Development of Fuzzy Relationships between Different Quality of Experience Parameters

Marko Matulin, Štefica Mrvelj
Department of Information and Communications Traffic
Faculty of Transport and Traffic Sciences, University of Zagreb
Zagreb, Croatia
mmatulin@fpz.hr, smrvelj@fpz.hr

Abstract—**User psycho-acoustic and visual perception of telecommunication service quality depends on a wide palette of objective and subjective parameters. This is especially relevant in the process of evaluation, assessment or prediction of user Quality of Experience (QoE) for specific service, because these processes require the revelation of the correlations between different parameters. Using the User Datagram Protocol (UDP) based video streaming service as an example, in this paper we investigate the impact of three objective parameters (the packet loss rate, the number and the duration of packet loss occurrences) on user perception by conducting subjective quality tests. We then use Fuzzy C-Means (FCM) clustering approach to cluster test subjects' ratings as well as to define degrees of membership of each rating to specific fuzzy cluster. Based on the obtained degrees of membership, the fuzzy relationships between different objective and subjective parameters are approximated using the normal distributions.**

Keywords- Quality of Experience, fuzzy logic, subjective and objective parameters, fuzzy relationships, evaluation

I. Introduction

One of the paramount objectives of todays' network operators and service providers is to understand the relationship between achieved network performance and user perception regarding the quality of a specific service. This is useful when attempting to improve network efficiency, reduce operating costs and maintain certain levels of user satisfaction. However, defining and modeling the relationships between different Quality of Experience (QoE) parameters has proven to be a difficult task. This is due to a large set of quantitative (objective) and qualitative (subjective) parameters which are important from user psycho-acoustic and visual perspective.

Understanding how a user perceives the quality of a specific service normally requires subjective quality testing. Subjective evaluation that involves test subjects is the only true mean of discovering user opinions about what is presented to them as a service. The subjective evaluation is usually conducted in a controlled, laboratory conditions, however some authors conducted real-life tests; for instance, see [1-6].

Though crucial for discovering the relationships between the parameters, it is clear that the subjective evaluation requires commitment of considerable resources. Therefore, the objective quality assessment models have been developed for different services. Some of these models can estimate subjective quality, as perceived by users, solely from objective quality measurement or indices. They are usually based on different relationships between the network and the application oriented parameters (such as network delay, jitter, packet loss, bitrate, framerate, video artifacts, re-buffering frequency etc.). In the light of this study, we have focused the following review of the related work on the objective video quality assessment models.

ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) in [7] grouped the objective video quality assessment models into three main categories: a) Full-Reference (FR) models for which both the reference input and the degraded (processed) output video signal have to be available, b) Reduced-Reference (RR) models that require access to the degraded output and some limited features extracted from the reference input video signal, and c) No-Reference (NR) models that require access only to the degraded output video signal.

In order to assess the quality of a video signal, the objective models employ statistical models that ITU-T in [8] classifies into: a) media-layer models which use media signals to quantify quality degradation, b) packet-layer models which use only the information extracted from packet headers, c) bitstream-layer models which, apart from packet header information, use payload information, d) hybrid models which combine the features of previously listed models, and e) planning models which include the quality planning parameters of networks or terminals.

It is noteworthy to mention that not all video quality assessment models use subjective data sets for the assessment. Some models or metrics (e.g., Mean Square Error, Peak Signal-to-Noise Ratio, Moving Picture Quality Metrics, Modified Sum of Absolute Difference, Structural Similarity Index and others) assess the video quality solely from analysing the properties of the output signal (image luminance, saturation, structure, etc.).

To model the relationships between objective and subjective parameters different methods may be applied. In [9] the authors showed how the random neural network can be trained with the subjective data set to assess the quality of a video stream at the receiver side. Similarly to [9], in [10] the authors also use the neural network for assessing the QoE of HTTP (Hypertext Transfer Protocol) video streaming. Another example of neural network application for QoE assessment for HTTP video streaming can be found in [11].

A machine learning approach was employed by Menkovski et al. to develop the objective model which can determine the extent of video quality degradations that may lead toward unacceptability of a video quality as perceived by test subjects (see [12-13]). In [14-15] the authors construct k-dimensional Euclidian space, where k represents the number of network dependent and independent parameters which may affect user QoE. The authors define zones in the k-dimensional space where different values of various parameters lead to the same QoE rating. Zhang et al. in [16] use fuzzy decision trees to predict user QoE from the log data collected from different Internet video service providers in China.

From this brief literature overview it can be seen that different techniques are employed for modelling the user perception. As discussed earlier, these type of models are based on the relationships between the objective and the subjective parameters. Thus, accurate assessment of user QoE means that these relationships have to be disclosed.

Using the User Datagram Protocol (UDP) based video streaming service as an example, in this contribution we show how the relationships between the values of the three objective network parameters (packet loss rate, the number of packet loss occurrences in the sequence and the duration of those occurrences) and the subjective perception (level of user annoyance) can be modelled. For this purpose, we use fuzzy logic. Based on the developed relationships, in our future research we will develop no reference objective video quality assessment model for assessing the user QoE for UDP-based video streaming service.

The paper is structured as follows. Chapter 2 brings short description of the subjective evaluation of video quality that was used to collect user opinions about specific video quality degradations. The fuzzification process is described in chapter 3. This chapter also presents the developed fuzzy membership functions. Discussion and concluding remarks can be found in chapter 4.

II. SUBJECTIVE EVALUATION OF VIDEO QUALITY

In this paper we develop fuzzy relationships between different objective and subjective parameters which may affect user QoE for UDP-based video streaming service. The obtained relationships, i.e. fuzzy membership functions, originate from the results of subjective evaluation of user QoE which was conducted in real-life environments. Therefore, it is meaningful to firstly discuss the research method and the obtained results from that study.

A. Research method

In order to test user QoE for the video streaming, 72 test sequences were prepared for rating. We used only one type of video (documentary film) that lasted one hour and it was chosen using the entertainment-oriented content selection [17]. The video was encoded using Advanced Video Coding (H.264/AVC) and Advanced Audio Coding (AAC). The video was coded at a bitrate of 9.8 Mbps and a framerate of 50 fps. The resolution of the video was 1920 x 1080 pixels, while the audio was coded at a bitrate of 256 kbps.

The video was streamed in an emulated network environment six times; each time different packet loss rates (PLRs) were introduced (0.05, 0.1, 0.5, 1, 1.5 and 2%) using the emulator client. Six incoming video signals were stored in the same format as the unprocessed video.

In the next phase, 1, 4, 7 or 10 short video clips from a degraded video signal were inserted into the original video signal. The duration of a single inserted clip, i.e., a single packet loss occurrence (PLO), varied between 1, 4 and 7 seconds. Different combinations of the number of PLOs and the duration of a single PLO resulted in different total duration of all PLOs in a test sequence that varied between 1, 4, 7, 10, 16, 28, 40, 49 and 70 seconds.

Test subjects who participated in the subjective evaluation were picked from student population of the University of Zagreb, Faculty of Transport and Traffic Sciences. This is due to two main reasons: a) video streaming services are generally used by users between the ages of 18 and 24, which corresponds with the age group of a typical student population [18], and b) this population was easy accessible for conducting such a survey, i.e. the convenience sampling method was used [19].

In order to deliver the prepared test sequences to the test subjects we considered employing several methods, namely: a) QoE crowdtesting, b) remote performance monitoring during the network streaming, and c) offline evaluation when previously prepared test sequences are delivered to the subjects (e.g., on Blu-ray or DVD disks) and they rate it offline. After analyzing each method, we have decided to employ the offline evaluation using DVDs. The offline evaluation allowed us to prepare the test sequences in the emulated network environment and then distribute them to the subjects for rating in real-life environment. We also concluded that it would be difficult task to pursue the subjects to participate in the survey if they would have to download or stream one-hour video in high definition resolution to their devices from their homes.

Prior to watching the video, the subjects were ignorant about what would be expected from them after watching. They were only asked to watch the sequence in the everyday conditions as they would normally watch television and to open the sealed envelope, that contained the questionnaire, after the screening.

After the data processing, in which some of the methods discussed in [20] were employed for the exclusion of unreliable responses, 602 questionnaires were accepted for further analysis.

B. Obtained results

In this study we analyze the impact of three objective parameters (the PLR, the number of PLOs in the sequence and the duration of those occurrences) on user QoE. In the questionnaire, used in the study, we asked our test subjects to evaluate the level of their annoyance in relation to: a) the observed quality distortions, i.e. video artifacts (we correlated these responses with the values of PLRs and the results can be found in Figure 1), b) the number of PLOs which they have noticed (results of this analysis are depicted in Figure 2), and c) the total duration of all quality distortions in the video (the results can be found in Figure 3).

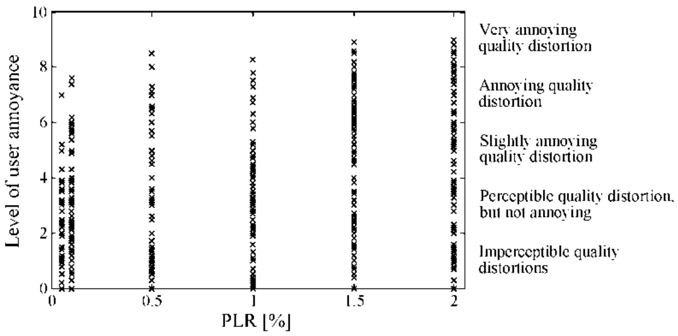

Figure 1. Level of user annoyance in relation to PLRs

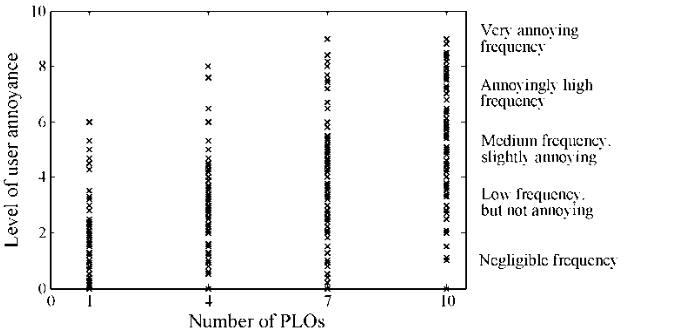

Figure 2. Level of user annoyance in relation to the number of packet loss occurrences (PLOs)

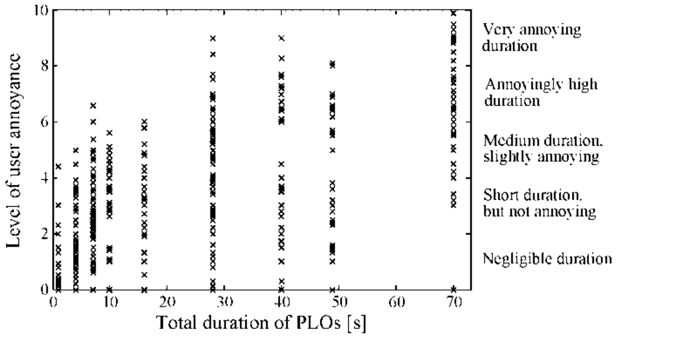

Figure 3. Level of user annoyance in relation to the total duration of packet loss occurrences (PLOs)

The subjects rated the level of their annoyance on an 11-point scales [21], which allow the linguistic meanings of different grades to be added as a help during rating. This feature makes the scale suitable for exploring user opinions, which are usually fuzzy in nature. The meanings that are used in this study are depicted on the secondary y-axes of Figure 1, 2 and 3.

We can observe how the level of user annoyance was sometimes spread over all the annoyance level categories. This is most conspicuous for the data presented in **Chyba! Nenašiel sa žiaden zdroj odkazov.**, and can be explained by agreeing that, e.g., PLR of 2% was perceived as *Imperceptible quality distortion* when the sequence contained only one PLO. Yet, the same PLR may be perceived as *Very annoying quality distortion* if the sequence contains 10 PLO, with 70 seconds of quality degradations. Similar results are reported in [22] where it is stressed how the correlation between the PLR and user opinions cannot be unambiguously defined.

III. RESULTS OF THE FUZZIFICATION PROCESS

A. *Fuzzy clustering*

In the fuzzification process we have used Fuzzy C-Means (FCM) clustering approach [23] for grouping the data points presented in Figure 1, 2 and 3. The objective was to group the data points presented on the figures into fuzzy clusters and to find centers of those clusters. The FCM method allows the clusters to overlap, i.e., specific data point can be a member of several clusters with different degrees of membership (u_i).

According to [23], the FCM procedure is based on minimization of the objective function J_m expressed by Eq. (1).

$$J_m = \sum_{i=1}^{L} \sum_{j=1}^{C} {u_{ij}}^m \cdot \left\|x_i - c_j\right\|^2 \tag{1}$$

The m in Eq. (1) represents any real number greater than 1 (we have set $m = 2$), u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i-th of d-dimensional measured data, c_j is the d-dimension center of the cluster and $|| * ||$ is any norm expressing the similarity between any measured data and the center. The fuzzy partitioning is carried out through an iterative optimization of the objective function J_m, with the update of membership u_{ij} (Eq. (2)) and the cluster centers c_j (Eq. (3)). The iterative process finishes when the stopping criteria ε is met (Eq. (4); k are the iteration steps). In our case, we used $\varepsilon = 10^{-5}$. Note that all the equations are taken from [23].

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|} \right)^{\frac{2}{m-1}}} \tag{2}$$

$$c_j = \frac{\sum_{i=1}^{N} {u_{ij}}^m \cdot x_i}{\sum_{i=1}^{N} {u_{ij}}^m} \tag{3}$$

$$max_{ij}\left\{\left|{u_{ij}}^{(k)} - {u_{ij}}^{(k-1)}\right|\right\} < \varepsilon \tag{4}$$

The data points presented in Figure 1, 2 and 3 are grouped into three clusters. The FCM procedure for the PLR, Number of PLOs and Total duration of PLOs required 24, 69 and 33 iterations, respectively, before stopping criteria ε was met. The results are presented in Figure 4, 5 and 6, respectively (the centers of the clusters are marked with circles).

eISSN: 2453-6571, cdISSN: 1339-5076
10.18638/rcitd.2016.4.1.85
ISBN: 978-80-554-1271-9

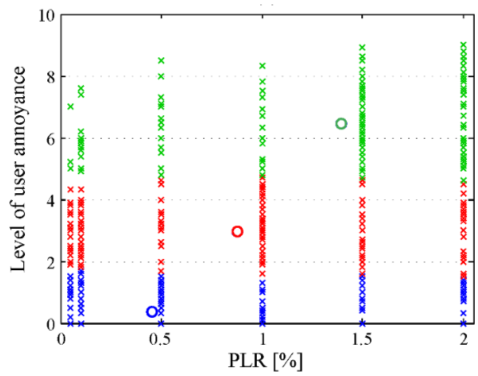

Figure 4. Results of the FCM clustering for the parameter: PLR

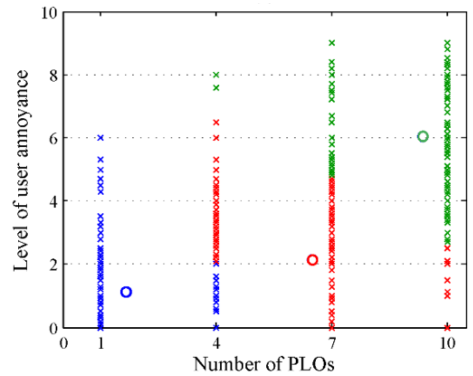

Figure 5. Results of the FCM clustering for the parameter: Number of PLOs

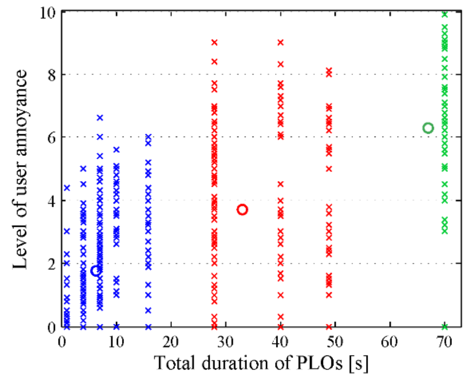

Figure 6. Results of the FCM clustering for the parameter: Total duration of PLOs

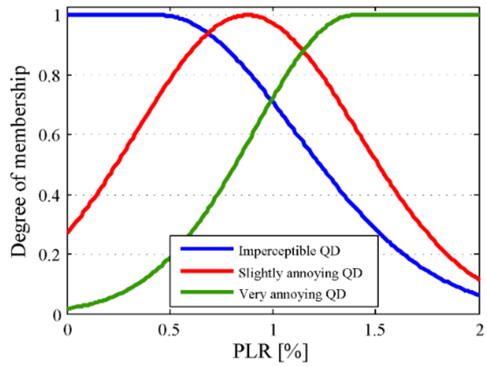

Figure 7. Membership functions of the three fuzzy clusters of the first objective parameter (QD stands for *quality distortion*)

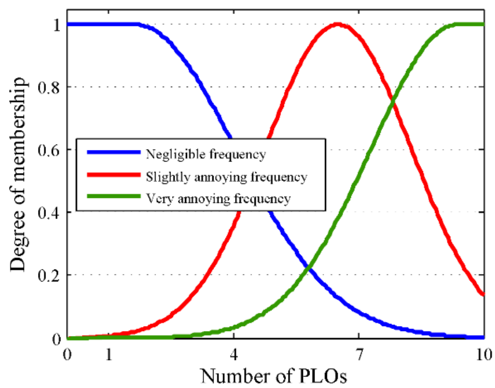

Figure 8. Membership functions of the three fuzzy clusters of the second objective parameter

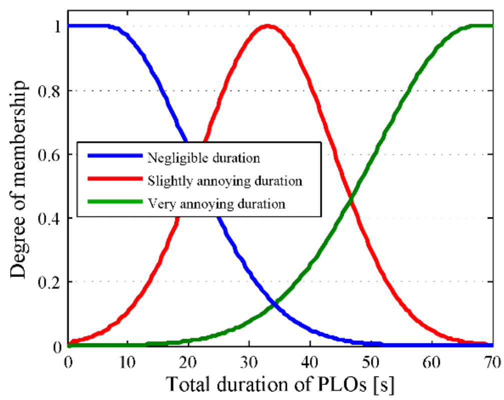

Figure 9. Membership functions of the three fuzzy clusters of the third objective parameter

B. *Fuzzy relationships*

By correlating the values of the data points, shown on the x axis of Figure 4, 5 and 6, with their degrees of membership u_{ij} to the specific cluster j, the boundaries of the clusters can be approximated. We have used the normal distribution for the approximation, since it is reported in [24] that normal distribution corresponds better with the changes in human perception. The transitions between different user opinions and attitudes are usually happening gradually ("smoothly") and bell-shaped functions can mimic that behavior better compared with, for instance, triangular functions. The boundaries of the clusters, i.e. the obtained fuzzy membership functions for each of the three objective parameters are presented in Figure 7, 8 and 9.

Each cluster shown in Figure 7, 8 and 9 is named since the fuzzy systems usually use linguistic variables to infer conclusions. Properties of the fuzzy membership functions shown in the figures can be found in Table 1. Note that the blue and the green Gaussian functions of each input parameter are modified so that u_{ij} would be equal to 1 in cases when $x_i < \bar{x}$ (for the blue functions) and $x_i > \bar{x}$ (for the green functions).

TABLE I. PROPERTIES OF THE MEMBERSHIP FUNCTIONS

Figure	Membership functions		
	1st membership function (blue)	*2nd membership function (red)*	*3rd membership function (green)*
7	$\bar{x} = 0.4545$ $\sigma = 0.6574$	$\bar{x} = 0.8758$ $\sigma = 0.5398$	$\bar{x} = 1.3937$ $\sigma = 0.4887$
8	$\bar{x} = 1.6513$ $\sigma = 2.4$	$\bar{x} = 6.5083$ $\sigma = 1.748$	$\bar{x} = 9.3728$ $\sigma = 2.061$
9	$\bar{x} = 6.4254$ $\sigma = 13.73$	$\bar{x} = 33.0713$ $\sigma = 10.92$	$\bar{x} = 67.1134$ $\sigma = 16.33$

The figures are showing how the clusters are overlapping. The overlapping is the most distinctive between the clusters of the first objective parameter PLR. This is due to the previously discussed fact how certain PLRs were evaluated differently by the subjects, depending on the number of PLOs in a test sequence and their total duration.

IV. DISSCUSSION AND CONCLUSIONS

From the obtained results and the developed fuzzy membership functions it can be derived that human perception of video quality degradations cannot be unambiguously correlated with the specific values of the objective (network) parameters. The distinctive overlapping between different fuzzy clusters serves as an evidence to that claim.

As previously discussed, the origin of the overlapping comes from the affiliated effect of the three objective parameters on user psycho-acoustic and visual perception. We recorded cases when higher PLRs were not adversely perceived by the subjects because the number of PLOs in the sequence remained low. However, the increase of PLOs and/or their total duration clearly increased the level of user annoyance, meaning that even lower PLRs may be perceived as *Annoying* or *Very annoying* by the subjects.

Incorporating this fuzziness of the user perception of service quality into the membership functions of the clusters is one of the distinguishing features of this study. This was achieved using the FCM clustering method. The applied method can be considered as a valuable tool for resolving cluster consensus problems in engineering, when a limited knowledge about the number of clusters often exists. Additionally, the method repetition, i.e. iterative optimization of the objective function J_m, is also awarding in terms of gaining the confidence about the significance of the achieved results.

Another possible application of the method would be for the dynamical clustering, a concept discussed in [25], when the membership functions could be updated over time. This could be especially useful, knowing that a given amount of change of the objective parameters has a different impact on resulting change of QoE, depending on the current level of QoE [6].

It is important to stress that the obtained results are highly influenced by the research method that was used. The subjective evaluation was conducted in a real-life test conditions and the subjects rated the video that lasted one hour. This means that test subjects were not focused on noticing and memorizing the quality distortions, as well as the effect on human short-term memory and recency effect must not be ignored. Furthermore, the video used in this study contained video subtitles. Video subtitles may affect user rating by diverting the user focus from the picture to the bottom of the screen. Some quality distortions, thus, remain unnoticed and that directly affects the rating [6].

With this contribution we made an effort towards developing the no reference objective video quality assessment model for assessing the user QoE for UDP-based video streaming service. The inference system of the model will be further built upon the relationships that were investigated in this paper. Since the subjective evaluation and the developed fuzzy relationships are reflecting real-life conditions, we believe that the developed model will stand out in a group of similar models that usually are using the subjective data sets collected in a controlled, laboratory environments.

REFERENCES

[1] P. Reichl, P. Fröhlich, L. Baillie, R. Schatz, and A. Dantcheva, "The Liliput prototype: a wearable lab environment for user tests of mobile telecommunication applications," in Human Factors in Computer Systems, New York, 2007, pp. 1833-1838.

[2] S. Jumisko-Pyykkö and M. Hannuksela, "Does Context Matter in Quality Evaluation of Mobile Television?," in 10th Conference on Human-Computer Interaction with Mobile Devices and Services, Amsterdam, 2008, pp. 63-72.

[3] N. Staelens et al., "Assessing Quality of Experience of IPTV and Video on Demand Services in Real-Life Environments," IEEE Transactions On Broadcasting, vol. 56, pp. 458-466, 2010.

[4] S. Ickin et al., "Factors Influencing Quality of Experience of Commonly Used Mobile Applications," IEEE Communications Magazine, pp. 48-56, April 2012.

[5] N. Staelens et al., "Subjective Quality Assessment of Longer Duration Video Sequences Delivered Over HTTP Adaptive Streaming to Tablet Devices," IEEE Transactions on Broadcasting, vol. 60, no. 4, pp. 707-714, 2014.

[6] Š. Mrvelj and M. Matulin, "Impact of Packet Loss on the Perceived Quality of UDP-based Multimedia Streaming: A Study of User Quality of Experience in Real-life Environments," Multimedia Systems, pp. 1-21, 2016. (Online first, doi:10.1007/s00530-016-0531-8)

[7] ITU-T, "User requirements for objective perceptual video quality measurements in digital cable television," International Telecommunication Union, ITU-T Rec. J.143 2000.

[8] ITU-T, "Reference guide to quality of experience assessment methodologies," International Telecommunication Union, ITU-T Rec. G.1011 2010.

[9] S. Mohamed and G. Rubino, "A Study of Real-Time Packet Video Quality Using Random Neural Networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 12, pp. 1071-1083, 2002.

[10] R. Wang, Y. Geng, Y. Ding, Y. Yang, and W. Li, "Assessing the Quality of Experience of HTTP Video Streaming Considering the Effects of Pause Position," in 16th Asia-Pacific Network Operations and Management Symposium, Hsinchu, 2014, pp. 1-4.

[11] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of Experience Estimation for Adaptive HTTP/TCP Video Streaming Using H.264/AVC," in IEEE Consumer Communications and Networking Conference, Las Vegas, 2012, pp. 127-131.

[12] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra Sánchez, "Predicting Quality of Experience in Multimedia Streaming," in 7th International Conference on Advances in Mobile Computing and Multimedia, Kuala Lumpur, 2009, pp. 52-59.

[13] V. Menkovski, G. Exarchakos, A. Liotta, and A. Cuadra Sánchez, "Estimations and Remedies for Quality of Experience in Multimedia Streaming," in 3rd International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services, Nice, 2010, pp. 11-15.

[14] M. Venkataraman, M. Chatterjee, and S. Chattopadhyay, "Evaluating Quality of Experience for Streaming Video in Real Time," in Global Telecommunications Conference, 2009, pp. 1-6.

[15] M. Venkataraman and M. Chatterjee, "Inferring Video QoE in Real Time," IEEE Network, vol. 25, no. 1, pp. 4-13, 2011.

[16] Y. Zhang, T. Yue, H. Wang, and A. Wei, "Predicting the Quality of Experience for Internet Video with Fuzzy Decision Tree," in IEEE 17th International Conference on Computational Science and Engineering, Chengdu, 2014, pp. 1181-1187.

[17] M.H. Pinson, K.S. Boyd, J. Hooker, and K. Muntean, "How to Choose Video Sequences for Video Quality Assessment," in 7th International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, 2013, pp. 79-85.

[18] P. Datta, L. Izdebski, N. Kumar, and K. Suh, "“It Came to Me in a Stream…” The Upward Arc of Online Video, Driven by Consumers," Cisco Internet Business Solutions Group, Cisco, white paper 2012.

[19] F. Farrokhi and A. Mahmoudi-Hamidabad, "Rethinking Convenience Sampling: Defining Quality Criteria," Theory and Practice in Language Studies, vol. 2, no. 4, pp. 784-792, 2012.

[20] T. Hoßfeld et al., "Best Practices for QoE Crowdtesting: QoE Assessment With Crowdsourcing," IEEE Transactions on Multimedia, vol. 16, no. 2, pp. 541-558, 2014.

[21] ITU-T, "Subjective Video Quality Assessment Methods for Multimedia Applications," International Telecommunication Union, ITU-T Rec. P.910 2008.

[22] D. Hernando, J. E. López de Vergara, D. Madrigal, and F. Mata, "Evaluating Quality of Experience in IPTV Services Using MPEG Frame Loss Rate," in International Conference on Smart Communications in Network Technologies, Paris, 2013, pp. 1-5.

[23] M. Alata, M. Molhim, and A. Ramini, "Optimizing of Fuzzy C-Means Clustering Algorithm Using GA," World Academy of Science, Engineering and Technology, no. 15, pp. 224-229, 2008.

[24] Timothy J. Ross, Fuzzy Logic with Engineering Applications, 2nd ed. Chichester: John Wiley & Sons Ltd., 2004.

[25] Y. Shang, "A combinatorial necessary and sufficient condition for cluster consensus," Neurocomputing, 2016. (Online first, doi:10.1016/j.neucom.2016.08.025)