Advanced Science and Technology Letters
Vol.31 (ACN 2013), pp.16-19
http://dx.doi.org/10.14257/astl.2013.31.04

Measurement Method for Mean Opinion Score in Actual Home Environments

Chin-chol Kim[1], Sang-yong Ha[1], Byeong-hee Roh[2],
Jong-pil Jeong[3], Jae-young Choi[3]

[1] NIA Bldg, Cheonggyecheonno 14, Jung-gu, Seoul, South Korea
[2]Ajou University, Suwon, Gyeonggi, South Korea
[3]SungKyunkwan University, Suwon, Syeonggi, South Korea
{cckim, yong}@nia.or.kr, bhroh@ajou.ac.kr, {jpjeong, jychoi1001}@skku.edu

Abstract. As the demand for high-quality broadcast content, such as from Internet Protocol Television (IPTV) and Video on Demand (VoD) has greatly increased, broadcasters are striving to improve consumer quality of experience (QoE) to differentiate themselves from competitors. Recently, subjective measurement methods have been internationally standardized as the most reliable approach for measuring and evaluating Mean Opinion Score (MOS). However, it is actually difficult to apply standard evaluation methods to assess real broadcast watching environment because a majority of these methods are performed in experimental environments. Therefore, this paper proposes a method that accommodates actual viewing environments likewise home. Using the mean opinion score, we experimentally analyze the effects of evaluation interval changes under actual conditions in which home IPTV service is provided.

Keywords: IPTV, Quality of Experience, Subjective Video Quality Assessment, Mean Opinion Score.

1 Introduction

Unlike data service, if quality is deteriorated, such as by discontinuation of audio or lagging of video, in the broadcast service—including in IPTV, in which audio and video are streamed in real time—consumers strongly react, which marks the decrease in overall quality of service (QoS). Therefore, interest in quality of experience (QoE), which represents a measure of quality that people experience, is greatly increasing as mean opinion score (MOS) [1]. QoE is affected by various factors, such as video digitalization, and encoding and decoding in the head-end system, including packet loss and delays in networks and terminals (STB, etc.). The quality of each variable affecting QoE can be measured by objective methods; however, it is not easy to

* This research was supported in part by contribution of Korea Communications Commission and Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No.2010-0020737). Corresponding author : Jae-Young Choi.

ISSN: 2287-1233 ASTL

objectively measure and evaluate quality without using original video at the current level. In the current method, video quality is evaluated by the mean opinion score (MOS) to quantify the QoE when consumers watch a broadcast service. Methods to objectively measure QoE for video are classified by whether or not the video is compared to the original video. These methods include full reference (FR), reduced reference (RR), and no reference (NR) approaches [2]. However, these methods have some limitations in operational or practical aspects when evaluating picture quality [3]. To address the above issues, this paper experimentally analyzes the effects of change in the evaluation cycle (interval) on MOS in environments in which IPTV service is provided. In this paper, we propose a measurement method that is suitable for an actual viewing environment.

The rest of this paper is organized as follows. In Section 2, our tests of an evaluation cycle appropriate for an actual home IPTV evaluation environment are outlined. Our conclusions are presented in Section 3.

2 Proposed Evaluation Cycle for Live Broadcast IPTV QoE

ACR evaluation method [4] that is suitable for most evaluation approaches should be developed to evaluate QoE for videos that consumers view in real time in an actual IPTV watching environment. Especially, the measurement cycle was the greatest disadvantage of the ACR, which is a representative QoE measurement method mainly used in existing laboratory environments. For our study, it was improved to be suitable for an actual home TV viewing environment. The MOS of evaluators was measured with the target IPTV service being provided as a live broadcast. As evaluators watched the IPTV broadcast content (Evaluation contents were divided by diverse characteristics into news, drama, and music broadcasts) scheduled in advance in real time, their QoE of the picture was measured with MOS. Because entire sectors in which services were provided, such as head-end, network, and STB, contained original quality content at the time it was produced, the entire measurement sector represented the evaluators' QoE.

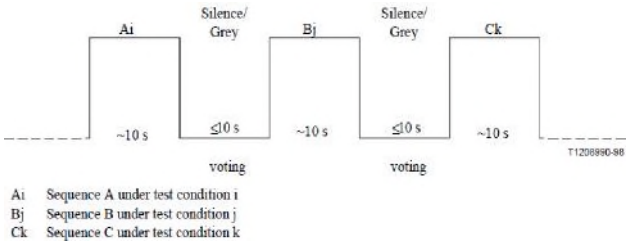

Fig. 1. ACR evaluation procedures

Evaluation procedures were carried out as follows. Picture QoE measurement software, with which viewers can rate MOS values in real time on their portable notebook or smart phone, was installed. The evaluation was performed according to instructions of the measurement program. If evaluators clicked the installed

measurement software prior to evaluation, the popup window where their basic information (gender, region, channel, and so on) is entered was opened and the alarm signaled the start of the evaluation. If users began the evaluation, the measurement software marked the evaluation time as a cycle of 10 minutes. Users could rate MOS values for their experience level about videos they watched within 20 seconds. Total evaluation time was 30 to 60 minutes based on the time needed to play the TV content. In addition, to increase the accuracy of evaluation results, users could not change the channel after the evaluation began. Once they gave their ratings with MOS values, the result would be transmitted in real time to a remote collection server. Measurement values which did not comply with the above procedures were excluded from the final data, which increased the accuracy of measurement results.

For the ACR measurements, all videos were separated by an interval of 10 seconds and a blank image was added to each of them for the 10 second evaluation period. On the contrary, in the IPTV experiments proposed for analysis with cycles of 2, 5, 10, and 30 minutes, the video continued to play during the evaluation without the addition of a blank image. The sound of an alarm at the end of each viewing window marked the start of its respective evaluation period.

The measurement scale was the same five-point rating system described earlier. The results of these measurements were compared with ACR and MOS values at 2, 5, 10, and 30 minute cycles.

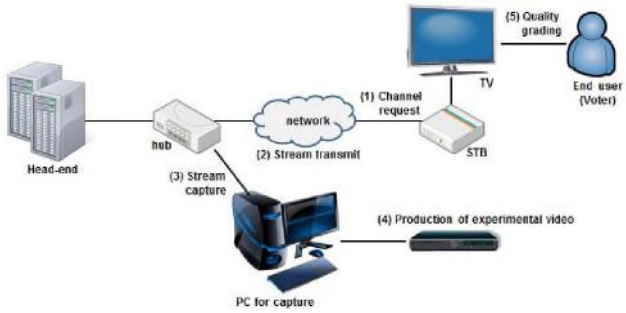

Fig. 2. Experimental scenario and topology

For the comparison, the number of MOS values measured during one evaluation cycle (2, 5, 10, and 30 minutes) was 1. However, in the case of the ACR, because every 10-second interval within the cycle of each evaluation was measured, and several values were derived, it was difficult to directly compare them to each other. Therefore, in this study, the average evaluation value for 10 seconds measured within the cycle was used as the ACR result value. However, most ACR result values consisted of values for evaluating the original video in which degradation was not applied when the average was calculated. Therefore, the deteriorated result value of the degraded video was not reflected. To address this problem, only the ACR values of a 10-second cycle in which degradation was applied were extracted, and the average of these values was used. Figure 2 shows the configuration in which the experimental video was produced and the experiment was carried out.

Values measured by ACR methods were obtained according to the test methods described in the above. These methods measured video quality in 10-second cycles

for three content types (documentary, drama and sports), and the MOS quality values were measured with video evaluation cycles of 2, 5, 10, and 30 minutes.

To obtain our experimental results, the video evaluation cycles appropriate for an actual IPTV viewing environment were selected. The averages for each content type and for entire videos were then calculated for analyzing the correlation of values measured by ACR methods and video evaluation cycles. This method was used to prove the validity of the selected evaluation cycle.

This experiment compared the evaluation cycle of 10 seconds, which was the standard method of ACR, with the changed evaluation cycle. It showed that the correlation between the measured MOS values was the largest in the evaluation method using a 10-minute cycle. The evaluator discomfort measurement and occurrence of experimental errors can be reduced by replacing the evaluation cycle that was found to have the greatest issues (when ACR was applied in an actual home environment) with a practical cycle that considers and applies evaluation in home environments. In addition, even if supplemented cycles were applied, the evaluation values of the entire MOS showed the same result value pattern obtained in a laboratory environment of the existing ACR. Based on this information, the high reliability of our experiment was confirmed. Therefore, an efficient and practical assessment was made in a real home environment by adhering to an actual evaluation cycle using evaluation techniques with existing ACR standards (e.g., absolute evaluation, rating evaluation)

3 Conclusion

To compensate for the evaluation cycle that is regarded as the drawback of the ACR standard, an evaluation cycle applicable to an actual evaluation environment was derived and verified. The findings showed that the correlation with the measurement result from the ACR method, which was standard in its cycle of 10 minutes, was the highest. This result implied that the result had the same reliability as measurements with a 10-second cycle, even if it was completed with a 10-minute cycle and users comfortably measured it. Finally, it was found that the drawback of the ACR could be complemented and utilized to measure QoE for video in real time.

References

1. A. Takahashi, D. Hands, V. Barriac, “Standardization Activities in the ITU for a QoE Assessment of IPTV,” IEEE Communications Magazine, Vol.46, No.2, pp.78-84, 2008.
2. ITU-T J.143 User Requirements for Objective Perceptual Video Quality Measurements in Digital Cable Television Series J: Transmission of Television, Sound Programme and Other Multimedia Signals Measurement of the Quality of Service, 2000.
3. K. Yamagishi, T. Hayashi, “Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services,”, IEEE ICC’2008, 2008.
4. ITU-T Rec. P.910, “Subjective video quality assessment methods for multimedia applications,” ITU-T, 1999.

