Towards an Overall QoE Model for 360-Degree Video

Huyen T. T. Tran*, Nam Pham Ngoc†, Truong Cong Thang*
*The University of Aizu, Aizuwakamatsu, Japan
†VinUniversity, Vietnam

2020 IEEE Eighth International Conference on Communications and Electronics (ICCE) | 978-1-7281-5471-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCE48956.2021.9352076

***Abstract*—Although 360-degree video is becoming more and more popular on the Internet, understanding of Quality of Experience (QoE) of 360-degree video is still limited. In this paper, we aim to investigate for the first time the impacts of two QoE aspects, namely perceptual quality and presence, on the overall QoE of 360-degree video. By using a subjective dataset, statistical analysis shows that the overall QoE is significantly affected by the perceptual quality and presence. In addition, we make the first attempt to model the overall QoE of 360-degree video taking into account the impacts of the QoE aspects. In particular, two approaches are investigated to predict the overall QoE. The result shows that one of the approaches is applicable to predict the overall QoE of 360-degree video over different videos and rendering device sets.**

***Index Terms*—Quality of Experience, Acceptability, Quality model, 360-degree video, Virtual Reality**

I. Introduction

Thanks to cost decreases and usability increases of virtual reality (VR) devices, 360-degree video (or 360 video for short) has gradually been gaining popularity on streaming platforms such as YouTube and Facebook in recent years. Different from traditional video, 360 video is capable of providing a 360-degree view of a scene, and so immersive viewing experience to users. However, user experience when watching 360 video is rather complex and so not fully understood yet. Also, to provide excellent immersive experience, it is recommended that 360 video should have high quality and high resolution, resulting in a bulky size [1]–[3]. Therefore, for effective generation, storage and transmission of 360 video, it is crucial to obtain in-depth understanding of Quality of Experience (QoE) of 360 video.

For traditional video, perceptual quality and acceptability are two key QoE aspects, which have been extensively studied in the literature [4]–[7]. Perceptual quality refers to the degree of user satisfaction of video quality being displayed on rendering devices. Acceptability indicates whether a service or application is acceptable to users. In some previous studies [6], [8], the acceptability is considered as the overall QoE of a service. For 360 video, in addition to the perceptual quality and acceptability, presence is another important QoE aspect. Presence refers to the sense of "being there" in the VR environment with interactions like in the real environment [9].

Although QoE of traditional video has been thoroughly investigated in previous studies [4], [6], [7], researches on QoE of 360 video are still very limited. Most existing studies focus on the perceptual quality aspect [10]–[13]. There are only a few studies on the presence aspect such as [14], [15]. Especially, there is no study on the acceptability aspect of QoE of 360 video as well as the impacts of QoE aspects on the overall QoE of 360 video.

There have been some attempts to build QoE models of 360 video such as [16]–[19]. However, previously proposed models are to model a specific QoE aspect such as cybersickness [16] and perceptual quality [17]–[19]. None of these models are for the overall QoE of 360 video which combine user perceptions of QoE aspects.

To fill this gap, we first examine the impacts of the two QoE aspects of perceptual quality and presence on the acceptability which is used to represent the overall QoE of 360 video in this study. Then, we investigate two approaches to model the acceptability based on the perceptual quality and presence. To the best of our knowledge, this study is the first that presents approaches for modeling the overall QoE of 360 video.

The remaining of this paper is structured as follows. In Section II, we investigate the impacts of the perceptual quality and presence on the acceptability. Section III presents and discusses our approaches for predicting the acceptability of 360 video. Finally, Section IV gives a conclusion for the paper.

II. Impacts of the Perceptual Qualiy and Presence

To investigate the impacts of the perceptual quality and presence on the acceptability, we use a subjective dataset presented in [15]. This dataset consists of totally 60 video versions generated from three original 360 videos. The content features of these videos are described in Table I. We can see that the used videos have different content features. In addition, two rendering device sets, denoted D#1 and D#2, are used to display video versions in this dataset. In particular, device set D#1 consists of a Samsung Galaxy S6 phone and a Samsung Gear VR. The Samsung Galaxy S6 has the display size of 5.1 inches and the screen resolution of 1440×2560 [20]. Device set D#2 consists of a Samsung Galaxy S5 phone and a Google Cardboard. The Samsung Galaxy S5 has the display size of 5.1 inches and the screen resolution of 1080×1920 [20].

In this study, we conduct Kruskal-Wallis tests over six subsets (S#1∼S#6) of this dataset to determine if there are statistically impacts of the QoE aspects on the overall QoE across different videos and device sets. In particular, the first subset (S#1) consists of all the video versions in the dataset.

978-1-7281-5471-8/20/$31.00 ©2020 IEEE

TABLE I

A DESCRIPTION OF CONTENT FEATURES OF THREE ORIGINAL VIDEOS IN THE DATASET [15]

Content	Content motion	Description
Content #1	Static camera, Few moving objects, Static background	The camera is fixed to the floor. Characters take part in a cooking contest in a kitchen room.
Content #2	Medium camera motion, Many moving objects, Dynamic background	The camera is held by a diver. Dolphins move around in the ocean.
Content #3	Fast camera motion, Many moving objects, Dynamic background	The camera is in a roller coaster moving at a high speed.

TABLE II

STATISTICAL RESULTS ABOUT THE EFFECTS OF THE PERCEPTUAL QUALITY AND PRESENCE ON THE ACCEPTABILITY. THE BOLD, ITALIC, AND UNDERLINED NUMBERS IN COLUMN η^2 RESPECTIVELY CORRESPOND TO "LARGE", "MODERATE", AND "SMALL" EFFECT SIZES.

Subset	Aspect					
	Perceptual quality			*Presence*		
	χ^2	p-value	η^2	χ^2	p-value	η^2
S#1 (Full dataset)	1071.3	<0.0001	**0.50**	975.1	<0.0001	**0.45**
S#2 (Video #1)	326.5	<0.0001	**0.45**	323.1	<0.0001	**0.45**
S#3 (Video #2)	400.1	<0.0001	**0.56**	341.9	<0.0001	**0.48**
S#4 (Video #3)	378.5	<0.0001	**0.53**	324.7	<0.0001	**0.45**
S#5 (D#1)	584.9	<0.0001	**0.54**	510.9	<0.0001	**0.47**
S#6 (D#2)	521.2	<0.0001	**0.48**	475.5	<0.0001	**0.44**

For the three next subsets (S#2∼S#4), each is composed of the versions generated from one of the three original videos. Specifically, subsets S#2, S#3, and S#4 respectively correspond to Videos #1, #2, and #3. For subset S#5, it is comprised of the versions rendered by device set D#1. Subset S#6 consists of the versions that are watched using device set D#2.

The obtained results of the Kruskal-Wallis tests are shown in Table II. Based on Cohen's conventions [21], eta-squared values η^2 can be used to interpret effect sizes. In particular, thresholds of η^2 are respectively 0.01, 0.06, and 0.14 for "small", "moderate", and "large" effect sizes. From Table II, it can be noted that both the perceptual quality and presence have statistically significant impacts on the acceptability in all the subsets (i.e., $p < 0.05$). In addition, the sizes of these effects are all "large" (i.e., $\eta^2 > 0.14$). Thus, for all the considered videos and rendering device sets, both the quality of displayed video and the presence must be satisfactory in order that a 360 video service is acceptable to users.

Besides, based on the subjective dataset, we investigate the impacts of two other factors of content motion and rendering mode on the acceptability, perceptual quality, and presence. Table III shows the statistical results from the Kruskal-Wallis test using subset S#1. It is found that there are no statistically significant effects of the content motion and rendering device on the acceptability (i.e., p>0.05). Meanwhile, the statistically significant effects of these factors with the "small" sizes are found for the presence and the perceptual quality (i.e., p<0.05, $0.06 > \eta^2 > 0.01$). This implies that, although the perceptual quality and presence have "large" impacts on the acceptability, factors having significant impacts on the presence and the perceptual quality may not cause considerable influences on the acceptability.

III. OVERALL QUALITY MODELS

In this section, we investigate two approaches to model the overall QoE (i.e., acceptability) of 360 video. Also, relationships between the QoE aspects and the overall QoE in the models are discussed.

A. Linear model

In the first approach, we simply use a weighted sum of the normalized QoE aspect values, which is a linear regression, to model the acceptability. The aim of this approach is to additionally investigate the impacts of QoE aspects. In particular, the acceptability Acc is given by

$$Acc = a \times \overline{PQ} + b \times \overline{PS} + c, \quad (1)$$

where a, b, and c are parameters to be trained, $\overline{PQ}$ and $\overline{PS}$ are respectively the normalized perceptual quality and presence values.

To determine the values of the parameters, the training process is conducted using six different training sets corresponding to the six subsets (S#1∼S#6) described in Section 2. For each training set, the corresponding test set consists of the remaining versions in the dataset. For example, the test set corresponding to training set S#2 is comprised of the versions of Videos #2 and #3.

Table IV shows the values of the parameters in the first approach for the different training sets. We can see that, for all the training sets, the value of b is always significantly higher than that of a. This reveals that the presence has a more important contribution to the acceptability than the perceptual quality. This can be explained that, when watching

TABLE III

STATISTICAL RESULTS ABOUT THE EFFECT OF THE CONTENT MOTION AND RENDERING MODE ON THE ACCEPTABILITY, PERCEPTUAL QUALITY, AND PRESENCE USING SUBSET S#1. THE BOLD, ITALIC, AND UNDERLINED NUMBERS IN COLUMN η^2 RESPECTIVELY CORRESPOND TO "LARGE", "MODERATE", AND "SMALL" EFFECT SIZES.

Factor	Aspect	χ^2	p-value	η^2
Content motion	Acceptability	0.81	0.668	0.001
	Perceptual quality	13.18	0.001	0.012
	Presence	23.83	<0.0001	0.022
Rendering device	Acceptability	0.09	0.763	<0.0001
	Perceptual quality	58.1	<0.0001	0.027
	Presence	24.64	<0.0001	0.011

TABLE IV

PARAMETERS OF THE FIRST APPROACH FOR DIFFERENT TRAINING SETS

Training sets	Parameters		
	a	b	c
S#1 (Full dataset)	0.39	1.33	-0.24
S#2 (Video #1)	0.10	1.69	-0.19
S#3 (Video #2)	0.51	1.23	-0.29
S#4 (Video #3)	0.45	1.31	-0.28
S#5 (D#1)	0.21	1.46	-0.26
S#6 (D#2)	0.85	1.02	-0.25

360 video, users expect to have better perception of the presence comparing to watching traditional video, and so they pay more attention to this QoE aspect.

Also, it is interesting that the values of the parameters are variable across different training sets. In particular, for training set S#2, the value of a is smallest (i.e., 0.10), whereas the value of b is largest (i.e., 1.69). This result suggests that the impact of the perceptual quality is smallest while the impact of the presence is largest for Video #1. This could be explained by the fact that Video #1 is created by a static camera that may cause negative influences on the presence of users. Consequently, Video #1 has the wost presence values among the three used videos [15]. This makes users give more notices to the presence when watching this video.

For training set S#3, the values of a and b are respectively 0.51 and 1.23. It can be seen that, comparing to training set S#2, the value of a is much more higher, while the value of b is considerably lower. The reason may be because the camera in Video #2 moves at a medium speed, resulting in the higher presence values comparing to that of Video #1 [15]. Therefore, when watching Video #2, users feel more satisfactory with the presence, and so place more emphasis on the perceptual quality comparing to watching Video #1.

In addition, we can see that the differences of the parameter values between training sets S#3 and S#4 are small. In particular, the value of a with training set S#3 is slightly higher than that with training set S#4 (i.e., 0.51 vs. 0.45). Similarly, the value of b for training set S#3 is a litter bit lower than that for training set S#4 (i.e., 1.23 vs. 1.31). This is because the camera in Video #3 moves at a fast speed, leading to the worse presence values than those of Video #2 [15]. Therefore, comparing to Video #2, Video #3 has a slightly larger impact of the presence and a smaller effect of the perceptual quality.

We can see that the value of a with training set S#5 is significantly lower than that with training set S#6. Meanwhile, the value of parameter b of training set S#5 is noticeably higher that of training set S#6. This result may be explained by the fact that device set D#1 has a substantially higher screen resolution comparing to device set D#2 (i.e., 1440x2560 vs. 1080x1920). Hence, device set D#2 has lower perceptual quality values than device set D#1 [15]. This makes users pay more attention to the perceptual quality when using device set D#2.

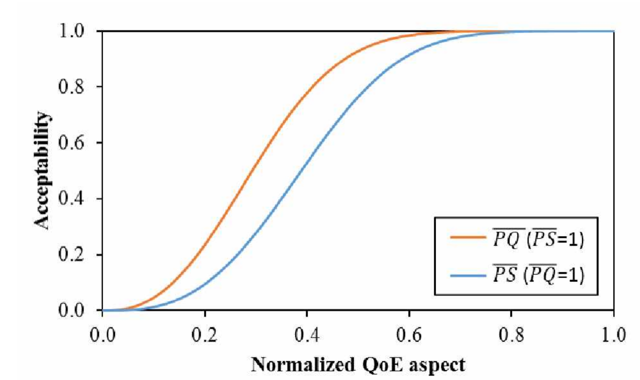

Fig. 1. Relationships between the QoE aspects and the acceptability using the second approach and training set S#1

B. Non-linear model

In the second approach, a non-linear regression is used to model the acceptability. Similar to [22], after examining different non-linear functions, we select the below function (i.e., Eq. (2)) since it achieves the highest correlation to the subjective acceptability values in the dataset. In particular, the acceptability is given by

$$Acc = \frac{1 - e^{-n \times \overline{PQ}^m}}{1 - e^{-n}} \times \frac{1 - e^{-k \times \overline{PS}^h}}{1 - e^{-k}}, \quad (2)$$

where $n, m, k,$ and h are parameters, $\overline{PQ}$, and $\overline{PS}$ are respectively the normalized perceptual quality and presence values.

Table V shows the performances of the two approaches in terms of Pearson Correlation Coefficient (PCC) and Root Mean Squared Error (RMSE) for the training and test sets. It can be seen that the second approach has significantly higher PCC and lower RMSE than the first approach for

TABLE V
PERFORMANCE OF THE TWO APPROACHES

Training set	First approach				Second approach			
	Training set		*Test set*		*Training set*		*Test set*	
	PCC	*RMSE*	*PCC*	*RMSE*	*PCC*	*RMSE*	*PCC*	*RMSE*
S#1 (Full dataset)	0.93	0.14	N/A	N/A	0.96	0.11	N/A	N/A
S#2 (Video #1)	0.94	0.13	0.94	0.17	0.97	0.09	0.95	0.15
S#3 (Video #2)	0.94	0.14	0.93	0.15	0.98	0.09	0.96	0.14
S#4 (Video #3)	0.94	0.13	0.93	0.15	0.97	0.09	0.96	0.12
S#5 (D#1)	0.96	0.12	0.93	0.18	0.98	0.07	0.95	0.16
S#6 (D#2)	0.94	0.14	0.95	0.19	0.97	0.10	0.96	0.16

TABLE VI
PARAMETERS OF THE SECOND APPROACH FOR TRAINING SET S#1

Parameter	n	m	k	h
Value	14.82	2.49	10.96	2.92

all the test sets. This indicates that the second approach better represents the relationship between the QoE aspects of the perceptual quality and presence and the subjective acceptability. In addition, the second approach always achieves PCC values higher than or equal to 0.95 for all the training and test sets. This means that this approach can be applied to predict the acceptability of 360 video over different videos and rendering device sets. This also reconfirms that the content motion and rendering device do not have significant impacts on the acceptability.

Table VI shows the parameters of the second approach when using training set S#1. Figure 1 shows the relationships between the two QoE aspects and the acceptability. Obviously, the decrease of the perceptual quality or the presence results in a significant reduction in the acceptability. Also, it can be seen that the acceptability is more dramatically decreased as the presence reduces. This again shows that the impact of the presence is more significant than that of the perceptual quality.

IV. CONCLUSIONS

In this paper, we have investigated for the first time the impacts of two QoE aspects of perceptual quality and presence to the overall QoE (i.e., the acceptability) of 360 video. Based on statistical analysis, it is found that the acceptability is strongly affected by both the perceptual quality and the presence. In addition, two approaches to model the acceptability have been investigated in this study. The result shows that the non-linear approach is better than the linear approach in predicting the acceptability of 360 video over different videos and rendering device sets. For future work, we intend to extend our approaches by additional taking into account the impacts of other QoE aspects such as cybersickness.

REFERENCES

[1] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "An optimal tile-based approach for viewport-adaptive 360-degree video streaming," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 29–42, Mar. 2019.

[2] D. V. Nguyen, H. T. T. Tran, and T. C. Thang, "A client-based adaptation framework for 360-degree video streaming," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 231–243, Jan. 2019.

[3] M. Hosseini and V. Swaminathan, "Adaptive 360 vr video streaming: Divide and conquer," in *IEEE International Symposium on Multimedia (ISM)*. San Jose, CA, USA: IEEE, Dec 2016, pp. 107–110.

[4] Y. Liu, S. Dey, F. Ulupinar, M. Luby, and Y. Mao, "Deriving and validating user experience model for DASH video streaming," *IEEE Transactions on Broadcasting*, vol. 61, no. 4, pp. 651–665, 2015.

[5] M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "Youtube qoe on mobile devices: Subjective analysis of classical vs. adaptive video streaming," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*. Dubrovnik, Croatia: IEEE, Aug. 2015, pp. 43–48.

[6] P. Spachos, W. Li, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, "Acceptability and quality of experience in over the top video," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. London, UK: IEEE, Jun. 2015, pp. 1693–1698.

[7] H. T. T. Tran, N. P. Ngoc, T. Hoßfeld, and T. C. Thang, "A Cumulative Quality Model for HTTP Adaptive Streaming," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. Sardinia, Italy: IEEE, May 2018.

[8] W. Song and D. W. Tjondronegoro, "Acceptability-based qoe models for mobile video," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 738–750, Apr. 2014.

[9] R. S. Kalawsky, "The Validity of Presence as a Reliable Human Performance Metric in Immersive Environments," in *3rd International Workshop on Presence*, Delft, Netherlands, 2000, pp. 1–16.

[10] A. Singla, S. Goring, A. Raake, B. Meixner, R. Koenen, and T. Buchholz, "Subjective quality evaluation of tile-based streaming for omnidirectional videos," in *10th ACM Multimedia Systems Conference (MMSys 2019)*. Amherst, MA, USA: ACM, Jun. 2019, pp. 232–242.

[11] H. Duan, G. Zhai, X. Yang, D. Li, and W. Zhu, "Ivqad 2017: An immersive video quality assessment database," in *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*. Poznan, Poland: IEEE, May 2017, pp. 1–5.

[12] H. T. T. Tran, N. P. Ngoc, C. M. Bui, M. H. Pham, and T. C. Thang, "An evaluation of quality metrics for 360 videos," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*. Milan, Italy: IEEE, July 2017, pp. 7–11.

[13] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang, and Z. Fei, "Subjective and objective quality assessment of panoramic videos in virtual reality environments," in *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. Hong Kong, China: IEEE, July 2017, pp. 163–168.

[14] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp, "An 'In the Wild' Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 4, pp. 1406–1414, 2016.

[15] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A Subjective Study on QoE of 360 Video for VR Communication," in *2017 IEEE International Workshop on Multimedia Signal Processing*, London-Luton, UK, oct 2017, pp. 1–6.

[16] H. G. Kim, H. Lim, S. Lee, and Y. M. Ro, "Vrsa net: Vr sickness assessment considering exceptional motion for 360 vr video," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1646–1660, April 2019.

[17] M. Huang, Q. Shen, Z. Ma, A. C. Bovik, P. Gupta, R. Zhou, and X. Cao, "Modeling the perceptual quality of immersive images rendered on head mounted displays: Resolution and compression," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6039–6050, Dec. 2018.

[18] H. T. T. Tran, T. H. Hoang, P. N. Minh, N. P. Ngoc, and T. C. Thang, "A perception-based quality metric for omnidirectional images," in *4th International Conference on Consumer Electronics Asia*, Bangkok, Thailand, Jun. 2019, p. pp.

[19] J. Li, R. Feng, Z. Liu, W. Sun, and Q. Li, "Modeling qoe of virtual reality video transmission over wireless networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018, pp. 1–7.

[20] Samsung official website, "Available online: http://www.samsung.com/ (accessed on 24 January 2017)."

[21] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences." Academic Press, 1977.

[22] Y. Ou, Y. Xue, and Y.Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *IEEE Transactions Image Processing*, vol. 23, no. 6, pp. 2473–2486, June 2014.