

Subjective Evaluation of Transmission Errors in IPTV and 3DTV

Jesús Gutiérrez*, Pablo Pérez[†], Fernando Jaureguizar*, Julián Cabrera*, and Narciso García*

*Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

[†]Alcatel-Lucent, María Tubau 9, 28050 Madrid, Spain.

Abstract—The increase of multimedia services delivered over packet-based networks has entailed greater quality expectations of the end-users. This has led to an intensive research on techniques for evaluating the quality of experience perceived by the viewers of audiovisual content, considering the different degradations that it could suffer along the broadcasting system. In this paper, a comprehensive study of the impact of transmission errors affecting video and audio in IPTV is presented. With this aim, subjective assessment tests were carried out proposing a novel methodology trying to keep as close as possible home environment viewing conditions. Also 3DTV content in side-by-side format has been used in the experiments to compare the impact of the degradations. The results provide a better understanding of the effects of transmission errors, and show that the QoE related to the first approach of 3DTV is acceptable, but the visual discomfort that it causes should be reduced.

I. INTRODUCTION

Lately, the number of multimedia services delivered over packet-based networks is increasing more and more, together with the quality expectations of the end-users. With the aim of satisfying the viewers, broadcasters, service providers and other sectors promote the research on evaluating the Quality of Experience (QoE) perceived by the users of multimedia contents. However, many factors are involved in the viewing experience of the users, such as audiovisual quality and aspects related to human perception. This makes the evaluation of QoE a very complex task which has led researchers to carry out numerous works [1].

Since several subjective factors of the observers are related to the perceived QoE, the most accurate methods for its measurement are those in which the content is evaluated by a number of subjects. These subjective experiments also provide a deeper knowledge of the human visual system (HVS) and consequently, make easier the development and improvement of techniques to process and deliver audiovisual content to households. Nevertheless, subjective assessment tests are not suitable for practical applications, like monitoring the QoE in a packet-based network, which is one of the most important purposes in broadcasting systems. Thus, objective quality metrics are developed, which estimate automatically the quality perceived by the end-users of multimedia services. The results obtained in subjective tests are essential to evaluate the performance of objective QoE metrics and to develop robust and efficient metrics, which is still an open issue.

In addition, the appearance of new audiovisual technologies, like 3DTV, keeps very active the research on QoE evaluation. Several studies showed that QoE related to 3D video content is influenced by more factors than those affecting conventional video, like depth perception and visual discomfort [2]. However, 2D video quality is essential to achieve high QoE from 3D video content. Therefore, a deep knowledge of the influence of degradations on monoscopic video is needed to study the quality of stereoscopic video.

In this paper, a comprehensive analysis of the effects of transmission errors in multimedia content delivered over packet-based networks is presented. Firstly, a comparison of the degradations of QoE caused in IPTV was carried out. Then, stereoscopic sequences

were used to achieve a first knowledge of the impact of these errors on the first generation of 3DTV [3]. Subjective experiments were made to evaluate the impact of typical distortions affecting video and audio, caused by packet losses or network outage. A novel single stimulus methodology of evaluation was used, which is based on standard methods and was designed to mimic real home environment viewing conditions.

This paper is organized as follows. In Section II a brief review of the state-of-the-art related to the effects of transmission errors on QoE is presented, together with a description of the type of errors considered in the current work. The experimental setup of the subjective tests is detailed in Section III. The obtained results are presented in Section IV and the final conclusions in Section V.

II. TRANSMISSION ERRORS IN AUDIOVISUAL CONTENT

Several distortions could degrade the QoE perceived by the final viewers of audiovisual content when it is broadcasted to households. Especially important are the effects of typical transmission errors in packet-based networks, like packet losses, jitter and outage. Therefore, some research works have been presented analyzing the influence of these distortions in the QoE. For example, a subjective study of the visibility of packet losses is presented by Lin *et al.* in [4] and in their previous works, considering MPEG-2 and H.264/AVC encoders, various coding settings, and different patterns of video losses. To analyze not only the visibility, but also perceptual effects on QoE, Boulos *et al.* [5] carried out subjective experiments taking into account various patterns including single and multiple losses. In [6], Staelens *et al.* compared the effects of packet losses with video freeze distortions.

In addition, most of the existing works only analyze the influence of transmission errors affecting video, while very few studies have been presented evaluating their impact on the audio of video sequences. In [7], Hands carried out various experiments to analyze the relationship between audio and video quality.

Subjective tests are extremely important, as they provide knowledge about the HVS, information about the performance of the processing techniques applied to broadcasted content, and a basis to develop objective quality metrics to be applied in network monitoring [4][8].

While many works exist in the literature analyzing the influence of transmission errors on the quality of monoscopic video, the research on the effects of these kind of distortions in 3D video content is taking its first steps. For example, a subjective experiment was carried out by Barkowsky *et al.* in [9], analyzing the impact of packet losses on 3DTV content in a simulcast broadcasting scenario.

In this paper, typical effects of the transmission of multimedia content over error-prone channels were taken into account. Different patterns of video and audio losses were considered, in addition to distortions caused by network outage, like video freezing. Moreover, consequences of reductions of the quality of service of the network

TABLE I
CONSIDERED TRANSMISSION ERRORS.

Error type	Video losses								Bitrate drop		Framerate drop		Video freeze			Audio losses						
Code	E1	E2	E3	E4	E5	E6	E7	E8	R1	R2	F1	F2	V1	V2	V3	A1	A2	A3	A4	A5	A6	AV
Duration	1	1	1	1	3	5	3	5	6s	6s	6s	6s	0.5s	2s	6s	Min	0.5s	2s	6s	Min	Min	6s
Pattern	Min	90%	90%	90%	90%	90%	25%	25%	1/2	1/4	1/2	1/4	-	-	-	S	S	S	S	3losses/2s	7losses/6s	Audio-Video loss
% Frame	100	25	50	100	50	50	50	50														

were taken into account, such as framerate drops or bitrate drops [8]. The impact of these degradations was studied for conventional video and side-by-side (SbS) 3D video, which is based on the spatial multiplex of left and right views into one single frame. Thus, the existing infrastructure for delivering IPTV could be used for 3DTV broadcasting.

The specific distortions considered in the tests are shown in Table I. For video losses, the duration indicates the number of consecutive GOPs, and the pattern represents the percentage of the frames of each GOP affected by packet losses. "Min" is used to indicate that only B frames not used as reference are affected. Also the percentage of the frame that has been corrupted is shown. When video packets are lost, the error concealment algorithm of the decoder substitutes the affected macroblocks for the corresponding ones of the previous frame, thus the effects of video losses propagates until the next I-frame. Concerning bitrate and framerate drops, the pattern represents the reduction with respect to the reference bitrate and framerate of the sequences. For audio losses, "Min" indicates that a single audio packet is lost. Finally, regarding the pattern, "S" indicates a single loss, while in the other cases it is shown the number of losses per time interval.

III. EXPERIMENTAL SETUP

A. Environment and Equipment

The subjective tests were carried out at the Universidad Politécnica de Madrid, in a lab especially fitted-out to satisfy the recommendations given in ITU-R BT.500-11 [10] for assessments in home environments. The display used for presenting the sequences was a 42" Panasonic stereoscopic television, with resolution of 1920x1080. For visualizing 3D sequences, the observers wore the active shutter glasses associated to this TV set. The subjects were positioned at a viewing distance of 3 times the height of the TV, following the indications of the manufacturer in agreement with the ITU recommendations.

B. Observers

In the subjective experiment participated 22 observers, all of them having normal visual and stereoscopic acuity, and color vision. The ages of the subjects were ranged between 24 and 48 years old, with an average age of 31. The results obtained in the subjective tests were screened according to the recommendation ITU-R BT.500-11[10], to discard outliers and observers who could have voted randomly. After this process, no observers were rejected.

C. Procedure

To evaluate the impact of transmission errors on the QoE perceived by the end-user of a broadcasting system, the assessment method should keep as close as possible the typical viewing conditions in real home environments. Furthermore, the effects of these distortions on video quality are highly dependent of time aspects, like their duration or the moment when they happen, so it is necessary to evaluate the quality of the sequence in a continuous way. Due to these reasons, the methodology used in the tests was based on the single stimulus methods.

The observers had to evaluate the distortions inserted in monoscopic and stereoscopic sequences filling a questionnaire while viewing them. As outlined in Figure 1, the sequences were divided into segments of around 6 seconds. Different types of transmission errors among those described in Section II were randomly introduced in alternate segments, resulting in 25 distorted segments to evaluate. In some cases, no errors were introduced, being these segments hidden references. A number (i) was printed in the right-bottom corner of the image in each undistorted segment. This number has the appropriate size to make easy its perception in the screen by the observers, since it indicates which square of the questionnaire they should fill. In that square they had to evaluate the effect of the distortion of the previous segment (T_i) using the five-grade impairment scale (5 imperceptible, 4 perceptible but not annoying, 3 slightly annoying, 2 annoying and 1 very annoying) defined in [10].

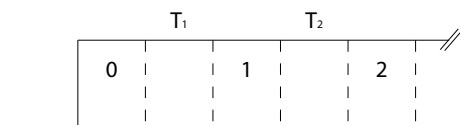


Fig. 1. Diagram of the final structure of the test sequence.

The segment labeled with a 0 triggers the start of the viewing time and provides a reference of coding quality of the video. Finally, all the segments were stitched together again, obtaining continuous sequences. As the segments with printed numbers have no errors, the observers can avert their gaze to the questionnaire if needed. The proposed assessment methodology is similar to the Absolute Category Rating described by the Video Quality Experts Group in [11], but instead of inserting grey segments to make the evaluation, segments with no degradation are introduced. This allows to keep the continuity of the sequence and get closer to real viewing conditions.

The evaluation tests were carried out organizing different assessment sessions, in which a maximum of 4 subjects participated. When monoscopic and stereoscopic versions of a video source were available, both of them were evaluated consecutively. Each session started with an explanation of the tests focused on clarifying the objective of the experiments and familiarizing the subjects with the evaluation methodology and the possible transmission errors. In total, the tests sessions lasted less than 30 minutes.

In addition, after seeing each stereoscopic video, the observers were asked to indicate their preference between this one and the same variant showed in monoscopic format. They also evaluated in a five-grade scale the degree of visual fatigue after viewing 3D video, and they could freely describe other types of discomfort felt during the tests, like headache or dizziness.

D. Test Material

As the aim of the subjective experiment is to evaluate the quality perceived by the users of a broadcasting system in their homes, three test sequences were selected taking into account content commonly watched by viewers. Following this same idea, the sequences should be longer than those used in evaluation of coding quality (about 10s), thus the duration was set to 5 minutes.

TABLE II
TEST SEQUENCES

Source	Format	Content
SRC1	1920x1080p 23.97fps	Movie. Some slow segments with dialogs. Some others with fast camera movement.
SRC2	720x576p 25fps	Documentary. Slow action. Some segments with camera panning. Only music as soundtrack.
SRC3	720x576p 25fps	Football match. Commentaries as soundtrack.

The main characteristics of the test sequences are shown in Table II. The sequences were firstly encoded in H.264/AVC, multiplexed in MPEG-2 TS, and packetized in RTP. A GOP length of 24 frames was used following the structure IBBBP. All I-frames were set as Instant Decoder Refresh (IDR) pictures, and only one slice was used in each frame. The bitrate used for encoding the videos were 8Mbps for SRC1, and 4Mbps for SDTV sequences. After that, they were processed to insert transmission errors randomly in the corresponding segments, and evaluate them as explained in Subsection III-C. Only 3D versions of the first and second video sources were available, thus SRC3 was not used in stereoscopic tests.

IV. RESULTS

A. Errors affecting video and audio

For the distortions affecting video, the MOS values obtained in the evaluation of the three monoscopic sequences are shown in Figure 2, together with a 95% confidence interval. Since the objective of the study is a comparison of the different type of distortions, the absolute MOS values are less important than the relation among the different transmission errors. In addition, the absolute MOS values could be affected by other factors which are not under study, like the coding quality of the sequences. This can be noticed with the yellow line represented in the figures, which indicates the MOS value obtained for the evaluated non-distorted segments.

From a global point of view, video freezing is the distortion which degrades more the QoE, and the longer the duration of the freezing, the worse the QoE perceived by the observers. The main cause of this, is that the natural rate of the video is altered, which is crucial for the quality perceived by the viewers [6]. In relation to this fact, it is shown that framerate drops, which also changes the flow of the sequence, are worse than other distortions of similar characteristics like bitrate drops. As it was expected, in both cases of bitrate and framerate drops, a greater reduction of the original rates produces higher degradations of the QoE. A framerate drop is the typical effect when the quality of service of the network does not satisfy

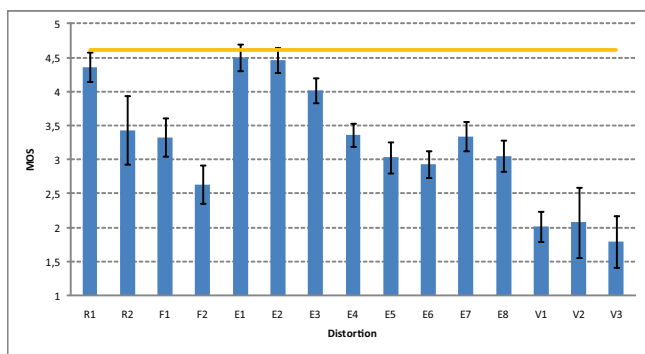


Fig. 2. Results for video distortions.

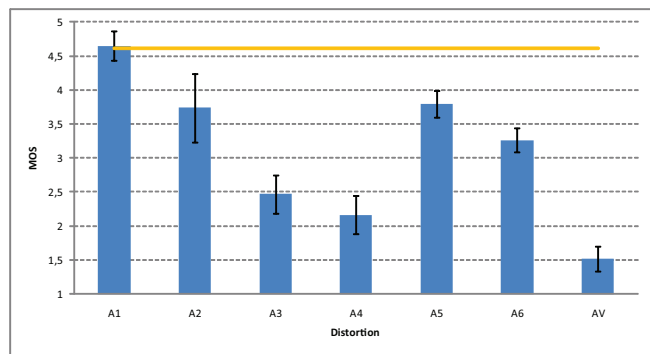


Fig. 3. Results for audio distortions.

the requirements. However, if versions of the broadcasted sequence with lower bitrates are available, delivering content with lower coding quality could be a better countermeasure for that situations.

Various conclusions could be extracted from the analysis of the effects of video packet losses (E1-E8). Firstly, it is shown that losses affecting B frames which are not used as reference for other frames (E1) are hardly perceptible, having a MOS value close to the result obtained for non-distorted segments. Regarding the fraction of the frame affected by the losses, it can be seen that the greater the distorted area, the worse the perceptual quality obtained. This is extracted from the comparison of distortions E2, E3, E4, and V1. When distortions E4 and V1 take place, the whole frame is lost, but in the case of V1 also the header could not be decoded, causing short video freezing. In addition, losing a great part of the GOP (E5-E7, E6-E8) produce worse results, and packet losses affecting more consecutive GOPs (E5-E6, E7-E8) also degrade more the QoE.

The results related to audio distortions are presented in Figure 3. As expected, the greater the duration of the losses, the worse the QoE perceived by the subjects (A1, A2, A3, A4). Also, a greater number of losses cause more severe degradations (A1, A5, A6).

The effects of outage could be the loss of audio, long video freezing or both. From the results obtained in the experiments, long video freezing (V2, V3) are more annoying than long audio losses (A3, A4). As it was expected, a total outage causing video freezing and audio loss (AV) is the effect of transmission errors which reduces more the QoE. The relations between the influence of the distortions on the quality of audio and quality of video should be explored in more detail in future experiments, because it is essential in the research on multimedia quality metrics.

B. Comparison between monoscopic and stereoscopic videos

After analyzing the effects of different transmission errors on the QoE associated to monoscopic videos, a comparison of the impact of these distortions in 3D content is presented. As the target is to compare the degradations in different video formats, the errors described in Table I have been grouped. Thus, 6 type of errors have been considered: bitrate drops (R), framerate drops (F), video losses (E), video freezing (V), audio losses (A), and video freezing combined with audio loss (AV). The MOS values obtained for the monoscopic and stereoscopic sequences of SRC1 and SRC2, with a 95% confidence interval are shown in Figure 4.

As it can be seen in Figure 4, very similar results were obtained for the same distortion in monoscopic and stereoscopic content. In fact, the confidence intervals of the MOS values overlap, questioning the statistical significance of the differences between the results. Therefore, comparisons between the results for 2D and 3D sequences were done applying statistical tests of significance. Specifically, after

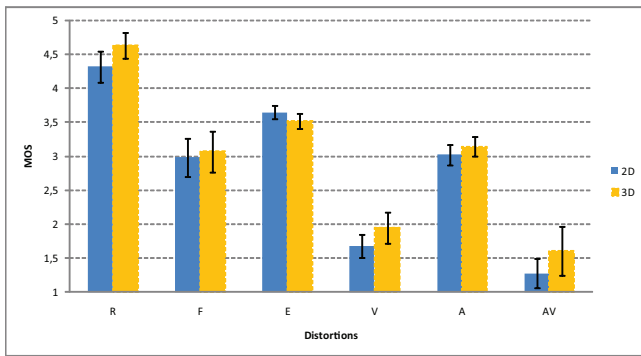


Fig. 4. Results for distortions in stereoscopic sequences.

checking that any of the samples of results were normally distributed by a Kolmogorov-Smirnov test (5% significance level), the Wilcoxon Signed-Rank test was used [12]. This non-parametric test is used to compare two non-normal samples of data looking for statistical significance in the difference between their mean values.

The Wilcoxon tests showed statistical significance for the slightly better results obtained with 3D content for bitrate drops (5% level), video freezing (1% level), and audio losses (2% level). This could be attributed to the added value provided by the factors related to the depth perception of 3DTV. On the other hand, the differences between the results for framerate drops and video freezing combined with audio loss are not significant. Probably caused by the drastic annoyance produced by the alteration of the natural video pace in the first case, and by the total loss of audiovisual information in the second one. The most remarkable result is related to the significant worse results obtained with 3D videos for video losses (1% level). The reason of this could be related to the frame-compatible format considered to broadcast 3D content. Using SbS 3D video implies encoding a frame containing both stereo views with H.264/AVC, and decoding it at the receiver side with a conventional decoder. This means that when video packets are lost in the network, a common algorithm of error concealment is used, which usually cause blocking artifacts in different regions of each stereo view, resulting in different content in corresponding regions of each view. These differences in the content visualized by each eye could difficult the fusion of both images by the HVS producing visual discomfort to the observer.

After seeing each sequence in both formats, the observers were asked to express their preference between the monoscopic and the stereoscopic versions of a video source. Regarding SRC1, a 52.63% of the observers preferred the 3D sequence, while only a 21.05% preferred the stereoscopic version of SRC2 (three of the viewers did not expressed their preferences). This aspect is heavily influenced by the quality of the production of the depth perception, and by the video content itself. As these results show, the viewers would only switch from 2D to 3D, only when their expectations related to 3D perception are notably satisfied.

Concerning the results related to visual discomfort, more than a half of the subjects felt visual discomfort in some degree during the visualization of 3D content. Moreover, a 15% of the observers felt headache or slight dizziness after seeing any of the 3D source sequences. Finally, some observers reported other annoying factors, like the reduction of luminance while seeing 3D video wearing shutter glasses, or a very high scene activity.

V. CONCLUSION

In this paper an exhaustive study of the impact of transmission errors on the QoE perceived by end-users of a broadcasting system

is presented. Specifically, various typical distortions related to IPTV were analyzed, affecting video and audio. Subjective assessment tests were carried out using a novel methodology, keeping as close as possible home environment viewing conditions. The results showed that those distortions affecting the natural video rate cause worse degradations of the QoE, like video freeze or severe framerate drops. The relations between the different errors will help to develop objective metrics for quality monitoring of IPTV networks.

The analysis of the effects of these distortions was extended to side-by-side broadcasted 3DTV. It is worth noting that the same distortions were less annoying for 3DTV than for the same content in monoscopic format, except in the case of video losses. The reason is that these losses produce blocking effects affecting differently the left and right views, thus causing binocular rivalry and considerable visual discomfort. These results show that the first generation of 3DTV based on frame-compatible formats provides acceptable QoE, but the research on more optimal techniques for processing and delivering 3D content are needed to improve the final QoE.

The study presented in this work provides a better understanding of the effects of transmission errors on QoE associated to audiovisual content, which will help to obtain robust methods to subjectively evaluate this type of distortions, and to develop and evaluate objective metrics for quality monitoring in broadcasting packet-based networks.

ACKNOWLEDGMENT

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV), and by the Ministerio de Industria under project ITEA2-JEDI.

REFERENCES

- [1] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008.
- [2] Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, "Video quality assessment: From 2D to 3D - Challenges and future trends," *Proc. IEEE Int. Conf. on Image Processing*, pp. 4025–4028, Sept. 2010.
- [3] ITU, "Features of three-dimensional television video systems for broadcasting," *Report ITU-R BT.2160*, 2009.
- [4] T. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. C. Cosman, and A. R. Reibman, "A versatile model for packet loss visibility and its application in packet prioritization," *IEEE Trans. on Image Processing*, vol. 19, no. 3, pp. 722–735, Mar. 2010.
- [5] F. Boulos, B. Parrein, P. Le Callet, and D. S. Hands, "Perceptual effects of packet loss on H.264/AVC encoded videos," *Proc. Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2009.
- [6] N. Staelens, S. Moens, W. Van Den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van De Walle, and P. Demeester, "Assessing Quality of Experience of IPTV and Video on Demand services in real-life environments," *IEEE Trans. on Broadcasting*, vol. 56, no. 4, pp. 458–466, Dec. 2010.
- [7] D. Hands, "A Basic Multimedia Quality Model," *IEEE Transactions on Multimedia*, vol. 6, no. 6, pp. 806–816, Dec. 2004.
- [8] P. Pérez, J. Macías, J. Ruiz, and N. García, "Effect of packet loss in video Quality of Experience," *Bell Labs Technical Journal*, vol. 16, no. 1, June 2011.
- [9] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnström, R. Olsson, and P. Le Callet, "Subjective quality assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission errors," *Int. Packet Video Workshop*, Dec. 2010.
- [10] ITU, "Methodology for the subjective assessment of the quality of television pictures," *Rec. ITU-R BT. 500-11*, 2002.
- [11] VQEG, "Validation of Reduced-Reference and No-Reference objective models for standard definition television, Phase I," 2009.
- [12] H. Coolican, *Research Methods and Statistics in Psychology*, 2nd ed. Hodder & Stoughton, 1994.