G.1070 Model Extension at Full HD Resolution for VP9/HEVC Codec

Debajyoti Pal and Vajirasak Vanijja
IP Communications Laboratory, School of Information Technology,
King Mongkut's University of Technology Thonburi,
Toongkru, Bangkok, Thailand.
debajyoti.pal@mail.kmutt.ac.th

***Abstract*—HD video streaming, which is gaining in popularity these days requires a large amount of bandwidth. This has resulted in the emergence of newer codecs like H.265/High Efficiency Video Coding (HEVC) and VP9 from Google. These codecs are supposed to provide an excellent video compression to quality ratio. ITU-T describes a standardised parametric model called the G.1070 Opinion Model, which estimates the Quality of Experience (QoE) of any multimedia content. The model estimates three parameters viz. the speech quality alone (S_q), the video quality alone (V_q) and the overall multimedia quality (M_q) of the input video. However, it needs to be trained separately for different codecs, video formats and certain other parameters, which can be obtained by carrying out suitable subjective tests. Our contribution in this paper is threefold. First, we carry out a subjective test according to the Recommendation P.910 to estimate the video quality for VP9 codec. Second, for the first time we use the results obtained from the subjective test to find out a set of coefficients that enables us to extend the G.1070 model for VP9 codec at Full HD resolution. Third, we provide an answer as to which is the better codec from H.265/HEVC and VP9 by evaluating their performance against scores obtained from different standard objective tests like the G.1070 model, Video Quality Metric (VQM) model and the Peak Signal to Noise Ratio (PSNR) model.**

***Index Terms*—H.265; MOS; Objective Test; QoE; Subjective Test; VP9.**

I. INTRODUCTION

Currently as per a report published in [1], video traffic constitutes more than 55% of the overall Internet traffic and is predicted to be continuously increasing. Predominantly, video streaming is being done on mobile devices having screen size ranging from 5 inches to 10 inches, either over mobile networks or over Wi-Fi [2]. This ever-increasing demand for bandwidth puts a serious strain on the existing network infrastructure and poses a serious challenge of end-user service quality to the Internet Service Providers (ISP's), especially the mobile ISP's. This situation has led to the emergence of newer codecs like H.265/HEVC from the Joint Collaborative Team on Video Coding (JCT-VC) and VP9 from Google [3] [4]. H.265/HEVC is able to save up to 50% of the bandwidth when compared against its immediate preceding generation codec i.e. H.264/AVC without any perceptual degradation of the video quality [5]. Similar observations are made in [6] where the comparison is done between VP8 and VP9 codecs. Models that can accurately predict the QoE of these newer codecs are extremely important in order to understand their advantages over older generation codecs [7] [8].

There is a twofold research challenge in this area [9] [10] [11]. First, we need to have proper QoE models in place that can accurately estimate the video quality of the current generation codecs when compared to the subjective results. Second, we need to find out the extent to which the current generation codecs can improve the perceived video quality. The first point is of particular importance as subjective tests are not always easy to carry out and they are very expensive too. Hence reliable/accurate mapping of objective data to subjective data is always desirable and vice versa. In this paper, we try to address the issues that have been mentioned above that will enable other researchers to reap the potential benefits of H.265/HEVC and VP9 codecs and use them efficiently for online video streaming purposes.

We carry out a subjective test by using the ITU-T Recommendation P.910 [12]. For objective measurements ITU-T Recommendation G.1070 is used [13]. The objective video quality V_q is based on the main assumption that for a particular experimental setup, it follows a Gaussian distribution in case of no packet loss and follows a decaying exponential pattern under the condition of packet loss. In particular, we used the results obtained from the subjective test to train the G.1070 model for VP9 codec by estimating the coefficients v_1 to v_{12}. By doing so, we are able to extend the ITU-T G.1070 model to support the VP9 codec. We also analyse the validity and robustness of the model from the data gathered. Next, we evaluate the performance of VP9 and H.265/HEVC codecs by benchmarking them across multiple standard objective methods namely, G.1070, PSNR and VQM. Thus, for a given set of conditions we are able to clearly answer as to which is the best performing codec.

II. LITERATURE REVIEW

Existing literature suggests that both H.265/HEVC and VP9 codec can save more that 50% of the bandwidth while maintaining either the same or better perceived video quality when compared to the older generation codecs like H.264/AVC or VP8 [14] [15]. However, in case of VP9 codec, there is some doubt as different researchers have different opinion about the effective bitrate savings offered by the codec [16] [17] [18]. In fact researchers in [16] even show that the VP9 encoder is inferior to the previous generation H.264 encoder in terms of bitrate savings for the same perceptual video quality. However, different subjective studies carried out in [19] and [20] show VP9 performs better. Thus, there is a need to do a comprehensive performance analysis of the codecs under consideration.

A lot more research has been done with H.265/HEVC codec both in terms of subjective and objective tests when compared to VP9 [21] [22]. Current studies have shown that H.265 can attain high compression efficiency, especially for Ultra High Definition (UHD) videos. However, not much work has been done that investigates the suitability of using H.265 in real time applications like video streaming. Also, the performance of H.265 or VP9 under low bandwidth conditions has not been accounted for. There has been one research done by [23] that evaluates the performance of H.265 encoded video content at 360p resolution and 200-400 Kbps bit rate. But, that is not a representative of the current generation standard where most of the video contents are being produced and transmitted at resolution of 720p and upwards. Thus, a comprehensive performance evaluation of the codecs needs to be done.

For subjective video quality assessments (sVQA), standard methods are provided by ITU-T [12] [24]. There are different types of test methods and experimental design viz. Absolute Category Rating (ACR), Absolute Category Rating with Hidden Reference (ACR-HR), Degradation Category Rating (DCR), Pair Comparison Method (PC), Single Stimulus Continuous Quality Evaluation (SSCQE) and Double Stimulus Continuous Quality Scale (DSCQS) to name a few. While ACR and SSCQE ask the viewers to rate only the impaired video stream, DCR and DSCQS present both the original as well as the degraded video sequence to the viewers and ask them to rate accordingly. Since ACR and SSCQE do not present the reference videos; hence, they can be carried out faster. However, they suffer from memory effect problem, which can be reduced by randomising the video orders [25].

Objective video quality assessments (oVQA) can also be classified into three broad categories viz. Media Layer Models, Packet Layer Models and Parametric Models [26]. Media layer models are based upon the analysis of the video contents. These metrics can further be classified into Full Reference (FR), Reduced Reference (RF) and No Reference (NR) models. Packet layer models are based upon the network information i.e. IP packets only. Parametric models, on the other hand combine some reduced set of parameters from the media layer model and the packet layer model. Peak signal to noise ratio (PSNR), Structural Similarity Index (SSIM), Video Quality Metric (VQM) and Motion-Based Video Integrity Evaluation Index (MOVIE) are all examples of media layer models. Historically, the PSNR metric has been the most widely used one, although it does not quite match the actual "perceived" quality by human observers [27 [28]. VQM gives a better prediction as it incorporates the characteristics found in any Human Visual System (HVS) [29].

For a good oVQA model, it must predict the video quality that is in agreement with the sVQA techniques. Current research shows that all the objective assessments that have been done for H.265/HEVC or VP9 video codec tend to underestimate the quality as compared to the subjective tests [30]. This means that we have to fine tune the models so as to improve their prediction accuracy for these newer codecs. Most of the QoE models assess the video quality based upon the PSNR, SSIM and VQM video metrics [28] [31] [32] [33]. All these models take into account the effect of video quality distortion in the received video as compared to the original one. However, they do not consider anything about the video delivery system i.e. the underlying network. Hence, the prediction accuracy for these models will be less accurate for online video streaming purpose, which is the main essence of our research. Hence, we use the parametric model ITU-T G.1070 for our purpose.

G.1070 model provides three main outputs. It provides us with a speech quality index (S_q), video quality index alone (V_q) and an overall multimedia quality index (M_q) that takes into account both S_q and V_q with any introduced audiovisual delay. Speech quality index S_q is based upon the Simplified E-Model [34]. Video quality index V_q depends upon the application and network layer parameters like bit rate (BR), frame rate (FR) and packet loss rate (PLR) of the encoded video. For this research, we scope our work only to V_q. For a given set of condition (BR, FR, codec type, video format and video display size) V_q follows a Gaussian distribution, while for the PLR factor, it follows a decaying exponential distribution [13]. The overall V_q for these two separate cases is expressed in the form of twelve coefficients v_1 to v_{12}. The value of these coefficients depends upon the type of codec, video format and video display size. To date, the G.1070 model has been trained for MPEG2, MPEG4-Part2, H.264/AVC and H.265/HEVC video codecs with resolutions ranging from CIF to Full HD [35] [36] [37] [38].

Literature review suggests that there are a number of gaps in the current research. In this paper, we try to fill up those gaps. First, we carry out a subjective test as per the ITU-T P.910 Recommendation for the VP9 codec. Second, we use the result obtained from the subjective test to extend the G.1070 model by extracting the optimised set of coefficients v_1 to v_{12} at full HD resolution. Third, we analyse the validity and robustness of the G.1070 model for the VP9 codec from the data that we have gathered. Fourth, we take some other commonly used objective models like PSNR and VQM and compare their performance with the G.1070 model for current generation H.265/HEVC and VP9 codecs based upon the subjective data that we have. Thus, we are able to suggest the overall suitability of a particular model towards online video streaming applications.

III. RESEARCH METHODOLOGY/EXPERIMENT DESIGN

In the first stage, we carried out the subjective test as per the ITU-T P.910 standard and used those results for training the ITU-T G.1070 model. The overall workflow for this phase is shown in Figure 1. Once the G.1070 model has been extended to support the VP9 codec, in the second stage we exhaustively evaluated the performance of the current generation codecs H.265/HEVC and VP9 across multiple OVQA models. This is shown in Figure 2.

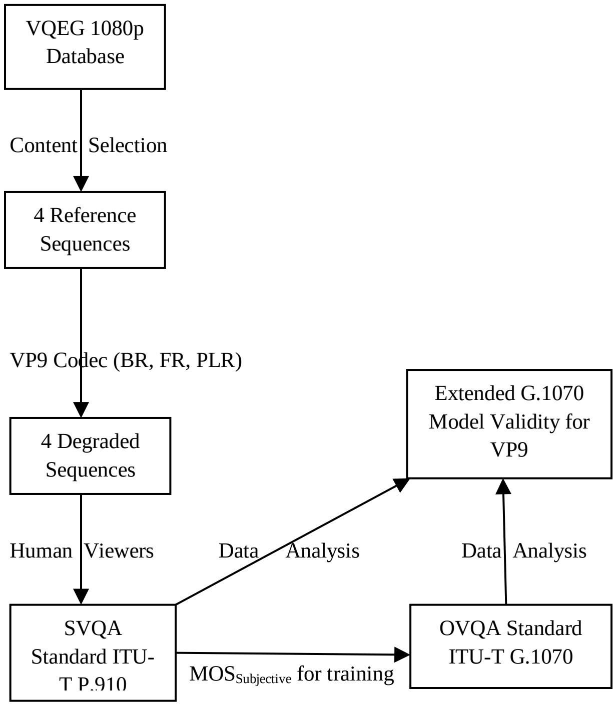

Figure 1: Research workflow for first phase

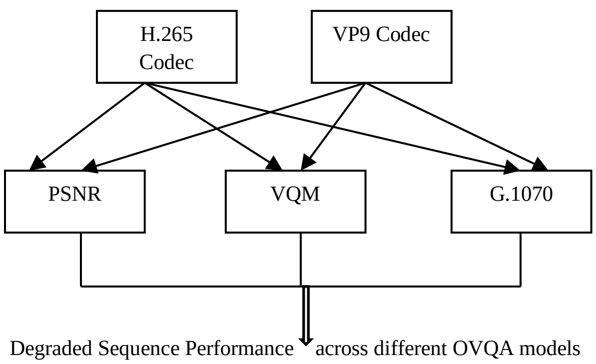

Figure 2: Research workflow for second phase

The availability of publicly available video dataset for research at Full HD and upward resolution is really very limited. While carrying out our subjective tests, we used the publicly available SVT High Definition Multi Format Test Set maintained by the Video Quality Experts Group (VQEG) [39]. We selected four different reference videos, each having different levels of spatial (SI) and temporal information (TI). The relevant details of the video clips are presented in Table 1. The SI and TI values were calculated as per the recommendation provided in [12] and included in Table 1. From the table, it can be observed that the SI and TI values vary over a very wide range depending on the selected video content. As the perceived video quality depends on the video content, which has been established by researchers in [40]; hence, we selected videos having a wide variety of content level to cover the entire gamut possible.

Table 1
Details of selected video sequences

Sequence No	Name	Resolution	Frame Rate	SI,TI Values
1	CrowdRun	1920 × 1080	30 fps	2,12
2	DucksTakeOff	1920 × 1080	30 fps	7,6
3	OldTown	1920 × 1080	30 fps	10,24
4	ParkJoy	1920 × 1080	30 fps	9,14

Each of the selected video sequences is of length 10s. All the four reference videos that have been selected are presented in the raw YUV 4:2:0 formats.

VP9 compression was performed as per the implementation provided by the latest version of the ffmpeg encoder (ver. 3.1.3). The encoding quality preset was set to “best” and we used the 2-pass encoding option to give the maximum quality although the encoding process was very slow. Standard values for the initial, optimal and maximum buffer levels were used as per the recommendation. The constrained quality (CQ) level was kept the same as that of the quantisation parameter (QP) value for the best quality. We disabled the adaptive quality mode as this is a VP9 only exclusive feature and not available for other codecs like H.265/HEVC. A summary of the encoder configuration is given in Table 2.

Table 2
VP9 encoder configuration

Parameter	Details
Encoder used	ffmpeg
Encoder version	3.1.3
Encoding quality	Best
No of passes	2
Bit-rate control mode	Variable bit rate (VBR defined by target bitrate)
Constrained quality (CQ) level	Kept Same as Quantization Parameter QP
Initial, optimal and maximum buffer level	4000 ms, 5000 ms, 6000ms
GOP size	Auto
GOP length (Intra Period)	320
Adaptive quality mode (Aq)	Off (Set to 0)
Bit depth	8

For each of the sequences under the condition of no packet loss, we used five different bit rate and frame rate combination totalling to 100 different test conditions. However, in the case of packet loss in order to limit the number of test conditions, we selected five specific combinations of bit rate and frame rate and combined them with four different packet loss levels to obtain a total of 80 test conditions. Thus, a total of 180 test conditions per user were prepared.

The subjective experiment was carried out as per the ITU-T Recommendation P.910 in a controlled laboratory environment. All the video samples were presented before the users on a Samsung Galaxy Note 5 having a screen resolution of 2K (1440 × 2560) pixels, 64 GB of internal storage, 4GB of RAM and running the latest version of Android Marshmallow (6.0.1). We selected this device as it has inherent support for displaying the latest generation codecs H.265/HEVC as well as VP9. All the videos were preloaded into the mobile and flight mode was turned on while carrying out the experiment. The detailed experimental setup is provided in Table 3.

Table 3
Experimental details

Parameter	Details
Video codec	VP9
Encoder version	ffmpeg version 3.1.3
Video format	Full HD progressive (1080p)
Video bit rate (kbps)	500, 1000, 2000, 4000, 8000
Video frame rate (fps)	5, 15, 25, 30, 60
Packet loss rate (%)	0.5, 1, 3, 5
Packet loss pattern	Random
Video sequences	CrowdRun, DucksTakeOff, OldTown, ParkJoy
Display device	Samsung Galaxy Note 5
Viewing distance and angle (measured from screen)	80 cm and 30º

We conducted the subjective test on 24 subjects. All the subjects were between 18-35 years of age, balanced in gender, non-experts in the field of video quality assessment and did not have any visual impairment like colour blindness or myopia. Before selecting the subjects, we asked them to describe the colours shown in a given image and gave them some training videos for quality comparison so as to judge their suitability for the experiment. No one was disqualified during this process. We also conducted a training session with a demo video so as to familiarise the participants with the actual test conditions. Since the subjective assessment is a very tedious and high-concentration task, we divided the entire session into two parts of 15 minutes each, keeping aside the demo session. During the assessment, the participants were left alone in order to minimise the unwanted effects of being supervised [41]. We adopted the 5-point ACR method as outlined in [12]. The subjects were provided with scoring sheets where they would input their assessment after watching a particular video. After the test, all the offline scores were manually entered into a computer for the purpose of data analysis. The scores were cross-checked by two different people so as to avoid any data entry error.

After finishing the subjective test and the corresponding data analysis, we used these results to train the objective G.1070 model. Accordingly, we propose the values of twelve set of coefficients (v_1 to v_{12}) that enable us to use the G.1070 model for VP9 codec. For the other objective methods PSNR and VQM, we used the standard Video Quality Measurement Tool (VQMT) maintained by the Multimedia Signal Processing Group (MMSPG) [42]. We used VQMT as it is an open source tool implemented in OpenCV (C++) and shows better performance than Matlab in terms of runtime. The oVQA scores were recorded in CSV files for further analysis.

We used the curve fitting toolbox offered by Matlab (version R2015b) for the purpose of curve fitting and regression analysis to map the subjective test results to extend the G.1070 model. Data analysis was carried out in IBM SPSS Statistics Desktop version 22. Our data analysis includes any outlier detection and their consequent removal that can happen during the sVQA phase, checking the validity and robustness of the extended G.1070 model for the VP9 codec and comparing the sVQA and oVQA scores obtained from the different models.

IV. sVQA Result Analysis

We recorded a total of 4320 subjective MOS scores (180 video sequences × 24 subjects). To begin with, we performed the process of outlier detection in order to remove any data inconsistency. If we represent the score obtained by any subject as S_{ij}, where i denotes a particular test sequence and j denotes the score obtained for that particular test sequence, then S_{ij} will be considered as an outlier if $S_{ij} > q_3 + 1.5(q_3 - q_1)$ OR $S_{ij} < q_1 - 1.5(q_3 - q_1)$, q_1 and q_3 being the 25th percentile and 75th percentile respectively of the score distribution [43]. This range is approximately equal to 99.3% of the normally distributed data. A subject can be considered to be an outlier and all his/her entries removed, if more than 20% of his/her scores are outliers [43]. In our experiment, following the above rules, we did not find any outlier. The mean opinion score (MOS) has been calculated as:

$$\mathrm{MOSi} = \sum_{j=1}^{n} \mathrm{Sij}/\mathrm{N} \quad (1)$$

where, N = number of valid subjects and Sij denotes the score by subject j for the test condition i.

Figure 3 show the MOS scores obtained from the subjective test for H.265 and VP9 codec at different bit rates. The subjective data for the H.265 codec has been taken from our previous work in [38]. The result shows that the VP9 codec performs marginally better than that H.265 codec at lower bit rates up to 1000 kbps. At higher bit rates above 1000 kbps, the performances of both the codecs are comparable.

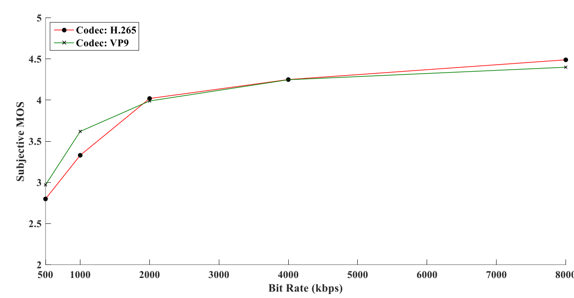

Figure 3: MOS from subjective test for H.265 and VP9

We carried out a t-test to investigate the statistical superiority of a particular codec, the results of which are shown in Table 4.

Table 4
Result of t-test for H.265 and VP9 codec

Bit Rate (kbps)	t-test Result
500	$t(7) = 6.254$, $p = 0.015$
1000	$t(7) = 4.258$, $p < 0.001$
2000	$t(7) = 2.821$, $p = 0.037$
4000	$t(7) = 1.964$, $p = 0.081$
8000	$t(7) = 2.671$, $p = 0.078$

Based on the t-test result, we can conclude that VP9 codec is statistically superior to the H.265 codec ($p = 0.015$, $p < 0.001$ and $p = 0.037$) at bit rates 500 kbps, 1000 kbps and 2000 kbps respectively. For, higher bit rates the superiority of VP9 ceases to exist. Also, for the bit rate 2000 kbps, we observe an anomaly where the observed quality does not match the statistical calculated value. Next, we use the MOS subjective scores to train the G.1070 model for VP9 codec.

V. oVQA Result Analysis

The video quality estimation function V_q in the G.1070 model is described by Equations (2) to (8).

$$V_q = 1 + I_{Coding} \exp\left(-\frac{P_{Pl}}{D_{Ppl}}\right) \quad (2)$$

where, I_{Coding} represents the video quality affected by the coding distortion only, P_{Pl} represents the % of packet loss and D_{Ppl} represents the degree of video quality robustness due to packet loss. If P_{Pl} is 0, equation (2) reduces to:

$$V_q = 1 + I_{Coding} \quad (3)$$

where, I_{Coding} is expressed as:

$$I_{Coding} = I_{Ofr} \exp\left\{-\frac{(\ln(Fr)-\ln(O_{Fr}))^2}{2D_{Fr}{}^2}\right\} \quad (4)$$

where, I_{Ofr} represents the maximum video quality for a particular bit rate B_r, F_r represents the frame rate, O_{Fr} represents the optimal frame rate that maximises the video quality for a particular bit rate B_r and D_{Fr} represents the degree of video quality robustness due to the frame rate F_r. O_{Fr}, I_{Ofr} and D_{Fr} are expressed as:

$$O_{Fr} = v_1 + v_2 \times B_r, \qquad 1 \leq O_{Fr} \leq 30 \quad (5)$$

$$I_{Ofr} = v_3 - \frac{v3}{1+\left(\frac{Br}{v_4}\right)^{v_5}} \qquad 0 \leq I_{Ofr} \leq 4 \quad (6)$$

$$D_{Fr} = v_6 + v_7 \times B_r \quad (7)$$

D_{Ppl} is expressed as:

$$D_{Ppl} = v_{10} + v_{11}\exp\left(-\frac{Fr}{v_8}\right) + v_{12}\exp\left(-\frac{Br}{v_9}\right), 0 < D_{Ppl} \quad (8)$$

Our aim is to estimate the coefficients v_1 to v_{12} as described by the above equations from the subjective MOS that we have. First, we estimated the coefficients v_1 to v_7 under the case of no packet loss. To do this, we performed the first curve fitting to the subjective data that we have from the experiment carried out; hence, we obtained the values of I_{Ofr}, O_{Fr} and D_{Fr} for every bit rate B_r. This is shown in Table 5. Applying these values, we performed the second curve fitting to equations (5) to (7) and obtained the value of the coefficients v_1 to v_7.

Figure 4 shows the plot of O_{Fr} vs. the bit rate B_r. There is no threshold value of O_{Fr} beyond which it saturates; in fact it increases in a linear fashion with increasing bit rates. However, the upper limit of O_{Fr} exceeds 30; so we consider the boundary condition of 0-30 fps invalid for our case of VP9 codec at full HD resolution and revise it to 0-60 fps for all further calculations.

Table 5
O_{Fr}, I_{Ofr} and D_{Fr} vales for different bit rates

Bit rate (kbps)	O_{Fr}	I_{Ofr}	D_{Fr}
500	29.57	2.11	1.63
1000	30.80	2.45	1.83
2000	31.24	2.85	2.22
4000	32.14	3.31	3.00
8000	33.93	3.84	4.56

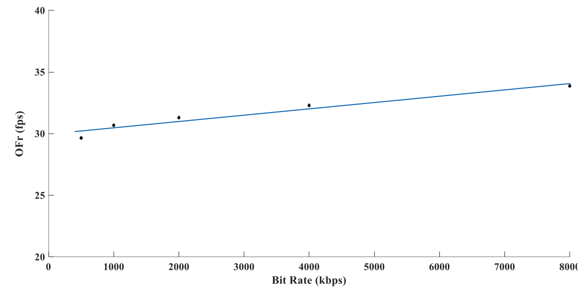

Figure 4: Optimal frame rate (O_{Fr}) vs. Bit rate (B_r)

Figure 5 shows the plot of I_{Ofr} vs. the bit rate. We obtained a reasonable fit for all the bit rates. Also, the assumption of $0 \leq I_{Ofr} \leq 4$ is found to be true.

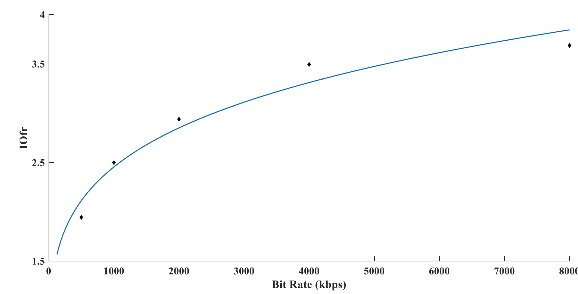

Figure 5: Maximum video quality (I_{Ofr}) vs. Bit rate (B_r)

Figure 6 shows the plot of D_{Fr} vs. the bit rate B_r. We observed that D_{Fr} increases linearly with an increase in bit rate B_r which is in agreement with equation 7.

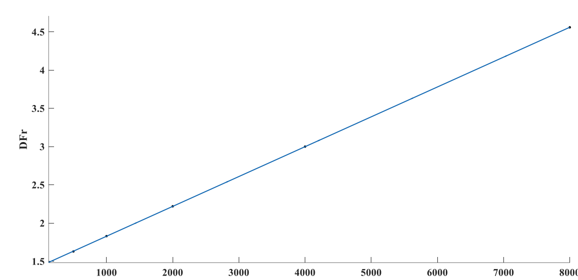

Figure 6: Degree of video quality robustness (DFr) vs. Bit rate (B_r)

Figure 7 shows the variation of MOS (Objective) with frame rate F_r. In case of video quality distortion due to the coding artifact only; V_q follows a Gaussian distribution as evident from equations 3 and 4. Figure 7 confirms this general trend. For every bit rate B_r, we have a maximum/optimal frame rate (O_{Fr}) that corresponds to the maximum video quality, thereafter the value of MOS decreasing.

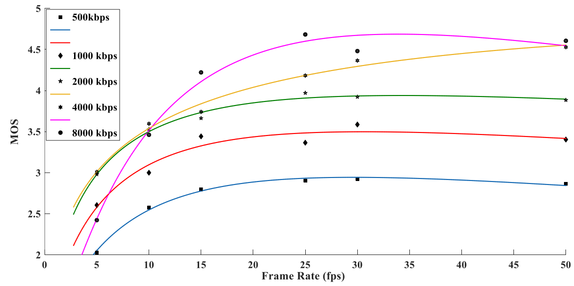

Figure 7: Variation of MOS for different bit rates

For the condition of no packet loss, we can therefore confirm from the above results that our set of data for the VP9 codec at full HD resolution fits well enough with only one exception viz. the threshold value of the optimal frame rate (O_{Fr}) being increased from 30 fps to 60 fps for our experiment.

Next, we find out the coefficients v_8 to v_{12} by considering the case of packet loss. In order to save time and limit the total number of testing conditions; we selected some specific combinations of bit rate and frame rate along with the packet loss rate, which are shown in Table 6. Video quality V_q is represented by equation 2 in case of packet loss. Since, we have already calculated the I_{Coding} values for every bit rate and frame rate combinations in the first part of our experiment (also shown in table 6 for easy reference), hence we do a curve fitting to equation 2 for finding out the D_{Ppl} values for our selected combination. We used these D_{Ppl} values in equations 9 and 10 to obtain the coefficients v_8 and v_9.

Table 6
I_{Coding} value for Bit rate (B_r), Frame rate (F_r) and Packet loss rate (P_{Pl}) combination

Bit rate/Frame rate combination	Packet loss rate (%)	I_{Coding}
1000 kbps, 30 fps	0.5, 1, 3, 5	2.45
2000 kbps, 30 fps	0.5, 1, 3, 5	2.85
8000 kbps, 15 fps	0.5, 1, 3, 5	3.78
8000 kbps, 25 fps	0.5, 1, 3, 5	3.83
8000 kbps, 30 fps	0.5, 1, 3, 5	3.84

$$D_{P_{pl}} = a + b \exp\left(-\frac{Fr}{v_8}\right) \tag{9}$$

$$D_{P_{pl}} = c + d \exp\left(-\frac{Br}{v_9}\right) \tag{10}$$

After v_8 and v_9 are known, we again carried out a curve fitting to equation 8 to obtain the values of the remaining coefficients v_{10} to v_{12}. The values of all the coefficients v_1 to v_{12} are shown in Table 7 that enables us to extend the G.1070 model to include the VP9 codec at Full HD resolution.

In case of video quality distortion due to packet loss; V_q follows an exponentially decaying pattern as evident from equation 2. Figure 8 that shows the MOS (Objective) vs. the packet loss rate for our selected combination of bit rate/frame rate confirms this trend. However, from the figure, it is evident that the model fails to give a correct estimation of the video quality for higher values of packet loss rate.

Table 7
Coefficients for the VP9 codec at Full HD resolution for G.1070 model

Coefficients	Value
v_1	45.44
v_2	3.25×10^{-3}
v_3	0.5497
v_4	16.68
v_5	0.2229
v_6	2.1
v_7	-2.71×10^{-5}
v_8	0.1067
v_9	0.2599
v_{10}	4.388
v_{11}	0.07597
v_{12}	0.2399

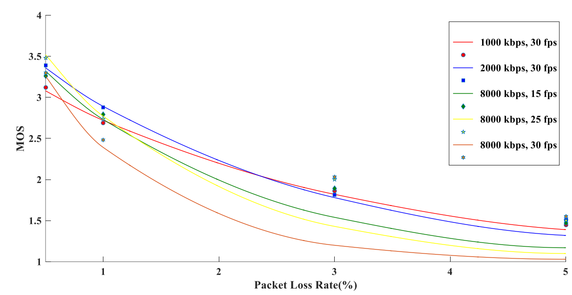

Figure 8: Variation of MOS with different packet loss rates

Next, we calculated the G.1070 model accuracy for the VP9 codec. Figure 9 shows the prediction accuracy of the G.1070 model for VP9 codec under both conditions. We calculated the overall model accuracy, accuracy due to the coding artifact only and accuracy due to the packet loss. R^2 values of 0.640, 0.7055 and 0.4668 were obtained for the different conditions respectively. Under similar conditions, the values of Pearson Correlation Coefficient obtained are 0.762, 0.805 and 0.556 respectively. From the results, we can see that the accuracy of the G.1070 model under packet loss is quite poor. The accuracy analysis was done using the same set of data that was used for training the G.1070 model.

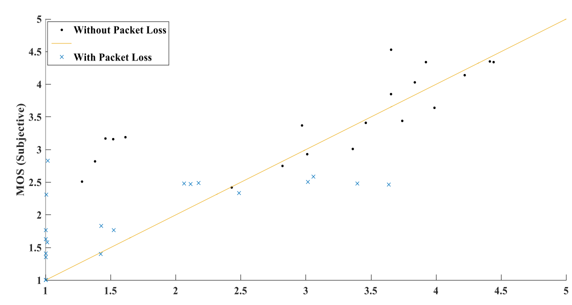

Figure 9: Prediction accuracy of the G.1070 model for VP9 codec

VI. Performance of HEVC and VP9 Codecs Across Different oVQA Models

In this section, we examine the performance of the H.265/HEVC and VP9 codecs across three popular objective models; name PSNR, VQM and the G.1070 Opinion Model.

The PSNR model assesses the video quality on a scale of 0 dB to 100 dB, higher meaning better quality. In case of VQM, it gives a rating ranging from 0 to 5, with 0, indicating the best quality, while 5 the worst quality. For the G.1070 model, it gives a MOS score ranging from 0 to 5, where 0 indicates the lowest and 5 indicates the highest quality. Figure 10, 11 and 12 show the PSNR score, VQM score and the MOS score as a function of the bit rate respectively in the absence of packet loss.

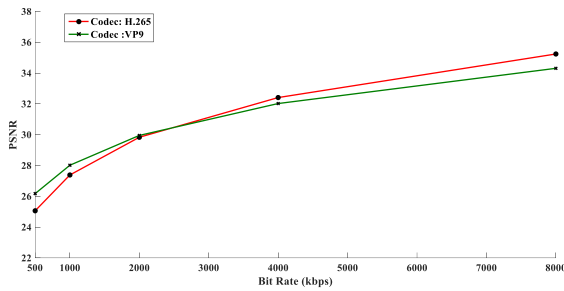

Figure 10: Variation of PSNR score with bit rate

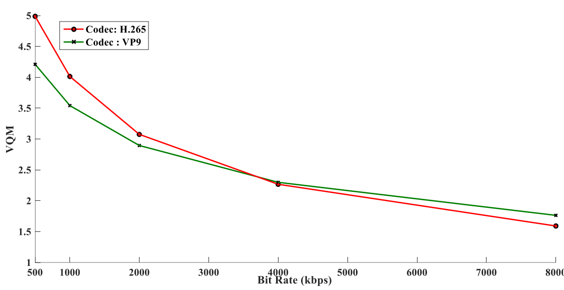

Figure 11: Variation of VQM score with bit rate

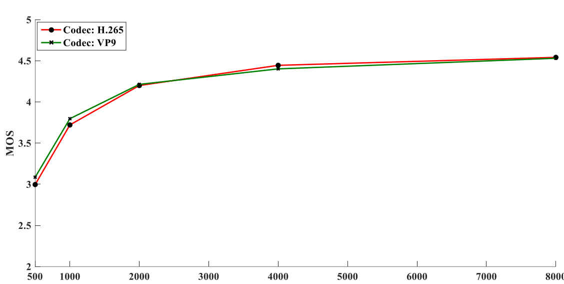

Figure 12: Variation of MOS score with bit rate

The PSNR, VQM and MOS scores are highly consistent with each other and all of them show a similar trend when compared against the MOS (Subjective) scores depicted in Figure 3. We observed that up to 2000 kbps, VP9 codec has a clear advantage over the H.265 codec. However, as the bit rate increases beyond 2000 kbps, the performance of H.265 and VP9 codec becomes comparable. This observation is in agreement with the results from the subjective test.

Next, Figure 13 to 20 show the subjective and objective video quality under the condition of packet loss.

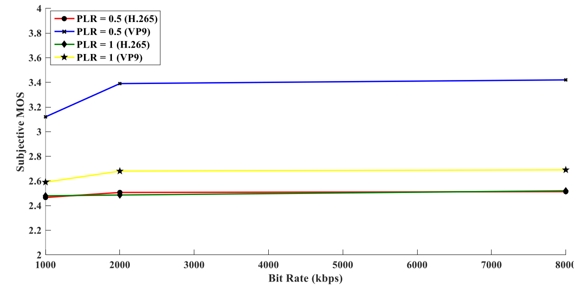

Figure 13: Variation of Subjective MOS score with bit rate under packet loss (0.5% and 1%)

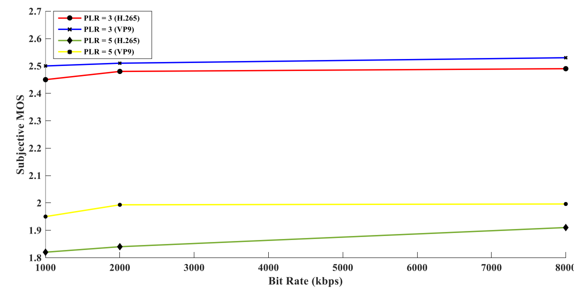

Figure 14: Variation of Subjective MOS score with bit rate under packet loss (3% and 5%)

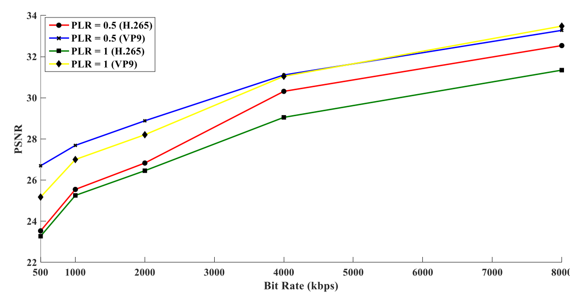

Figure 15: Variation of PSNR score with bit rate under packet loss (0.5% and 1%)

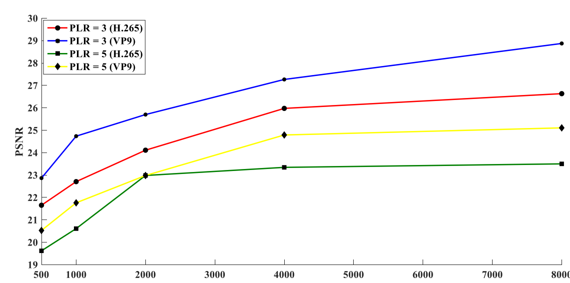

Figure 16: Variation of PSNR score with bit rate under packet loss (3% and 5%)

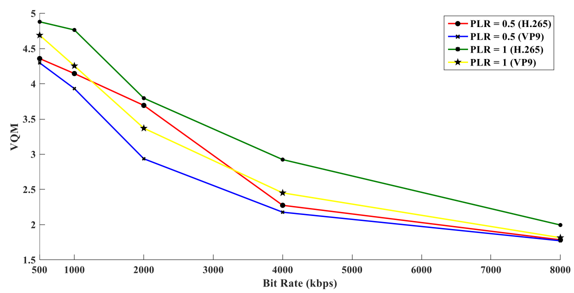

Figure 17: Variation of VQM score with bit rate under packet loss (0.5% and 1%)

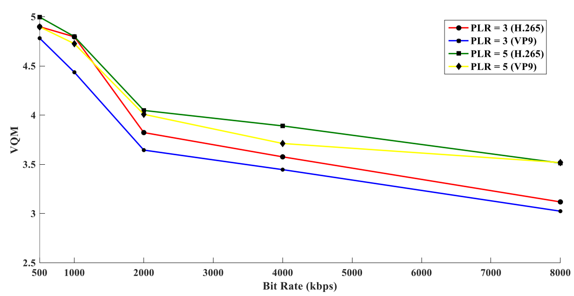

Figure 18: Variation of VQM score with bit rate under packet loss (3% and 5%)

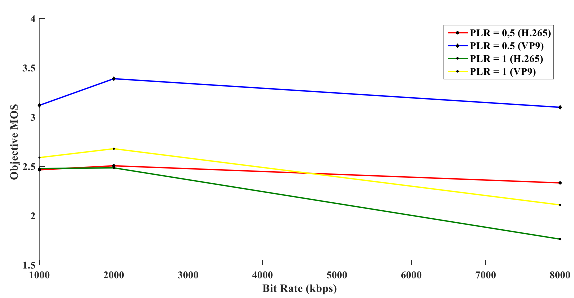

Figure 19: Variation of Objective MOS score with bit rate under packet loss (0.5% and 1%)

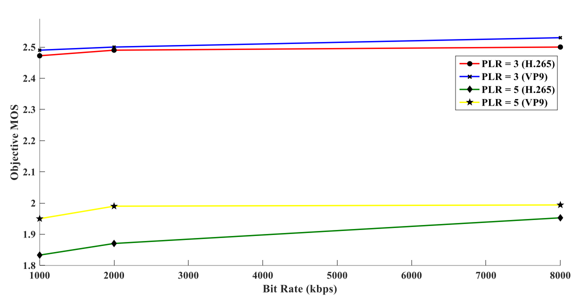

Figure 20: Variation of Objective MOS score with bit rate under packet loss (3% and 5%)

The figures show that from both the subjective as well as the objective test results obtained from the three models, for packet loss up to 1%, VP9 is the better codec across all bit rates. For higher values of packet loss, there is no noticeable difference in the video quality between the two codecs, both of them being quite poor. This result is somewhat different from the condition of no packet loss, where VP9 had a significant advantage only for bit rates up to 2000 kbps. Hence, we can conclude that the VP9 codec is more resistant to lower value of packet losses, which should give it a preference to be used for online video streaming applications over H.265 codec.

VII. CONCLUSION

In this paper, we have extended the G.1070 Opinion model for the current generation VP9 codec at Full HD resolution, by extracting an optimised set of parameters v_1 to v_{12}. For this purpose, we conducted a subjective experiment and used those results in our calculation. The accuracy of the extended G.1070 model is also fairly good under the condition of no packet loss. However, for the condition of packet loss, it could estimate only within a narrow range of 0% to 2%.

We also chose two other popular objective video metrics PSNR and VQM, and studied the performance of the videos coded with VP9 and H.265 across all the three models. When we compared these results with those obtained from the respective subjective tests, the results showed consistency and a common general trend. Under the condition of no packet loss, VP9 was seen to be performing better than H.265 in the lower bit rate region of up to 2000 kbps. Higher bit rates yielded the same result for both the codecs. However, quite surprisingly for low values of packet loss (less than 2%), VP9 was seen to be performing better across all values of bit rate. But, with an increase in packet loss, the performance of both codecs was found to be very poor and was not distinguishable from one another.

In this paper, we did not investigate the effect of other common parameters like the video resolution, nature of video content, effect of the display size, etc. that can affect the video quality. We plan to include these parameters along with any other(s) for our future work and predict our own video quality estimation model. Also, the effect that the video content can have for different codecs on the overall video quality needs to be investigated in detail, which we propose in our future work.

REFERENCES

[1] Cisco Global Mobile Data Traffic Forecast Update Report, 2014-2019, Cisco white paper (2016)

[2] Ooyala, Q1 2013 Video Index- TV is no longer a single screen in your living room, 2013

[3] Sullivan G.J, Ohm J.R, Han W.J and Wiegand T, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.12, pp. 1649-1668, December 2012, doi: 10.1109/TCSVT.2012.2221191

[4] D. Mukherjee et al., "The latest open-source video codec VP9- An Overview and Preliminary Results," *Picture Coding Symposium* (PCS), pp.390-393, San Jose, CA, 2013 doi: 10.1109/PCS.2013.6737765

[5] B. Bross et al., "High Efficiency Video Coding (HEVC) Text Specification Draft 10," *Document JCTVC-L1003*, JCT-VC, Geneva, Switzerland, January 2013

[6] Liang Hu, Yan Li, Tengfei Li, Hongtu Li and Jianfeng Chu, "The Efficiency Improved Scheme For Secure Access Control of Digital Video Distribution," *Multimedia Tools and Applications*, vol.75, no.20, pp.12645-12662, October 2016, doi: 10.1007/s11042-015-2450-7

[7] Kjell Brunnstrom et al., "Qualinet White Paper on Definitions of Quality of Experience," *Output from the 5th Qualinet Meeting*, 2013

[8] Menkovski V, Exarchakos G and Liotta A, "The Value of Relative Quality in Video Delivery," *Journal of Mobile Multimedia*, vol.7, no.3, pp. 151-162, September 2011

[9] V. Menkovski and A. Liotta, "Intelligent Control for Adaptive Video Streaming," *2013 IEEE International Conference on Consumer Electronics (ICCE)*, pp.127-128, Las Vegas, 2013 doi: 10.1109/ICCE.2013.6486825

[10] Song W, Tjondronegoro D and Himawan I, "Acceptability-based QoE Management for User-centric Mobile Video Delivery: A Field Study Evaluation," *22nd ACM International Conference on Multimedia (MM'14),* pp.267-276, New York, 2014 doi: 10.1145/2647868.2654923

[11] Agboma F and Liotta A, "User Centric Assessment of Mobile Contents Delivery'" *4th International Conference on Advances in Mobile Computing and Multimedia*, pp.104-107, Indonesia, 4 December, 2014

[12] Subjective Video Quality Assessment Methods for Multimedia Applications, ITU-T P.910, April 2008

[13] Opinion Model for Video-telephony Applications, ITU-T G.1070, July 2012

[14] Sullivan G, Topiwala PN and Luthra A, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the fidelity Range Extensions," *49th Annual Meeting on Optical Science and Technology*,pp.454-474, no.0001

[15] Ohm JR, Sullivan GJ, Schwarz H, Tan TK and Weigand T, "Comparison of the Coding Efficiency of Video Coding Standards-Including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.12, pp.1669-1684, December 2012 doi: 10.1109/TCSVT.2012.2221192

[16] Grois D, Marpe D, Muyaloff A, Itzhaky B and Hadar O, "Performance Comparison of H.265/MPEG-HEVC, VP9 and H.264/MPEG-AVC Encoders," *Picture Coding Symposium (PCS)*, pp.394-397, San Jose, 2013 doi: 10.1109/PCS.2013.6737766

[17] Mukherjee D et al., "A Technical Overview of VP9- The Latest Open-source Video Codec," *SMPTE Motion Imaging Journal*, vol.124, no.1, pp.44-54, January 2015 doi: 10.5594/ji8499

[18] Kufa J and Kratochvil T, "Comparison of H.265 and VP9 Coding Efficiency for Full HDTV and Ultra-HDTV Applications," *25th International Conference on Radioelektronika*, pp.168-171, 2015 doi: 10.1109/RADIOELEK.2015.7128999

[19] Uhrina M, Bienik J and Vaculik M, "Procedure for Mapping of Objective Video Quality Metrics to Subjective MOS Scale of VP9 Compression Standard for Full HD Resolution," *International Conference on Information and Digital Technologies (IDT)*, pp.288-294, Rzeszow 2016 doi: 10.1109/DT.2016.7557188

[20] Rerabek M et al. , "Quality Evaluation of HEVC and VP9 Video Compression in Real-Time Applications," *7th International Workshop on Quality of Multimedia Experience (QoMEX)*, pp.1-6, Nestoras, 2015 doi: 10.1109/QoMEX.2015.7148088

[21] Baroncini V, Ohm J.R and Sullivan G.J, "Report on Preliminary Subjective Testing of HEVC Compression Capability*," JCT-VC of ITU-T SG16 WP3* and *ISO/IEC JTC1/SC29/WG12*, 2012

[22] Bae SH, Kim J, Cho S and Choi J.S, "Assessments of Subjective Video Quality on HEVC-Encoded 4K-UHD Video for Beyond-HDTV Broadcasting Services," *IEEE Transactions on Broadcasting*, vol.59, no.2, pp.209-222, June 2013 doi: 10.1109/TBC.2013.2247171

[23] Garcia R and Kalva H, "Subjective Evaluation of HEVC and AVC/H.264 in Mobile Environments," *IEEE Transactions on Consumer Electronics*, vol.60, no.1, pp.116-123, February 2014 doi: 10.1109/TCE.2014.6780933

[24] Methodology for The Subjective Assessment of the Quality of Television Pictures, ITU-R BT.500-13, January 2012

[25] Pinson M and Wolf S, "Comparing Subjective Video Quality Testing Methodologies," *Proceedings of SPIE*, pp.573-582, 2003

[26] Joscowicz J, Sotelo R and Ardao J.C.L, "Towards a General Parametric Model For Perceptual Video Quality Estimation," *IEEE Transactions on Broadcasting*, vol.59, no.4, pp.569-579, December 2013 doi: 10.1109/TBC.2013.227795

[27] Quan Huynh-Thu and Ghanbari M, "The Accuracy of PSNR in Predicting Video Quality for Different Video Scenes and Frame Rates," *Telecommunication Systems Journal*, vol.49, no.1, pp.35-48, January 2012 doi: 10.1007/s11235-010-9351-x

[28] Song W and Tjondronegoro D.W, "Acceptability-Based QoE Models for Mobile Video," *IEEE Transactions on Multimedia*, vol.16, no.3, pp.738-750, April 2014 doi: 10.1109/TMM.2014.2298217

[29] Catelliar A and Pinson M, "Characterization of the HEVC Coding Efficiency Advance Using 20 Scenes, ITU-T Rec. P.913 Compliant Subjective Methods, VQM and PSNR," *IEEE International Symposium on Multimedia (ISM)*, pp.282-288, Miami, 2015 doi: 10.1109/ISM.2015.38

[30] Zeng Kai et al., "From H.264 to HEVC: Coding Gain Predicted by Objective Video Quality Assessment Methods," *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, February 2013

[31] Whore A and Ziou D, "Image Quality Metrics: PSNR vs. SSIM," *20th International Conference on Pattern Recognition (ICPR)*, pp.2366-2369, 2010 doi: 10.1109/ICPR.2010.579

[32] Anegekuh L and Lingfen S, "Encoding and Video Content Based HEVC Video Quality Prediction," *Multimedia Tools and Applications*, vol.74, no.11, pp.3715-3758, June 2015 doi: 10.1007/s11042-013-1795-z

[33] Nightingale J and Wang Q, "Deriving Video Content Type from HEVC Bit stream Semantics," *International Conference on Real Time Image and Video Processing*, May 2014 doi: 10.1117/12.2051757

[34] The E-Model, A Computational Model for Use in Transmission Planning, ITU-T G.107, March 2005

[35] Belmudez B and Moller S, "Extension of the G.1070 Video Quality Function for the MPEG2 Video Codec," *2nd International Workshop on Quality of Multimedia Experience (QoMEX)*, pp.7-10, Trondheim , 2010, doi 10.1109/QOMEX.2010.5518225

[36] Yamagishi K and Hayashi T, "QRP08-1: Opinion Model for Estimating Video Quality of Videophone Services," *IEEE Globecom*, pp.1-5, San Francisco, 2006 doi 10.1109/GLOCOM.2006.458

[37] Bilel Jamaossi, "Verification of Video Quality Estimation Model for Videophone Services," NTT, Japan, Rep.COM12-C34-E, Jan.2008

[38] Pal D, Triyason T and Vanijja V, "Extending the ITU-T G.1070 Opinion Model to Support Current Generation H.265/HEVC Video Codec," *International Conference on Computational Science and Its Applications (ICCSA)*, pp.106-116, Beijing, July, 2016 doi: 10.1007/978-3-319-42108-7_8

[39] VQEG Standard Database, maintained at http://www.its.bldrdoc.gov/vqeg/downloads.aspx

[40] Joscowiz J. and Ardao J.C.L, "Enhancements to the Opinion-Model for Video Telephony Applications," *5th International Latin American Networking Conference (LANC)*, pp.87-94, New York, 2009 doi: 10.1145/1636682.1636697

[41] Song W, Tjondronegoro D and Docherty M, "Exploration and Optimization of User Experience in Viewing Videos on a Mobile Phone'" *International Journal on Software Engineering Knowledge*, vol.20, no.8, pp.1045-1075, 2010

[42] Hanhart P. (2013). *VQMT: Video Quality Measurement Tool* [Online]. Available: http://www.mmspg.epfl.ch/vqmt

[43] Simone D, Goldmann F, Lee L and Ebrahimi T, "Towards High Efficiency Video Coding: Subjective Evaluation of Potential Coding Technologies," *Journal of Visual Communication and Image Representation*, vol.22, no.8, pp.734-748, 2011