucial:

– quality of incoming signals;

– in service, real time, and continuous monitoring of the QoS on network in operations, etc.

This Annex describes the recommended generic method for the assessment of video quality over digital television broadcasting networks.

### 2 General concepts

The design and the development of a video quality meter bring to consider a general structure of the measurement procedure. Several layers compose this structure.

– *Measurement methodology* defines the class or the strategy relative to the application requirement;

– *Measurement method* is composed of a set of modules, algorithmic and associated ones, implemented to process inputs such as original signals or processed reference data, and provide output results such as processed reference data, level of impairment or final quality notation;

– *Algorithmic module(s)* is the basic block of signal processing functions composing the method. It composes the core of the method from which the final objective qualification is delivered;

– *Associated module(s)* is an additional function that aids the algorithmic module(s) in its operation by addressing such issues as dating, synchronization, presentation of data, etc.

# 3 Technical approach for the QoS network monitoring

## 3.1 Measurement methodology

Three classes representing different measurement strategies for the assessment of the quality of video signals have been defined by Telecommunication Standardization SG 9 (Temporary Document 36. J.OVQ – Methodologies for video quality assessment on networks in operation) and adopted by Radiocommunication JWP 10-11Q:

– methodology using the complete video reference;

– methodology using reduced reference information;

– methodology using no reference signal.

Each one of those methodologies offers a specific approach relative to its operational use. Each methodology is well adapted to a specific field of applications. This is due to its technical implementation constraints:

– relevance of the measured data,

– availability of the reference,

– synchronization of original and impaired signals or data,

– transmission channel for reference data,

– real-time implementation,

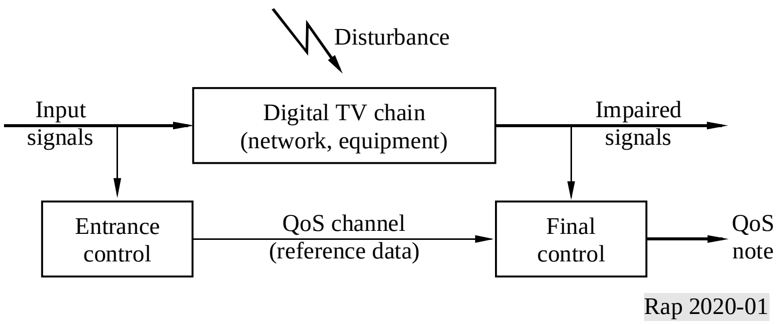
– in-service use, etc.

The methodology classification allows differentiating between each solution in order to identify its domain of application regarding its technical approach and constraints. This will also bring to the choice of specific procedure for the performance check with a consideration to the context of use.

To be implemented, the first measurement strategy needs a large amount of reference information provided at final comparison point. If that is possible in laboratory tests, this is not realistic on in-service operations and all over the network without strong compression, sub-sampling and synchronization with reference data. Those processes imply the use of heavy algorithms implemented on the final test equipment. This fact increases the complexity of the target hardware. The adequacy between cost and performances will be at a disadvantage of this first strategy.

FIGURE 1

**Methodology with reduced reference data**



The application of the second strategy using entrance and final control points is well adapted for an automatic and continuous monitoring of the quality of signals on DTV networks. It is useful to measure in service the deviation between input and output signals. It offers the advantage of the survey of the transparency of transmission links. On the other hand, the bit rate of the reference data is quite limited. The strategy with a reduced reference appears as an evidence for network monitoring. A reduced reference data stream, composed by relevant features and impairment characteristics represents the original signal. This information is easily dispatched by operating the QoS channel and provided to all final control points.

The third strategy will be considered, as an option of the second one to be used when no reduced reference data is available. This is the case when there is no possibility of access to the original signals or when the channel transmitting the reference data is missing or is broken.

The methodology proposed to be standardized for the monitoring of video quality over DTV network is “double-ended with reduced reference information” with an optional configuration without any need for reference.

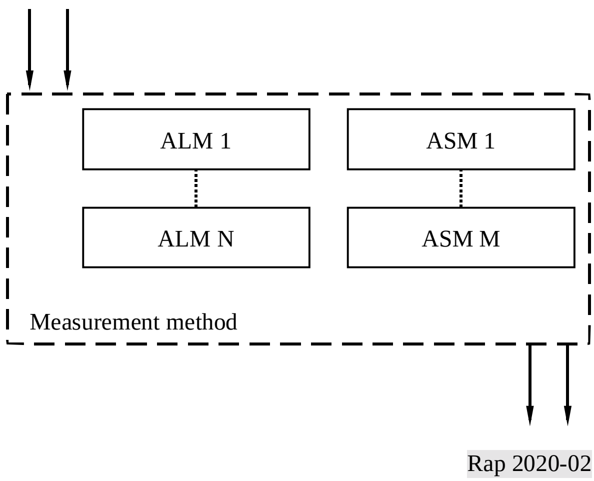
## 3.2 Measurement method

The goal of this proposition is to describe a conceptual representation of the measurement method to be used and recommended. It is as generic as possible, in order to allow the definition of general structure easy to be adopted and followed by the proponents and by standardization bodies. This generic description will be also as open as possible allowing improvement of algorithms and techniques.

To follow the measurement strategy recommended in § 3.1, a structure of the measurement method has to be described as the processing core implemented at the entrance and final control measurement points. A typical configuration for method is the structure shown hereafter and composed by algorithmic (ALM) and associated modules (ASM).

FIGURE 2

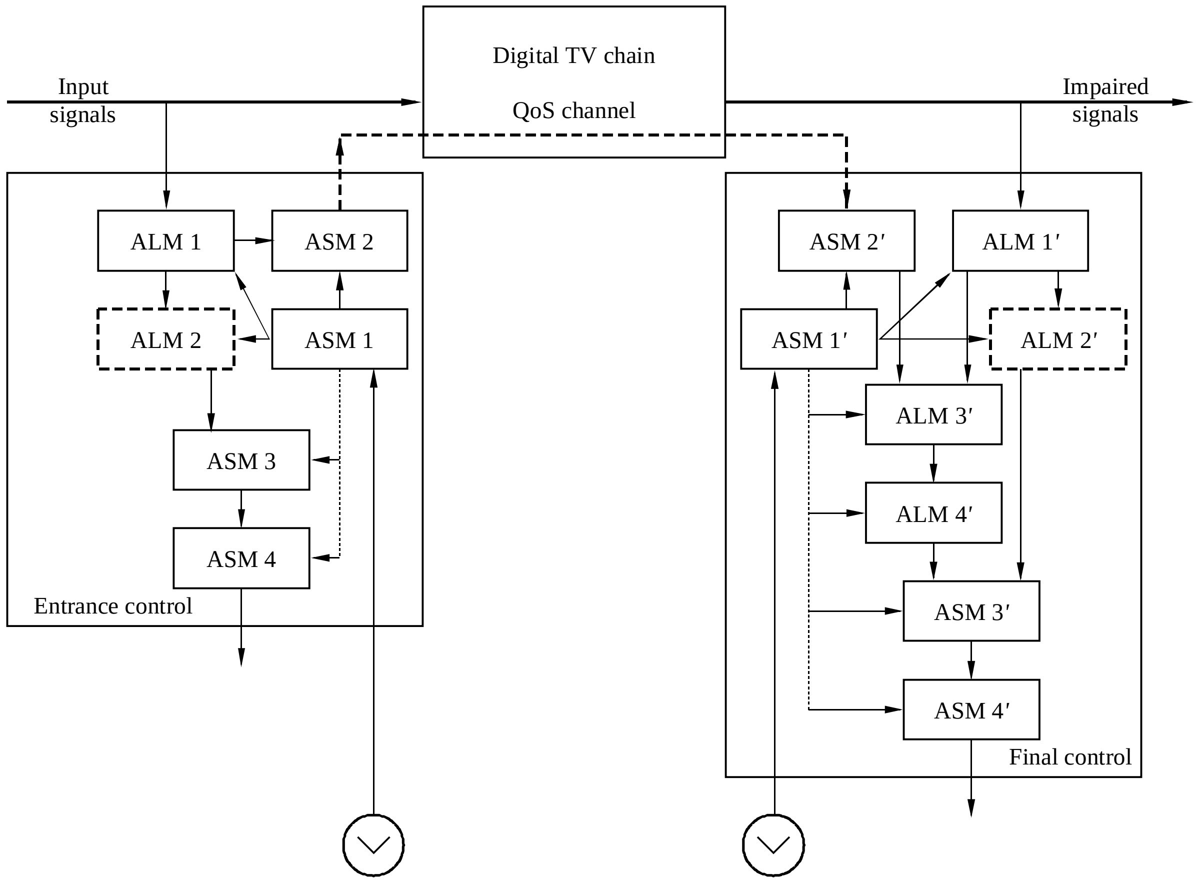
**Structure of the measurement method**




The proposed generic representation at entrance control point and final control one is the measurement structure shown hereafter. ALM and ASM are respectively the algorithmic and the associated modules composing the measurement method. The Indexes N and N′ indicate that the module is located at entrance control for N or at final control measurement point for N′.

FIGURE 3

**Detailed structure of the measurement method**




ALM 1 and 1′: signal representation
ALM 2 and 2′: quality assessment model without reference (optional)
ALM 3′: feature synchronization and comparison
ALM 4′: quality assessment model with reduced reference

ASM 1 and 1′: time stamping
ASM 2 and 2′: reference data handling
ASM 3 and 3′: result representation
ASM 4 and 4′: interfaces

Rap 2020-03

## 3.3 Algorithmic modules

The following paragraphs describe the main functions related to the algorithmic modules. They give, also, some examples proposed into the literature.

### 3.3.1 ALM 1 and 1′: Signal representation

This ALM concern modules where specific features are extracted from the signal. This is operated at the entrance control and at final control depending on the position of the measurement point or on the retained measurement methodology. The purpose of this algorithm is to use the extracted features to measure the impact of the impairments introduced by pieces of equipment or by transmission link on the video signal. At the end of the chain, a comparison between the processed features can provide an indicator of the video quality (see § 3.4.2). The performances of signal representation algorithmic module depend not only on the relevance of the representation, but also on the amount of extracted information to be transmitted. The limitation of the assigned bit rate of the reference data channel imposes the choice of the technique.

Signal representation is one or a set of transformations applied to the video in order to change its representation from luminance values of pixel matrix to another domain. This latter, can be a new matrix of transformed values or a vector of features, obtained on the initial domain or even after a mathematical transformation or filtering.

*Mathematical transformations of signal*

In the case of transformations, banks of band-pass filters are used. Each filter having selectivity to spatial frequency, direction, and temporal frequency [Watson, 1990]. Different types of transformations are used: discrete cosine transform, gabor functions, cortex transform, quadrature mirror filters and wavelets, etc. [Brétillon *et al.*, 1999; Daly, 1992; Teo and Heeger, 1994]. Based the decomposition of visual information, multi-resolution techniques are employed, allowing a good localization of information in time and space. Using such a transform scheme leads to represent the information in a way compatible with the human visual system.

*Feature extraction*

With the representation in vectors of features, very low bit-rate reference data can be reached. Indeed, this approach summarizes drastically the image content into the vector. The extracted parameters are expected to be sensitive to impairments introduced on signal by the systems (equipment, network, etc.). Most approaches tend to base the characterization of a video sequence on parameters relative to the spatial and temporal content or on parameters representing the impairment itself (block effect, blur, etc.). The other advantage of this approach is to allow a real-time implementation of the video analysis, in response to the randomness of transmission errors. This demands efficient, computationally tractable algorithms.

For spatial and temporal features, one approach consists in extracting edges from each frame of input and output video by one of the numerous existing methods (usually Sobel or Laplacian filters) [Webster *et al*., 1993; Ardito and Visca, 1995]. On the filtered images the mean, the standard deviation, the maximum, the minimum or the energy of the pixel values are calculated and then used as signal representation features: spatial and temporal information, etc [Webster *et al.*, 1993; Baïna and Goudezeune, 1999; Baroncini, 1999].

On the other hand, the other type of algorithms uses specific characteristics of video impairments obtained in relation with the standardised encoding techniques (e.g. MPEG-2). Those characteristics can be representative of one or several encoding functions implemented into codecs: the splitting grid of $8 \times 8$ pixel blocks, the discrete cosine transform, the quantization of its coefficients, the motion estimation applied to the macro-blocks, etc. [Baroncini, 1999; Lauterjung, 1998].

Finally, signal representation block is a module having as input the original signal and as an output the new representation.

After a mathematical transformation of video signal or feature extraction, signal representation is achieved as an output of the ALM 1. Then, the representations of reference and impaired pictures can be compared into ALM 2, to derive a distortion feature.

### 3.3.2 ALM 3′: Feature synchronization and comparison

In order to evaluate the impact of video systems on original signals and to measure the quality of the final video, a synchronization and the comparison between features are operated on the representation of input signals and output ones (see § 3.3.1). The result is considered as diagnostic measurement.

The proposed quality assessment method is a comparative one. In order to evaluate the quality of signals over one or several links of the network; it requires the comparison of two signal representations operated at different points. This approach imposes a precise synchronization between measurements done at entrance control point and those realized at final control ones.

Furthermore, it is indispensable to ascertain the veracity of the time stamping information concerning each measurement to be synchronized. This is to ensure the corresponding date between input representation and output ones. The synchronization and the comparison result is computed on a sample by sample, frame by frame or component by component basis depending of the signal representation (matrix or vector).

The comparison is defined as the difference or absolute difference between the signals or their representations. The amplitude of the comparison signal and its statistical properties carries information about the characteristics of the generated distortions. Many different parameters are commonly extracted from the error signal.

Several comparison functions have been used for this issue. Some ones are much more close to error signal calculations SNR, MSE, PSNR. They consist of a general comparison of signal values.

Other types of comparison functions have been introduced to be more adapted to the comparison of feature components: log_ratio, error and error_ratio. The mean and the maximum value applied on the previous functions allow having six different parameters obtained from features. A large set of objective parameters can be derived from spatial and temporal parameters and impairment parameters [Brétillon *et al.*, 1999; Webster *et al.*, 1993]. The results are a representation of the degradation between two video sequences. This output will feed the following module that is the perceptual model.

#### 3.3.3 ALM 2, 2′ and 4′: Quality assessment model

This section merges the previously defined features into a single quality prediction value, since several impairment types that can occur simultaneously influence the subjective judgement. The aim of the model is to compute a perceptual quality score. For example, it can deal with results of features comparison, and weight the result with *a priori* knowledge about human vision. So, the global quality score is computed by the way on signal parameters.

Since quality is influenced by multiple degradation types [Martens and Kayargalde, 1996], the most relevant measurements of individual impairments must be jointly used to evaluate perceptually objective quality by a single value. To this end, a combination model is set up. Several kinds of models are available to represent this process, based on a learning approach.

The most common model is a linear combination of the individual impairment measurements. An optimising process finds the weighting coefficients. This process seeks to minimize the distortion, like the mean square error, between the objective evaluation and a subjective evaluation. This process is operated on a large training set of sequences. Such an approach is implemented in several quality evaluation methods [Algazi *et al.*, 1994; Webster *et al.*, 1993]. The performance of the model for a given sequence is the correlation coefficient. A variant to this approach has been proposed for analogue systems (Recommendation UIT-R BT.654 – Subjective quality of television pictures in relation to the main impairments of the analogue composite television signal; ITU-R Handbook on Subjective assessment methodology in television) and applied to digital picture [Xu and Hauske].

The linear combination approach essentially suffers from the problem that the impairment combination may influence perceived quality in a non-linear way. Thus, ad hoc methods have to be proposed. Advanced learning techniques, like neural networks, have been employed. As an example, such a possibility has been investigated to predict quality on the normalized 5-grade scale [Kotani *et al.*, 1995; Brétillon *et al.*, 1999].

The learning methods have the great advantage of being simple and easy to be implemented, once the model has been set up. A general drawback is that the performance of the model is linked to the relevance of the training sequences used to set up the model.

The quality assessment model can be applied to the double-ended methodologies types 1 and 2 as ALM 4′, or to specific impairments for a single-ended methodology. In this latter case the model is considered as an option, as ALM 2 and 2′, implemented at measurement points providing quality measurement when no reference data is available.

### 3.4 Associated modules

The following paragraphs describe the main functions related to the associated modules. They give, also, some examples proposed into the literature.

#### 3.4.1 ASM 1 and 1′: Time stamping

The measurement points need a reference time in order to synchronize the processed quality parameters. The comparison obtained between the parameters has to be calculated for the same time stamp. The ASM 1 and 1′, are the pilots of the quality meters at the measurement points. They represent the schedulers of the whole system. The time stamping modules launch all measurement the processes. They, also, provide time stamps for the formatting of, on one hand, the reference data transmitted through the QoS channel with ASM 2 and 2′; on the other hand, the QoS information transmitted to the supervision system with ASM 4 and 4′. The time stamping modules have to use a common unified clock retrievable all over the network chain.

One example of a unified time reference is MPEG-2 internal system time clock (STC). It is useful to be used as a reference time for the generation of time stamps for the quality meter. This internal MPEG-2 reference clock is available at each measurement point localized everywhere onto the network [Baïna and Goudezeune, 1999]. The operations achieved at input and output equipment are then synchronous, to carry out the comparison of the parameters. The system labels all measurements with specific time stamps, which are present in the MPEG-2 stream.

### 3.4.2 ASM 2 and 2′: Reference data handling

In order to make available signal reference information all over digital television networks it must be transmitted to the final control stations.

One solution is to transmit input parameters in-band with the digital TV programmes, in a dedicated QoS channel multiplexed into the MPEG-2 transport stream. The bit-rate required to transmit the parameters is in the order of a few kbits/s, to be affordable. In this way, the parameters are easily broadcast to all final control measurement points [Baïna and Goudezeune, 1999].

For this purpose, the creation of a QoS channel was suggested and standardized in DVB (Recommendation for the usage of a user defined signalling channel embedded in a MPEG-2. Document TM1957(Rev.3) – Transport Stream under the Packet Identifier PID 0x001D. DVB). The multiplexing of the QoS channel into the MPEG-2 TS was proposed. DVB has edited the recommendations for the use of specific MPEG-2 packet identification number (PID) reserved to the QoS channel. Several applications are given to use this QoS channel.

One example of implementation proposed describe how to set up the test-data-section with information about QoS concerns the PID of elementary stream for which the measurement values are valid; the description of the content is in clear text, the processed measurement values and time stamps. This reference information is multiplexed at entrance control point and handled through the network until provided to the final control point (Recommendation for the usage of a user defined signalling channel embedded in a MPEG-2. Document TM1957(Rev.3) – Transport Stream under the Packet Identifier PID 0x001D. DVB).

The quality meter has to manage this QoS channel for reference information and to generate the quality parameters for network monitoring. For this purpose, inserter and extractor means have to be implemented as associated modules at entrance and final measurement points ASM 2 and 2′. An example of implementation has been proposed in [Lauterjung, 1999; Veillard and Negru, 1999; Brétillon and Baïna, 1999].

### 3.4.3 ASM 3 and 3′: Result representation

This module is needed to propose different types of measurement representation: graphical charts, curves. It also concerns the issue of short-term and long-term representation with the problems of time collapsing and statistical representation of objective quality measurement.

### 3.4.4 ASM 4 and 4′: Interfaces

Several solutions are possible for connecting the quality meter to the supervision system. This will allow gathering quality information for the monitoring of the network and providing it if it is requested.

To achieve the interfacing, HTML page server with JAVA applets is one possible solution. The other one is to use simple network management protocol (SNMP) agents or specific external PROXY’s. The agents allow to connect the pieces of equipment to the management information base (MIB) and consequently to the supervision system [Lipski, 1999].

## 4 Evaluation procedures

Model evaluation techniques are usually based on the comparison of objective measurements with subjective scores, and the choice of the subjective methodology has a major influence on the final results.

The proposition here is to study the way of adapting each methodology in both subjective and objective approaches in order to reach optimized performances. The goal here is to propose a procedure for the evaluation of model performances taking into account not only technical considerations but also implementation problem.

The goal of standardization is to provide techniques matching the needs of users (broadcasters, television industry and T&M manufacturers). The recommended methods have to be not only efficient from the theoretical point of view, but also suitable for an implementation of a realistic hardware platform in order to target the real world. Example: for the continuous monitoring of plenty of programmes there is a need for very low-cost pieces of equipment. This kind of constraint is a criterion much more important than an improvement of 0.05 for the correlation coefficient.

## 5 Conclusions

The goal of this contribution is to propose a methodology and method for the monitoring of digital television broadcasting network.

The proposition is a conceptual model for the description of perceptual objective assessment methods for signal quality. This will lead to finding common agreement and consensus around this generic approach well adapted to cover a large set of solutions. On the other hand, this proposition will bring to the comparison of algorithmic and associated modules in the right way. Indeed, a comparison of basic functions specified by their input and output and their internal operations is much easier than trying to compare the results obtained by the whole measurement methods from signal representation to quality notation providing. This simplification is imposed by the complexity of the techniques and by the difficulty to identify a unique winner trough a comparison tests procedure.

Further more the proposition is re-enforced by the current ITU-T and DVB standards. This in order to provide an outcome with a coherent solution with a consideration to the standardisation bodies.

The proposed method and generic structure can be extended to describe quality meters for audio and video signals or a combined audiovisual service.

REFERENCES

ALGAZI, V. R., OHIRA, H., KOTANI, K. and MIYAHARA, M. [March, 1994] Important distortion factors in the encoding of very high quality images. *SPIE,* Vol. 2298. Proc. Human Vision, Visual Processing and Digital Display V, San Diego (CA), USA.

ARDITO, M. and VISCA, M. [September, 1995] Correlation between objective and subjective measurements for video compressed systems. IBC Convention, Amsterdam, Netherlands.

BAÏNA, J. and GOUDEZEUNE, G. [September, 1999] Equipment and strategies for signal quality monitoring for digital television networks. *Proc. SMPTE J.*

BARONCINI, V. [December 15, 1999] Single-ended objective video quality assessment. International Workshop on Quality of Service for Digital Television, Munich, Germany.

BRÉTILLON, P. and BAÏNA, J. [December 15, 1999] Quality monitoring of broadcast audio and video signal. International Workshop on Quality of Service for Digital Television, Munich, Germany.

BRÉTILLON, P., BAÏNA, J., JOURLIN, M. and GOUDEZEUNE, G. [September, 1999] Method for image quality monitoring on digital television networks. *SPIE,* Vol. 3845, p. 298-305.

DALY, S. [1992] The visible difference predictor: An algorithm for the assessment of image fidelity. *SPIE,* Vol. 1666. Human Vision, Visual Processing and Digital Display III, p. 2-15.

KOTANI, K., KATAYAMA, M., MURAI, T. and MIYAHARA, M. [1995] Objective picture quality scale for color image coding. ICIP'95, Vol. 3, p. 133-136, Washington, DC, USA.

LAUTERJUNG, J. [October 6-7, 1998] First results of Digital Video Quality Measurements in DVB Networks. International Workshop on Quality of Service for Digital Television, Metz, France.

LAUTERJUNG, J. (Rohde & Schwarz) [December 15, 1999] The integration of a quality-of-service channel inserter and extractor in a test and measurement instrument. International Workshop on Quality of Service for Digital Television, Munich, Germany.

LIPSKI, J. L. [December 15, 1999] SNMP applied to DTV network equipment supervision. International Workshop on Quality of Service for Digital Television, Munich, Germany.

MARTENS, J. B. and KAYARGALDE, V. [1996] Image quality prediction in a multidimensional perceptual space. *Proc. IEEE*. Int. Conf. on Image Proc., ICIP'96, Vol. I, p. 877-880, Lausanne, Switzerland.

TEO, P. C. and HEEGER, D. J. [1994] Perceptual image distortion. *SPIE,* Vol. 2179. Human Vision, Visual Display and Digital Display, p. 127-141.

VEILLARD, M. (CCETT) and NEGRU, O. (THOMCAST) [December 15, 1999] Interoperability in a multi-application DVB MG measurement point. International Workshop on Quality of Service for Digital Television, Munich, Germany.

WATSON, A. [October, 1990] Perceptual component architecture for digital video. *J. of the Optical Society of America*, Vol. 7, **10**, p. 1943-1954.

WEBSTER, A., JONES, C. T., PINSON, M. H., VORAN, S. D. and WOLF, S. [June, 1993] An objective video quality assessment system based on human perception. *SPIE*, Vol. 1913, p. 15-26.

XU, W. and HAUSKE, G. Picture quality evaluation based on error segmentation. *SPIE*, Vol. 2308, p. 1454-1465.