

## Research Article

# A Model for Video Quality Assessment Considering Packet Loss for Broadcast Digital Television Coded in H.264

Jose Joskowicz<sup>1</sup> and Rafael Sotelo<sup>2</sup>

<sup>1</sup> Universidad de la República, 11300 Montevideo, Uruguay

<sup>2</sup> Universidad de Montevideo, 11600 Montevideo, Uruguay

Correspondence should be addressed to Jose Joskowicz; josej@fing.edu.uy

Received 31 October 2013; Revised 23 January 2014; Accepted 28 January 2014; Published 3 April 2014

Academic Editor: Fabrice Labeau

Copyright © 2014 J. Joskowicz and R. Sotelo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a model to predict video quality perceived by the broadcast digital television (DTV) viewer. We present how noise on DTV can introduce individual transport stream (TS) packet losses at the receiver. The type of these errors is different than the produced on IP networks. Different scenarios of TS packet loss are analyzed, including uniform and burst distributions. The results show that there is a high variability on the perceived quality for a given percentage of packet loss and type of error. This implies that there is practically no correlation between the type of error or the percentage of packets loss and the perceived degradation. A new metric is introduced, the *weighted percentage* of slice loss, which takes into account the affected slice type in each lost TS packet. We show that this metric is correlated with the video quality degradation. A novel parametric model for video quality estimation is proposed, designed, and verified based on the results of subjective tests in SD and HD. The results were compared to a standard model used in IP transmission scenarios. The proposed model improves Pearson Correlation and root mean square error between the subjective and the predicted MOS.

## 1. Introduction

Television is by far the communications service with higher penetration in society. It reaches every household in many countries of the world. Broadband Internet access does not have such an extended penetration in some countries. While in developing countries 72.4% of households have a TV set, only 22.5% have a computer and only 15.8% have Internet access (compared to 98%, 71%, and 65.6%, resp., in developed countries) [1].

In order to care for bringing quality to the people through communications services it is essential to be concerned about the quality delivered by the TV services. Nowadays great attention is given to Digital Television (DTV). Many countries in the world have already made the so-called “analog switch-off” which means that all analog TV transmitters in a given country or city are finally switched off, paving the way to new usage of the spectrum. Some other countries are planning it for some time between 2015 and 2020.

The completion of the analog switch-off will imply the end of the migration from analog TV to DTV.

The above mentioned applies to Terrestrial Digital Television, that is, the traditional free-to-air TV service. However, there are also some other transmitting mediums used for TV services, mainly for Pay TV, which were digitalized many years ago. Those are the cases of satellite TV and cable TV. Besides, digitalization has enabled the emergence of TV services over IP networks, allowing new network and service operators to offer Pay TV.

Although nowadays DTV is a relatively novel topic for the public and for policy makers in many countries of the world, it has been among us over other mediums, different from terrestrial, from some time ago. We call the set of DTV standards used up to the moment as the First Generation DTV Standards. They allow us to watch television and to access complementary multimedia content or applications but with a limited scope.

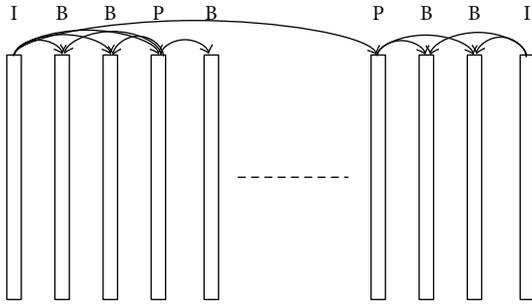


FIGURE 1: The MPEG-2 GoP and dependencies between I, P, and B frames.

In DTV, the video signal is coded in a certain codec, mainly MPEG-2 and H.264, and packetized in preparation for transmission. These Transport Stream packets (TS packets) are 188 bytes long. Its structure is defined in [2]. TS packets may be grouped and encapsulated in IP packets for streaming or IPTV services and are transmitted independently in terrestrial DTV.

DTV signal is subject to degradations in its path from the transmitter to the viewer. No matter which transmission medium is used (i.e., terrestrial, coaxial, fiber or satellite), noise is added to the original signal, leading to potential packet loss. Besides, when TV signals are transmitted over IP networks, packet loss may occur due to network issues, such as congestion. In order to control the quality of service the DTV operator offers to its audience, it is necessary to understand the way these packet losses affect the quality perceived by the TV viewer. Quality of experience (QoE) is a concept coined to represent how the viewer perceives not only the video or audio of a program but also the whole multimedia experience. Quality of service (QoS), a term commonly used in telecommunications, refers to many network-related parameters of a certain service that must be considered in order to fulfill a set of requirements that are related to quality. Beyond QoS, QoE evaluates the impact that different parameters affecting the joint transmission of audio, video, and associated data or applications have on the final audiovisual experience that the user has.

In this paper we present an approach to evaluate the impact of packet loss on video quality perceived by the user for broadcast television transmission environments. We will briefly describe video coding and transmission techniques in the current DTV standards in Section 2. Then in Section 3 we review the topic of packet loss impact on QoE. Firstly, we consider IP transmission and then Terrestrial DTV transmission. We show essential differences that must be taken into account between IP and Terrestrial DTV, when measuring impact of packet loss on video quality perceived. In Section 4 we present an overview of different techniques to achieve video quality evaluation. We performed some subjective tests that are presented on Section 5, along with the discussion of the results, where we report that the percentage of packet loss and its distribution are not sufficient to describe perceived quality. Section 6 introduces a new metric, representing the *weighted percentage* of slice loss, which is correlated with

the perceived degradation introduced by TS packet loss. Using this metric and a previously published model to predict video quality in the absence of packet loss, a new model is developed that estimates video quality for Terrestrial DTV coded in H.264. The verification of the model with a second round of subjective tests with clips obtained from actual DTV recordings is also presented in this section, along with a comparison to a standard model for video quality estimation. Finally, Section 7 has the conclusions and future planned work.

## 2. Video Coding and Transmission in Current DTV Standards

There are three main DTV families of standards that are being or have already been deployed all over the world:

- (i) Advance Television System Committee (ATSC) [3];
- (ii) Digital Video Broadcasting (DVB) [4];
- (iii) Integrated Services Digital Broadcasting (ISDB) [5].

There are particular standards for different mediums for DTV transmissions, for example, DVB-S for Satellite TV, ATSC, DVB-T and ISDB-T for Terrestrial TV, and DVB-C for Cable Television. In these standards the video signal is typically coded in MPEG-2 or H.264 [6] and packetized in small packets prior to being transmitted. In MPEG-2 video frames are grouped into sequences, called “Groups of Pictures” (GoP). Each GoP can include three different types of frames (see Figure 1): I (“Intra”), P (“Predictive”), and B (“Bidirectional predictive”). Type I frames are encoded only with spatial compression techniques. They are used as reference frames for the prediction (forward or backward) of other P or B frames. P slices are coded using prior information from frames I and other P frames, based on motion estimation and compensation techniques. The B frames are predicted based on information from previous (past) and subsequent (future) frames. The size of a GoP is given by the number of frames existing between two I frames. H.264 provides better compression techniques than MPEG-2 and each frame can be divided into one or more slices. In this case the GoP structure is related to slices rather than frames, and the encoding process is different. Another difference is that while in MPEG-2 GoP structure is fixed, in H.264 it can vary during time. In both MPEG-2 and H.264, the coded video can then be packetized in small Transport Stream (TS) packets, 188 bytes long. As shown in Figure 2, they consist of a header of 4 bytes and 184 bytes of payload. The header contains different fields, including a Packet Identifier (PID), the Program Clock Reference (PCR), a Transport Error Indicator (TEI) flag, and a 4-bit Continuity Counter, among others.

All these DTV standards have differences between them, but they share some characteristics on the channel coding. They have an initial stage for data randomizing followed by another one with a Reed-Solomon Encoder, a data interleaver, and an inner code. These blocks are defined in order to reduce the impact of errors on transmission. A DTV receiver, once it has demodulated the transmitted signal, undoes the processes

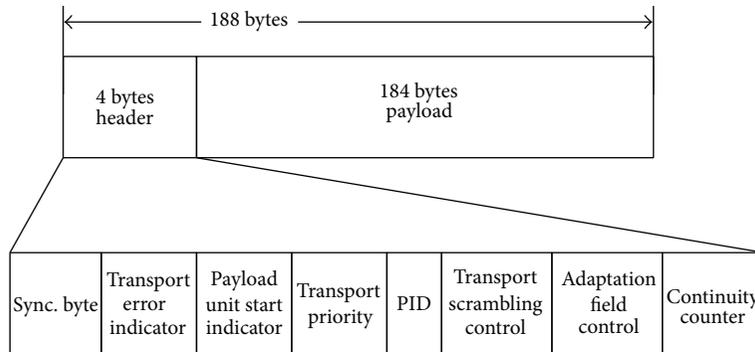


FIGURE 2: Structure of the MPEG-2 Transport Stream packet.

of channel coding achieved at the transmitter. The last stage in a DTV receiver is the Reed-Solomon (RS) decoder. In the case of ISDB-T, DVB-T, DVB-S, or DVB-C, the Reed-Solomon Code is (204, 188) and can correct errors in up to 8 bytes. On the other hand, ATSC uses a Reed-Solomon Code of (207, 187) capable of correcting errors in up to 10 bytes. Thus, some kind of bit errors in the TS packets at the initial stage of the decoder can eventually be corrected, and a valid TS packet can be presented to the video decoder, even if the packet had some errors at the front end of the decoder.

### 3. Packet Loss Impact on QoE

There has been considerable work published regarding the QoE impact of packet loss in video transmission over IP networks. Most of the published papers consider the effect on video quality with respect to the percentage of IP packet losses. There are different possible patterns for packet losses studied in the literature, both with a random distribution and with taking into account the effect of bursts [7, 8]. It is known that even small percentages of IP packets loss with uniform distribution can produce high effects in the perceived video quality. Reference [9] describes a test performed transmitting H.264 streams over IP networks with 0.02% random packet loss rate. The authors report that higher levels of packet loss severely damage the user experience by freezing the receiver for long periods of time. Authors of [10] have verified that noise structure affects perceived quality for a given packet loss rate. They experimented with 0.375% and 2% of packet loss rate with different burst densities (5%, 20%, 45%, 75%, and 100%). Reference [11] while studying how to assess quality of experience for high definition video streaming under diverse packet loss patterns it concludes that the perceived quality of HD video streaming is prone not only to packet losses but also to patterns of losses.

In IP networks, multimedia transport can be performed by encapsulation of TS packets in User Datagram Packets (UDP) or in Real Time Protocol (RTP) [12] over UDP. In any case up to seven TS packets of 188 bytes can be carried in a single 1500 bytes IP packet, in order to improve efficiency. Thus, in an IP transmission environment, a lost IP packet produces a burst of seven lost TS packets.

On the other hand, the common approach in DTV is to use the Bit Error Rate (BER) as a parameter related to the capability of the receiver to reconstruct the transmitted signal. For example, DVB-T [13] defines Quasierror Free (QEF) reception as less than one uncorrected error event per hour, corresponding to  $BER = 10^{-11}$  at the input of the MPEG-2 demultiplexer or a  $BER = 2 \times 10^{-4}$  after Viterbi. The common approach is that if Carrier to Noise relation ( $C/N$ ) is below a certain value, there will be a cliff effect (also called brick effect or “brick wall” effect or “fall off the cliff”) that will cause an immediate degradation of the signal [14–16].

Another concept called the Correct Reception Rate (CRR) has been used in [17, 18], but it also is used as a threshold (good/bad reception) and does not enable making any analysis of the perceived video quality as a function of this parameter. Other works have partially analyzed the effect of the signal fading in the video quality [19]. However this parameter is very difficult to measure at the receiver and cannot be included in a video quality estimation model.

Although the quality of the radio frequency transmission link is usually characterized by the BER, this approach is not sufficient to study the problem on how transmission link noise affects QoE from the user’s perspective in DTV. A decrease in received  $C/N$  will make the DTV receiver go from a clear and no error picture to a “picture freeze” (the so-called cliff effect). However, this transition is not extremely abrupt, and as  $C/N$  decreases from perfect reception, received picture experiences different kinds of degradations before reaching a complete “blockiness.” According to some literature [20, 21] this transition from no degradation to full degradation takes place in a received  $C/N$  fall of 1 to 3 dB. Reference [22] evaluates theoretically and practically the performance of DTV signals over Gaussian, Rician, and Rayleigh channels. It also evaluates video quality with a metric called DVQL-W. Their results for the Rician channel are that the evaluated picture quality indicates visible errors in the picture with the  $C/N$  equal to 22 dB or lower, while the “cliff-off” effect comes with the  $C/N$  equal to 20 dB or less in the transmission channel. That means that for this case there is an interval in which, while  $C/N$  varies between 22 and 20 dB, the quality perceived varies from acceptable to bad. On the other hand, for a Rayleigh channel, the evaluated picture quality based on DVQL-W metric indicates visible errors in the picture

with the  $C/N$  equal to 28 dB or lower. The “cliff-off” effect comes with the  $C/N$  equal to 23 dB or less in the transmission channel. That is, in this case, the transition from a good quality to a completely bad quality is a 5 dB fall in the  $C/N$  of the received signal. This range may include many households in a normal coverage area of a DTV station.  $C/N$  varies with environment conditions and varies with time. That increases the number of receivers that can be on the edge of this “cliff effect” in different periods of time. Besides, noise intensity may have different structures during time, depending on its origin (i.e., homogeneous or bursts). When the signal intensity decreases, some errors may be present in the TS packets. The error correction techniques provided by the standards (RS Codes) can eventually correct the errors, but the RS decoding algorithm may be overloaded and be unable to correct the packet. In this case the transport\_error\_indicator bit in the TSP header shall be set [23]. The decoder can decide what to do with the missing information. As for our experience, we have analyzed three different consumer type receivers that when recording a Transport Stream file and when an error occurs, they simply drop the TSPs marked as having had a transport error. That is, by checking the continuity counter of the header of the TSP of the TS file recorded, some missing TS may be found.

Taking into consideration the previous paragraphs, it is important to point out that, when studying losses related to DTV which have origin in channel noise, it is necessary to focus on degradations experienced by packet loss due to digital broadcasting transmission. Particularly, the problem of the video quality assessment in the transition from completely good reception (no packet loss after error correction) to completely bad reception (totally degraded picture) must be studied. Channel noise leads to errors in bits at the receiver front end; but since the Reed-Solomon Code can correct errors in packets, either bit errors are corrected or there are errors in TS packets at the output of the Reed-Solomon decoder, previous to the video decoder. This will affect each TS packet independently, leading to individual TS packet loss patterns, as opposed to what happens in IP networks, where the loss of an IP packet produces seven consecutive lost TS packets.

The distribution pattern of the packets loss at the output of the RS decoder may also be different with respect to the IP networks loss patterns. In IP networks, packets loss can be produced by network congestion or jitter-buffer overflow, with patterns distributions such as the Gilbert-Elliot or similar variants. This kind of packet loss distributions has been analyzed for multimedia services over IP [24, 25]. Nevertheless, in Terrestrial DTV scenarios, the packet loss is produced by a completely different cause than in IP networks, that is, low signal to noise ratio at the front end of the receiver, which is followed by the Viterbi decoder and followed by the Reed-Solomon decoder. Besides, in DTV systems, many other error prevention mechanisms are used, such as byte or bit interleaving before modulation and time interleaving or frequency interleaving while preparing the OFDM-frame structure (for systems using OFDM). Thus, it is not evident that the same typical loss patterns model can be applied in this case. In order to explore the type of degradations we need

to set up our base of video clips for the subjective tests; we have recorded many different intervals from free-to-air DTV from two different broadcasters from Montevideo, Uruguay, with low reception at the front of the receiver. The video clips used for subjective tests are usually 10 seconds long, so we the recordings we made from the free-to-air DTV have that exact duration. Analyzing the decoded TS packets, we have found the following characteristics.

- (i) There are periods with homogeneous losses, presumably corresponding to signal fading. These periods have a length higher than the 10 seconds used.
- (ii) There are periods without any lost packet and periods with many losses (bursts). In most cases, the periods with losses (bursts) are less than one second, followed by periods of more than one second without losses.
- (iii) Nearly, the 50% of the losses consist of individual TS packets (i.e., the continuity counter of the header of the TSP reports only one missing TS packet).

Each lost TS packet can carry video coded information related to a specific frame or slice type, that is, I, P, or B. Although it seems intuitive that the impact of the loss would be very different depending on the type of frame it corresponds to, there are few papers that take this approach. In [26] it is shown that “not all packets are equal.” Loss or damaged I frames affect much more than P or B frames, because the information of I frames is used to decode P and B frames of the entire GoP, so the effect is propagated to more frames. Loss of P frames affect more than B frames, because its information is used to decode other P and B frames, from the lost P frame until the next I frame. Finally, a lost or damaged B frame does not affect any other frames, just itself. Reference [27] studies QoE in DVB-H networks using frame loss pattern and video encoding characteristics. They define and use a parameter called the “loss rank” that weighs the impact of a lost frame on the perceived quality by counting how many other frames the error propagates.

#### 4. Video Quality Evaluation

Audiovisual content producers and TV operators try to offer the best possible video quality to their viewers. Better video quality has been one of the driving forces for the advent of DTV. However, some processes involved in DTV, such as digital video encoding and transmission systems, introduce degradations that may result in unsatisfactory perceived quality. Video quality depends on many aspects related to the encoding process, the transmission stage, the receiver, or even to the content itself.

TV operators can define an upper limit to the video quality by setting the bit rate assigned to each signal, the GoP structure and size, and other parameters, while controlling the encoding process. Besides, the selection of the transmission parameters can affect the way the signal is propagated and received, particularly, in relation to the robustness to noise, thus affecting the degradation introduced during the transmission stage. These configurable parameters such as the modulation used, the FEC, or the guard interval

have incidence on the signal to noise required for proper reception. The perceived quality can also be affected by the receiver. Different display sizes, display technologies (CRT, LCD, or LED), and error concealment strategies applied at the receivers can affect the perceived quality as experienced by the end user. Finally, given the encoding and transmission parameters and the receiver settings, different video contents can be perceived with different quality, depending on the amount of spatial and temporal activity of the sequence.

The most accurate methods for measuring the perceived video quality of a video clip are the subjective tests, where the video sequences are presented to different viewers and opinions are averaged. The Mean Opinion Score (MOS) or the Difference Mean Opinion Scores (DMOS) are the metrics typically used in these tests. Different kinds of subjective tests can be performed, based on recommendations ITU-R BT.500-13 [28], ITU-R BT.710-4 [29], and ITU-T P.910 [30]. These recommendations describe the methodology, environment, scales, and number of observers and conditions, among other aspects of the tests. In all cases, the tests are conducted in laboratories, using controlled environments, and with video sequences specially selected and prepared for this purpose. MOS varies from 1 (“bad” quality) to 5 (“excellent” quality).

## 5. Subjective Tests Performed and Analysis of Results

We performed subjective tests to verify the effect of packet loss on DTV signals. Five different video clips were used: “Fox & Bird”, “Football”, “Concert”, “Voile”, and “Golf”, obtained from [31]. These video clips span over a wide range of different spatial and temporal activity. Each clip’s length is ten seconds. The video clips were coded in H.264/AVC, High Profile, Level 4.1 for HD and Main Profile, Level 3.1 for SD, with no more than two consecutive B frames and key interval of 33 frames. One hundred different degraded video clips were generated in HD and one hundred in SD, varying the bit rate and the percentage of packet loss with different distribution patterns, including uniform distribution and different number of bursts, as detailed in Table 1.

Burst intervals were less than one second long, according to the observed bursts duration in real broadcasting signals, as described in the previous section. During the burst interval, we have decided to use a uniform distribution loss pattern for this study, because we have observed many individual TS lost packets inside the burst interval in real signal recording, as described in the previous section. The different percentages of packet loss inside each burst (0.1% and 10%) were selected in order to simulate scenarios with very different number of losses. The case of 0.1% simulates that most of the error can be corrected by the RS decoder. On the other hand 10% of loss simulates that most of the errors cannot be corrected by the RS decoder.

The subjective tests were performed according to the general guidelines of Recommendation ITU-R BT.500-13, using the five points Absolute Category Rating with Hidden Reference (ACR-HR) scale, as defined in Recommendation

TABLE 1: Packet loss patterns tested.

	Percentage of packet loss
No packet loss	0% along all the video clips
Uniform	0.3% along all the video clips
One burst	0.1% inside the burst; 0% outside the burst
One burst	10% inside the burst; 0% outside the burst
Two bursts	0.1% inside the burst; 0% outside the burst
Two bursts	10% inside the burst; 0% outside the burst
Three bursts	0.1% inside the burst; 0% outside the burst
Three bursts	10% inside the burst; 0% outside the burst

ITU-T P.910. A 42” Led TV was used, without any kind of postprocessing techniques. The test room allows up to four simultaneous people watching the video clips and voting. A special voting system was developed, allowing the evaluators to use a smart phone application, synchronized with the clip sequences, to enter each score. The voted scores are automatically stored in a database, associated with the test session, the specific video clip, and the user. The system is depicted in Figure 3. This application can be downloaded from [32].

Nineteen nonexpert viewers from 21 to 55 years old performed the evaluation for the SD clips. Three of them were identified as “outliers,” as the Pearson Correlation of their ratings compared to the average was below the threshold established in 0.75 in the test plan. Twenty-five viewers from 21 to 42 years old performed the evaluation for the HD clips. Four of them did not pass the normal vision tests. From the twenty-one remainders, one of them was identified as “outlier.” The data collected from the remainder viewers in the SD and HD tests was used to compute the MOS values. The tests are considered “formal” according to the definition of ITU-R BT.500.

Video quality is often evaluated according to [33]

$$V_q = 1 + I_c I_p, \quad (1)$$

where  $V_q$  is the predicted MOS (or MOSp),  $I_c$  is the video quality determined by the encoding process, and  $I_p$  is the video quality degradation introduced by the packet losses in the transmission process.  $I_c$  can vary from 0 to 4 and  $I_p$  from 0 to 1. When there is no packet loss,  $I_p$  equals 1, leading to a  $V_q$  dependence on  $I_c$  exclusively (determined by the encoding process). On the other hand, when packet loss extremely affects video quality,  $I_p$  equals 0, resulting in the worst possible value for  $V_q$ , 1. Thus,  $I_p$  reduces  $I_c$  by a factor related to the degradations introduced by packet loss in the transmission stage. Resulting values for MOS are appropriately from 1 to 5.

Using the data obtained from the subjective tests, we compute  $I_c$  and  $I_p$  for each coded clip (“Fox”, “Voile”...) in each degraded condition. Subjective tests produced a MOS value for each of the two hundred coded clips (including HD and SD). Since when there is no packet loss,  $I_p$  equals 1,  $I_c$  was obtained for each combination of bit rate and clip, using the degraded video clips without packet loss (i.e., the clips degraded according to the first row of Table 1). Then, for each

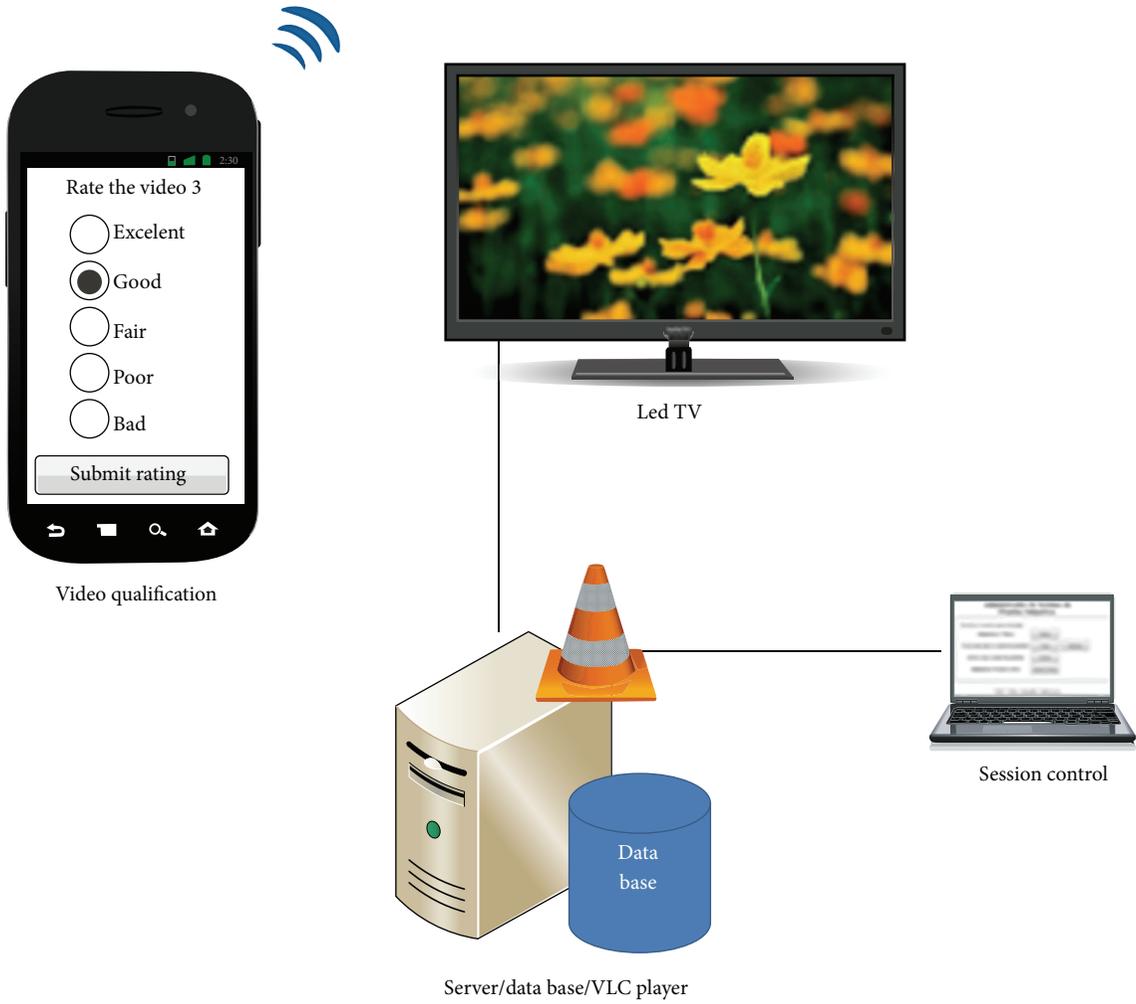


FIGURE 3: Automatic rating system for subjective tests.

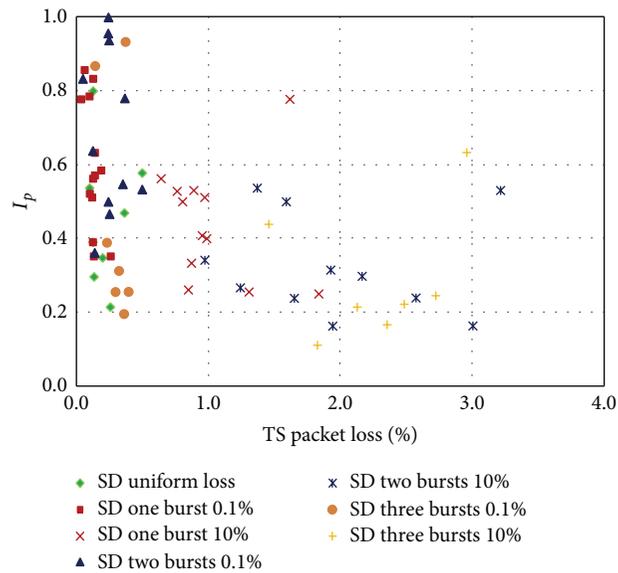


FIGURE 4:  $I_p$  versus percentage of TS packet loss for SD resolution.

other clip, where there are losses,  $I_p$  was derived comparing the MOS of the same clip at the same bit rate (MOS resulting from subjective tests), but with the corresponding packet loss degradation.

Packet losses were generated on individual TS packets. This represents a broadcast transmission scenario, not an IP transmission (where TS packets are lost in blocks of seven).

Figures 4 and 5 present the results of  $I_p$  with respect to the percentage of TS packet loss, for SD and HD, respectively, and in different conditions (random and burst distribution). Each point in the graphs represents a particular degraded video from the two hundred evaluated in the subjective tests. Video clips without packet loss are omitted in both graphs, for better readability. Many conclusions can be drawn from these charts.

**5.1. General Considerations.** Looking at Figures 4 and 5, it can be seen that, given a percentage of packet loss,  $I_p$  shows a very high dispersion, even for the same video clip (e.g., for 1.5% of TS packet loss,  $I_p$  varies approximately from 0.2 to 0.8). It is remarkable that, in some cases, even with very low percentages of packet loss, the video is highly degraded. This is expected as part of the mentioned “cliff effect.” Nevertheless, in many other cases, even with high percentages of packet loss, the video quality is only partially degraded, and acceptable values for the MOS are obtained.

**5.2. Uniform Loss Pattern.** With uniform TS packet loss (green squares in both graphs), the video quality is degraded ( $I_p$  is low) even with very low percentage of TS packet loss. The case of HD is especially illustrative: at 0.3% of TS packets loss,  $I_p$  varies between 0.03 and 0.56, with an average value of 0.14, as can be seen in Figure 6. This means that, even for “perfect” encoded videos (MOS = 5), at 0.3% of TS packet loss, the MOS drops in average to 1.6.

A ten seconds HD video clip, coded at 14 Mb/s, has about 15 000 TS packets, distributed in 600 video slices. A 0.3% of packet loss means 45 individual TS lost packets, randomly distributed along the entire video clip. These 45 lost packets can impact in up to 45 different slices, leading up to  $45/600 = 7.5\%$  of lost slices. In our tests, using 5 different degraded HD video clips coded at 14 Mb/s with 0.3% of TS packet loss, a minimum of 22 slices, and a maximum of 35 slices was affected, corresponding to 4.5% and 5.8% of lost slices. This explains the huge effect that random packet loss can have in the perceived video quality.

**5.3. Burst Loss Patterns.** With burst losses, as shown in Figures 4 and 5,  $I_p$  decays much slower than with a uniform loss pattern, even with much higher percentages of lost TS packets. As stated before, a ten seconds HD video clip, coded at 14 Mb/s, can have the order of 15.000 TS packets distributed in 600 video slices. One burst of no more than one second can contain the order of 1500 TS packet. A 10% of packet loss means 150 individual TS lost packets, randomly distributed inside the burst. The total number of affected TS packets in this case is higher than the obtained with 0.3% of uniform distribution loss. Nevertheless, this kind of

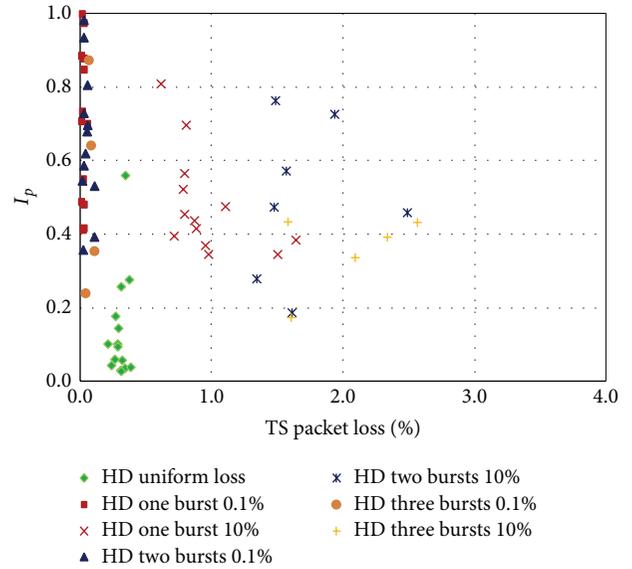


FIGURE 5:  $I_p$  versus percentage of TS packet loss for HD resolution.

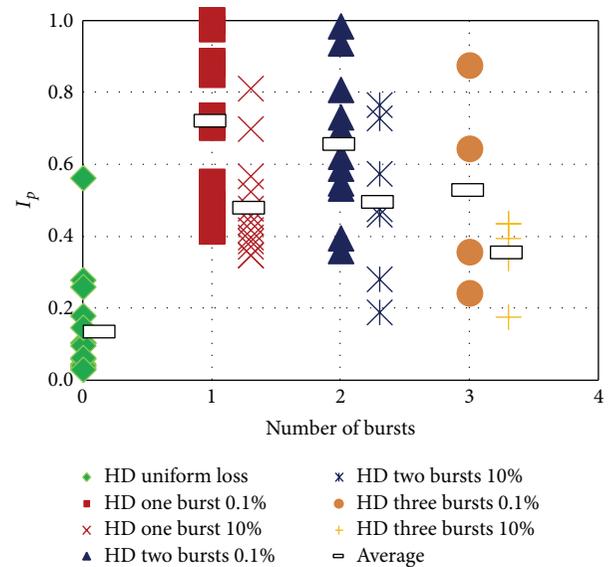


FIGURE 6:  $I_p$  versus number of bursts in HD video clips.

degradation affects much less the video quality, as can be seen in the average  $I_p$  values in Figure 6. This can be explained by the fact that many lost TS packets can be related to the same slice, because the lost packets are very close between them in time. As an example, in a particular video clip, 106 TS packets were lost in one burst, but only two slices were affected.

## 6. Modeling and Verification

Different models for video quality estimation were presented by different authors in recent years. Some of them include the video quality estimation in the presence of lost packets, but in an IP data network scenario. In a previous work [33] we have

made a comparison of ten of such models. Within the models that take into account packet loss, the one with the best performance was the proposed in Recommendation ITU-T G.1070 [34]. This model was developed for small display sizes, but a similar model for HDTV was proposed in [35]. In the ITU-T G.1070 model,  $I_p$  is expressed as

$$I_p = e^{-p/D_{pplv}}, \quad (2)$$

$$D_{pplv} = v_{10} + v_{11}e^{-f/v_8} + v_{12}e^{-b/v_9},$$

where  $b$  is the bit rate,  $f$  is the frame rate,  $p$  is the percentage of IP packet loss, and  $v_8$  to  $v_{12}$  are coefficients that must be calculated for each codec and display size (coefficients' names are presented as in the recommendation). In our work, the frame rate was fixed at 25 fps in all cases, so the number of coefficients can be reduced to three, defining  $v = v_{10} + v_{11}e^{-(25/v_8)}$ , and  $I_p$  can be expressed as

$$I_p = e^{-p/(v+v_{12}e^{-b/v_9})}. \quad (3)$$

We have tested this model in the Terrestrial DTV scenario, using  $p$  as the percentage of lost TS packets (instead of the percentage of lost IP packets). We have calculated the values of  $v$ ,  $v_9$ , and  $v_{12}$  that minimize the Root Mean Square Error (RMSE) between the actual  $I_p$  values (derived from the subjective tests) and the values derived from (3). With these values, the obtained Pearson Correlation was 0.60 and the RMSE 0.30 (in the 0-1 scale), which are not good results. These poor results were in some way expected. Figures 4 to 6 show that there are great variations in the  $I_p$  value for the same percentage of packet loss and even for the same number of bursts. Based on these observations, we can conclude that video quality cannot be properly estimated by any model that takes into account just the percentage of packet loss and number of bursts, as most of the published papers consider.

Taking into account the previous considerations, we can conclude that the impact of a particular loss has a great variance depending on the impacted slice (I, P, or B) and probably on the video content. As stated in Section 2, since B and P slices depend on information from I ones, loss of I slices affect much more than P or B ones, and, analogously, loss of P slices affect more than B slices. With these considerations, a new metric representing the *weighted percentage* of slice loss  $p_w$  can be defined, as detailed in

$$p_w = x_1I + x_2P + B, \quad (4)$$

where  $I$  is the percentage of affected I slices (i.e., the number of affected I slices with respect to the total number of slices in the video clip),  $P$  is the percentage of affected P slices,  $B$  is the percentage of affected B slices, and  $x_1$ ,  $x_2$  are two coefficients. The coefficient  $x_1$  can be interpreted as the average number of affected slices when there is an error in a I slice. Analogously,  $x_2$  can be interpreted as the average number of affected slices when there is an error in a P slice. There is no coefficient in  $B$  because the errors in affected B slices are not propagated.

We have found that  $I_p$  can be correlated with  $p_w$ , selecting the appropriate values for  $x_1$ ,  $x_2$ . This is depicted in Figure 7,

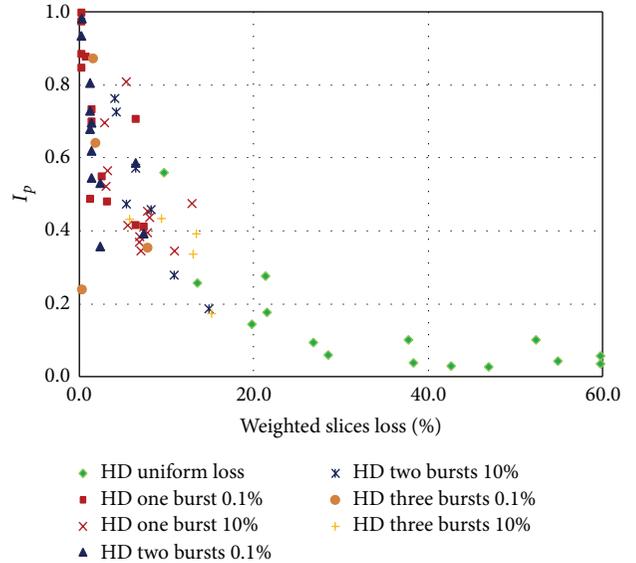


FIGURE 7:  $I_p$  versus  $p_w$  (weighted percentage of I, P, and B lost slices).

TABLE 2: Best values for  $k$ ,  $x_1$ ,  $x_2$ .

$x_1$	21.5
$x_2$	5.7
$k$	26.9

where  $I_p$  is plotted against  $p_w$  (for the case of HD video clips). The relation between  $I_p$  and  $p_w$  can be expressed as

$$I_p = \frac{1}{1 + kp_w}, \quad (5)$$

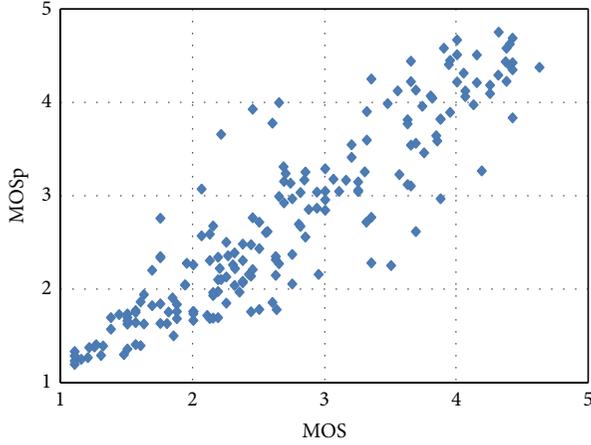
where  $k$  is a constant. We have calculated the values of  $k$ ,  $x_1$ ,  $x_2$  that minimize the RMSE between the actual  $I_p$  values (derived from the subjective tests) and the values derived from (5). The resulting values are presented in Table 2. With these values, the obtained Pearson Correlation (PC) is 0.84 and the Root Mean Square Error (RMSE) between the actual and the derived  $I_p$  values is 0.16. These values are much better than the obtained using the G.1070 model and that were mentioned above.

This corroborates the fact that the influence of a lost packet depends on which type of slice was impacted and gives some insights into the relative weights of the impact in the degradation perceived for each slice type.

In order to obtain an estimation of the MOS (MOS<sub>p</sub> or  $V_q$ ) for each clip using (1),  $I_c$  must also be evaluated. We have calculated  $I_c$  according to the formula and coefficient values previously presented in [36], showed in

$$I_c = 4 \left( 1 - \frac{1}{1 + (ab/v_4)^{v_5}} \right), \quad (6)$$

where  $I_c$  is the video quality due to the encoding process,  $a$  depends on display size,  $b$  is the bit rate, and  $v_4$  and  $v_5$

FIGURE 8: MOSp ( $V_q$ ) versus MOS for the first set of measurements.

are coefficients that depend on video content, according to (coefficient's name are presented as in [36])

$$\begin{aligned} v_4 &= c_1 s^{\zeta_2} + c_3 \\ v_5 &= c_4 s^{\zeta_5} + c_6, \end{aligned} \quad (7)$$

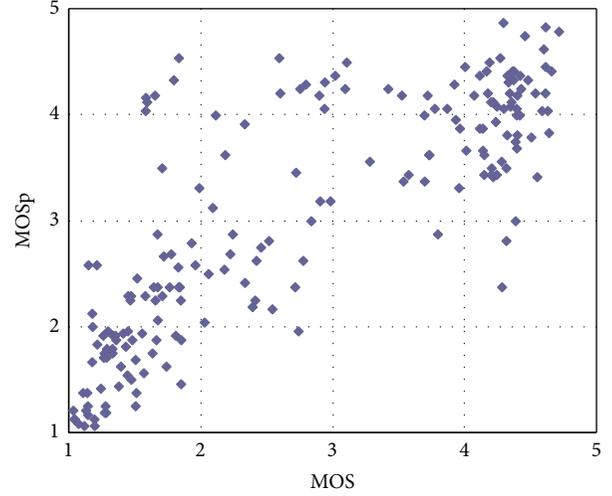
where the parameter  $s$  depends on video content and is the average Sum of Absolute Differences (SAD) of the video and  $c_1, \dots, c_6$  are fixed coefficients.

Figure 8 shows the MOSp ( $V_q$ ), obtained using (1), (4), (5), and (6) for each clip against the actual MOS obtained with the subjective tests. The overall Pearson Correlation (PC) is 0.91 and the RMSE is 0.42. Using the same values estimation for  $I_c$  according to (6), but the standard G.1070 estimation according to (3) for  $I_p$ , then the overall PC falls to 0.75 and the RMSE raises to 0.80. As can be seen, using the model for  $I_c$  presented in [36] and the model for  $I_p$  detailed in this paper, a good prediction of the MOS (and better than G.1070) can be obtained.

In order to validate the model, a second set of tests was achieved with actual video clips recorded from free-to-air DTV from two different broadcasters from Montevideo, Uruguay. The recorded clips selected had the following conditions.

- (i) Some of the signals were recorded in good reception conditions, without packet loss; some of the signals were recorded with low reception conditions, thus obtaining “real world” packet loss.
- (ii) It can be inferred that the original signal is of “excellent quality.”
- (iii) They are representative of various types of content, including movies, sports, music, advertising, cartoons, news, and documentaries.
- (iv) As far as possible, the selected clips cover different ranges of spatial and temporal activity.
- (v) Scene changes rates are typical from actual TV.

One hundred clips in HD format and one hundred clips in SD format, with 10 seconds of duration, were recorded. These

FIGURE 9: MOSp ( $V_q$ ) versus MOS for model validation.

recorded clips had differences with respect to the specially prepared ones for the first training phase of the model. The first difference is that one broadcaster used six slices per frame, and the other used one slice per frame (in the first sets of clips, all of them had one slice per frame). The second one is that one broadcaster had dynamic GoP structure, compared to the static GoP structure used in the first set of clips.

The subjective tests were performed in the same conditions as the first ones, according to ITU-R BT.500. The five points Absolute Category Rating method was used. Eighteen nonexpert viewers from 18 to 55 years old performed the evaluation for the SD clips. Two of them were identified as outliers, leaving sixteen evaluations for the MOS calculation. Twenty-seven viewers from 17 to 68 years old performed the evaluation for the HD clips, but three of them did not pass the vision test. There were not outliers in the remaining twenty-four evaluations. This second test is also “formal” according to ITU-R BT.500.

The dispersion between the subjective scores and the obtained with (1), (4), (5), and (6) was calculated, maintaining the same coefficient values used for the first set of clips, as shown in Figure 9. The PC was 0.81 and the RMSE was 0.80. Although these values are worse than the obtained using the “training” data, they are good enough to support the fact that the proposed model can be applied for real DTV signals, where many packet loss patterns may be present.

## 7. Conclusions and Future Work

This paper describes a model to assess video quality in Terrestrial Digital Television when packet loss is present. We justify that the correct approach in the case of DTV is losses of individual TS packets in contrast to the case of IP transmission, where seven TS packets are lost with each IP packet.

Results of subjective tests performed with two hundred video clips in HD and SD resolution are presented that show a different incidence of the packet loss when it is present

randomly than when in bursts. The results also show that there is a high variability on how quality is described by end users for a given type of error (i.e., a great variation in the final MOS values can be seen for the same number of bursts and with the same loss distribution pattern). This fact implies that there is practically no correlation with the type of error or the percentage of TS packets loss and the perceived degradation. This implies that none of the models based only on these parameters (percentage of lost packets and number of bursts) can produce appropriate results. A deeper inspection must be performed in the lost TS packet in order to understand the impact on video quality.

We have introduced a new metric, the *weighted percentage* of slice loss, which takes into account the affected slice type (L, P, or B) for each TS packet. This metric is correlated with the video quality degradation introduced by the packet losses in the transmission process of Terrestrial DTV. Using this metric and a previously published model to predict video quality in the absence of transmission degradations, a novel parametric model for video quality estimation in Terrestrial DTV was proposed. The results were compared to a standard model used in IP transmission scenarios, obtaining much better Pearson Correlation and RMSE between the subjective MOS and the predicted MOS using the proposed model. The model was also verified using actual video clips recorded from free-to-air DTV from two different broadcasters, obtaining satisfactory results. We conclude that the proposed model can be applied in real DTV environments.

Future planned work will improve the model, including the ability to explicitly take into account dynamic GoP structures and different number of slices per frame.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

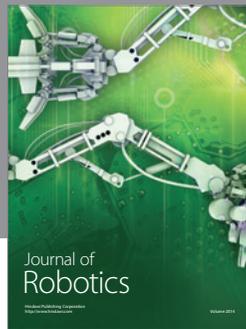
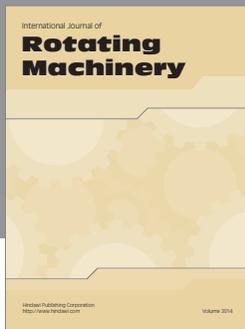
This work was supported by ANII-MIEM/Dinatel, FST\_1.-2012\_1.8147.

## References

- [1] ITU, "The World in 2010: ICT Facts and Figures," 2010, <http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf>.
- [2] ISO/IEC, "ISO/IEC, 13818-1: MPEG-2 Part1. Information technology— generic coding of moving pictures and associated audio information: systems".
- [3] ATSC, <http://www.atsc.org/>.
- [4] DVB, <http://www.dvb.org/>.
- [5] ARIB, <http://www.arib.or.jp/english/>.
- [6] G. J. Sullivan and T. Wiegand, "Video compression—from concepts to the H.264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18–31, 2005.
- [7] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: effect of burst losses and correlation between error frames," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 861–874, 2008.
- [8] F. You, W. Zhang, and J. Xiao, "Packet loss pattern and parametric video quality model for IPTV," in *Proceedings of the 8th IEEE/ACIS International Conference on Computer and Information Science (ICIS '09)*, pp. 824–828, June 2009.
- [9] M. H. Pinson, S. Wolf, and G. Cermak, "HDTV subjective quality of H.264 vs. MPEG-2, with and without packet loss," *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 86–91, 2010.
- [10] N. Chen, X. Jiang, and C. Wang, "Impact of packet loss distribution on the perceived IPTV video quality," in *Proceedings of the 5th International Congress on Image and Signal Processing (CISP '12)*, pp. 38–42, 2012.
- [11] M. Leszczuk, L. Janowski, P. Romaniak, and Z. Papir, "Assessing quality of experience for high definition video streaming under diverse packet loss patterns," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 903–916, 2013.
- [12] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RFC, 3550 RTP: A Transport Protocol for Real-Time Applications," 2003.
- [13] ETSI, "300 744 V1. 6. 1 (2009-01): Digital video broadcasting (DVB), framing structure, channel coding and modulation for digital terrestrial television," 2009.
- [14] F. C. C. De Castro, M. C. F. De Castro, M. A. C. Fernandes, and D. S. Arantes, "8-VSB channel coding analysis for DTV broadcast," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, pp. 539–547, 2000.
- [15] W. Fischer, *Digital Video and Audio Broadcasting Technology*, Chapter 12, Springer, Berlin, Germany, 2010.
- [16] T. Kratochvil and V. Ricny, "Simulation and experimental testing of the DVB-T broadcasting in the SFN networks," in *Proceedings of the 18th International Conference Radioelektronika*, April 2008.
- [17] G. Bedicks, F. Sukys, E. L. Horta et al., "Field measurements for ISDB-TB in the VHF band," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '11)*, June 2011.
- [18] C. Akamine, F. Yamada, G. Bedicks Jr. et al., "Field trial for Brazilian DTV using space diversity," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '10)*, March 2010.
- [19] D. Zegarra Rodriguez and G. Bressan, "Video quality assessments on digital TV and video streaming services using objective metrics," *IEEE Latin America Transactions*, vol. 10, no. 1, pp. 1184–1189, 2012.
- [20] S. R. Martin, "RF performance of DTV converter boxes—an overview of FCC measurements," *IEEE Transactions on Broadcasting*, vol. 56, no. 4, pp. 441–451, 2010.
- [21] T. S. Bonfim and M. M. Carvalho, "Video Quality Evaluation for a Digital Television Broadcasting Scenario," [http://enpub.fulton.asu.edu/resp/vpqm/vpqm10/proceedings\\_vpqm2010/vpqm\\_p41.pdf](http://enpub.fulton.asu.edu/resp/vpqm/vpqm10/proceedings_vpqm2010/vpqm_p41.pdf).
- [22] T. Kratochvil and R. Štukavec, "DVB-T digital terrestrial television transmission over fading channels," *Radioengineering*, vol. 17, no. 3, pp. 96–102, 2008.
- [23] DVB, "ETSI TR 101 290, Measurement guidelines for DVB systems," Digital Video Broadcasting.
- [24] G. Hasslinger and O. Hohlfeld, "The Gilbert-Elliott model for packet loss in real time services on the Internet," in *Proceedings of the 14th GI/ITG Conference on Measuring, Modelling and*

*Evaluation of Computer and Communication Systems (MMB '08)*, 2008.

- [25] N. B. Yoma, C. Busso, and I. Soto, "Packet-loss modelling in IP networks with state-duration constraints," *IEE Proceedings on Communications*, vol. 152, no. 1, pp. 1–5, 2005.
- [26] J. Greengrass, J. Evans, and A. C. Begen, "Not all packets are equal, Part 2: the impact of network packet loss on video quality," *IEEE Internet Computing*, vol. 13, no. 2, pp. 74–82, 2009.
- [27] K. D. Singh and G. Rubino, "Quality of experience estimation using frame loss pattern and video encoding characteristics in DVB-H networks," in *Proceedings of the 18th International Packet Video Workshop (PV '10)*, pp. 150–157, December 2010.
- [28] ITU, "Recommendation ITU-R BT. 500-13, Methodology for the subjective assessment of the quality of television pictures," January 2012.
- [29] ITU, "Recommendation ITU-R BT. 710-4 Subjective Assessment Methods for Image Quality in High Definition Television," November 1998.
- [30] ITU, "Recommendation ITU-T P. 910 Subjective video quality assessment methods," April 2008.
- [31] The Consumer Digital Video Library, <http://www.cdvl.org>.
- [32] <http://www2.um.edu.uy/ingenieria/vqi/subjectivetests.htm>.
- [33] J. Joskowicz, R. Sotelo, and J. C. L. Ardao, "Comparison of parametric models for video quality estimation: towards a general model," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '12)*, Seoul, Republic of Korea, 2012.
- [34] ITU-T, "Recommendation ITU-T G. 1070: opinion model for video-telephony applications," July 2012.
- [35] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 110–114, May 2008.
- [36] J. Joskowicz and J. C. L. Ardao, "Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model," in *Proceedings of the 6th IFIP/ACM Latin America Networking Conference (LANC '11)*, pp. 4–11, October 2011.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

