The Impact of Network Impairment on Quality of Experience (QoE) in H.265/HEVC Video Streaming

James Nightingale, *Member*, IEEE, Qi Wang, *Member*, IEEE, Christos Grecos, *Senior Member*, IEEE and Sergio Goma

Abstract — *Users of modern portable consumer devices (smartphones, tablets etc.) expect ubiquitous delivery of high quality services, which fully utilise the capabilities of their devices. Video streaming is one of the most widely used yet challenging services for operators to deliver with assured service levels. This challenge is more apparent in wireless networks where bandwidth constraints and packet loss are common. The lower bandwidth requirements of High Efficiency Video Coding (HEVC) provide the potential to enable service providers to deliver high quality video streams in low-bandwidth networks; however, packet loss may result in greater damage in perceived quality given the higher compression ratio. This work considers the delivery of HEVC encoded video streams in impaired network environments and quantifies the effects of network impairment on HEVC video streaming from the perspective of the end user. HEVC encoded streams were transmitted over a test network with both wired and wireless segments that had imperfect communication channels subject to packet loss. Two different error concealment methods were employed to mitigate packet loss and overcome reference decoder robustness issues. The perceptual quality of received video was subjectively assessed by a panel of viewers. Existing subjective studies of HEVC quality have not considered the implications of network impairments. Analysis of results has quantified the effects of packet loss in HEVC on perceptual quality and provided valuable insight into the relative importance of the main factors observed to influence user perception in HEVC streaming. The outputs from this study show the relative importance and relationship between those factors that affect human perception of quality in impaired HEVC encoded video streams. The subjective analysis is supported by comparison with commonly used objective quality measurement techniques. Outputs from this work may be used in the development of quality of experience (QoE) oriented streaming applications for HEVC in loss prone networks* [1].

Index Terms — H.265, HEVC, subjective evaluation, packet loss, quality of experience (QoE), benchmarks.

I. INTRODUCTION

The recent adoption by the ITU-T of High Efficiency Video Coding (HEVC) [1] as H.265 [2], the latest video coding standard, is a significant milestone for the consumer electronics community. Standardisation of H.265, which consumes up to 50% less bandwidth than the currently dominant encoding standard H.264/AVC [3], will drive the development of new products and services in areas such as video conferencing, broadcasting, home and mobile entertainment. In an increasingly connected world much of the video content encoded using the H.265 standard will be delivered over converged IP networks.

Given the growing trend [4] for users to consume content on mobile devices, which are often deployed in error-prone wireless network environments, it is important and timely to consider the potential impact of network impairments on perceived video quality (and the resultant user satisfaction level). The greater compression ratio of HEVC, when compared with pre-existing encoding standards, means that each network packet carries a bigger percentage of the encoded stream than was the case with previous encoding schemes. This study reports on a set of streaming experiments and subjective evaluations designed to quantify the impact of packet loss in HEVC on perceptual quality.

These experiments were conducted on a realistic testbed consisting of a hybrid wired/wireless network. Video sequences were encoded using the HEVC reference software version HM8.0 [5] and transmitted over the testbed where random packet loss rate between 1% and 5% were introduced. The HM8.0 decoder was modified to firstly overcome robustness issues related to packet loss and secondly to provide two alternative methods of error concealment. A panel of human subjects evaluated the impact on perceived quality of packet loss, video sequence content, error concealment method, spatial resolution and encoded bitrate. Each factor was ranked according to the significance derived from the scores given by the viewers who were shown to consider packet loss to have the most significant impact on quality. Packet loss rates of 3% or more were shown to cause high levels of user dissatisfaction. The choice of error concealment method and the content of each sequence were also shown to have a high level of significance in viewers' perception of quality. The other factors were shown to each have a measurable, but less noticeable impact on

[1] This work was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant No. EP/J014729/1: Enabler for Next-Generation Mobile Video Applications.

James Nightingale, Qi Wang and Christos Grecos are with the Centre for Audio Visual Communications and Networks, University of the West of Scotland, Paisley, UK. (e-mail: james.nightingale@uws.ac.uk, qi.wang@uws.ac.uk, christos.grecos@uws.ac.uk).

Sergio Goma is with the Multimedia Research Lab, Qualcomm Inc., San Deigo, USA. (e-mail: sgoma@qti.qualcomm.com).

Contributed Paper
Manuscript received 03/27/14
Current version published 06/23/14
Electronic version published 06/23/14.

0098 3063/14/$20.00 © 2014 IEEE

quality. The results of subjective testing were augmented by the application of objective testing methods to the same test data set. The Peak Signal-To-Noise Ratio (PSNR) method widely used in video streaming comparison did not accurately reflect the views of the panel, while the alternative VQM metric produced results closer to those of the viewers although there was still some significant divergence. The work presented in this paper provides a timely analysis of the factors that influence video quality in HEVC encoded streams, identifies and weights the main factors influencing quality and sets threshold levels for user tolerance to network impairment.

The rest of this paper is organized as follows. Section II discusses background on HEVC streaming and quality evaluation. Section III introduces the testbed setup, streaming experiments and subjective evaluation parameters. In Section IV the results of both subjective evaluations and objective testing are presented and analysed. Finally Section V concludes the paper.

II. Background and Related Work

As the H.265 standard has only very recently been approved, the majority of published work has concentrated on the design of encoding tools. Other than the contribution of packet loss software [6], which does not handle error concealment, no outputs from the standardization efforts for H.265/HEVC have looked at the effects of network impairment on HEVC encoded bitstreams. Existing work on HEVC streaming under impaired network conditions [7] has been limited to the use of objectively measured video quality assessment using the PSNR metric, which is known to be an imperfect measure of human quality perception for video [8]. The HEVStream proposal [7] did not specifically address the impact of Packet Loss Rates (PLRs) or other important factors such as error concealment methods, content type or spatial resolution on video quality.

There are currently only a few published subjective evaluations of HEVC. The work reported by Ohm *et al*. [9] was conducted in support of the aims of the standardisation effort and focused on comparing the subjective quality of HEVC with that of previous encoding standards; it did not consider any form of transmission channel impairment. A separate study by Garcia and Kalva [10] considered the use of HEVC in mobile environments and compared it with the current H.264/AVC standard, concluding that there was little difference in perceptual quality at typical mobile phone bandwidths. The only previously published work [11] to evaluate packet loss in HEVC streams subjectively reported a limited set of early results from this study.

III. Subjective Evaluation

A total of twenty four volunteers took part in the subjective testing program. The proportion of male and female volunteers was approximately equal. Ages ranged from 22 to 50. Occupations were PhD students (19), academic staff (1) and support staff (4). All of the volunteers were in good health and wore appropriate corrective lenses where required. The breakdown of participant ages is shown in Fig. 1. Volunteers were recruited by means of an email campaign within the university. Ethical approval was sought and obtained for the conduct of this study using human subjects.

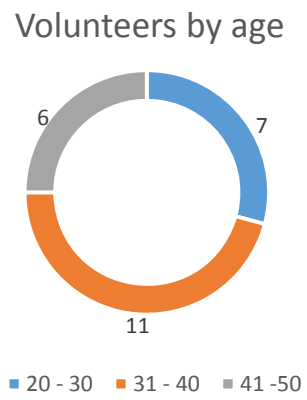


Fig. 1. Distribution of participants by age range.

Testing was conducted over a period of two weeks in July 2013. The testing environment consisted of a single testing station that was designed to be compliant with the requirements of ITU-R BT500-13 [12]. The room area surrounding the testing station was mid-grey in colour. The testing station consisted of a 22" LCD monitor and a high specification workstation running widely used commercial video quality evaluation software. The method of assessment is the Double Stimulus Impairment Scale (DSIS) method specified in ITU-R BT500-13 [12]. In this method the viewer was presented with the reference video for 10 seconds, used a mouse on the workstation to manually acknowledge the end of the sequence and was then presented with the impaired video, again for 10 seconds duration. In each case the reference video was the encoded and locally decoded test sequence which was then transmitted over an imperfect network to obtain the impaired version.

The encoder configuration (Quantisation Parameter, bitrate, spatial resolution etc.) was the same for both reference and impaired sequences. As has been previously highlighted by Reiter *et al* [13], two types of distortion (encoder and transmission) are encountered in video streaming. This work focused on evaluation of transmission distortions only. The reference video was of the same quality as the 0% packet loss case with no bias introduced by encoder distortion. After clicking on the mouse at the end of the impaired sequence, the participants were asked to give their opinion of the quality of the impaired video in comparison to that of the reference video. Viewers were offered the option to review both reference and impaired videos before making a decision on relative quality.

The subjective opinion of each viewer was recorded using the Absolute Category Rating (ACR) scale in which viewers were asked to give their opinion of quality on a scale from 1 to 5 representing the categories of opinion as shown in TABLE 1. Each individual testing session lasted no longer than 60 minutes, with participants being asked to assess a number of test sets, each containing between 10 and 15 pairs of examples

(impaired and reference) of the same sequence. Participants were encouraged to take short breaks between tests sets, each of which contained tests from a different video sequence/spatial resolution combination. All participants were required to sit at the distance (for the spatial resolution under test) specified in ITU-R BT500-13 [12].

TABLE I
5 POINT ABSOLUTE CATEGORY RATING SCALE

Category Rating	Opinion Score
Very Annoying	1
Annoying	2
Slightly Annoying	3
Perceptible, but not annoying	4
Imperceptible	5

Testing was conducted using six test sequences drawn from those used for conformance testing of HEVC. Two different spatial resolutions were used, chosen to represent typical resolutions for mid-range smartphones (416*240) and tablet devices (832*480) respectively. Some sequences were encoded with a temporal resolutions of 30 frames per second (fps) and others containing a higher degree of motion at 50 fps.

TABLE II
LIST OF ALL TEST SEQUENCE CONFIGURATIONS

Class	Sequence	Frame Rate	Spatial Resolution	QP	Kbps
Low	Keiba	30fps	416*240	24	521
			832*480	42	731
	Racehorses	30fps	416*240	29	443
			832*480	37	463
	Flowervase	30fps	416*240	17	460
			832*480	27	478
	BlowingBubbles	50fps	416*240	30	519
	BasketballPass	50fps	416*240	30	459
	BasketballDrillText	50fps	832*480	38	462
Med.	Keiba	30fps	416*240	17	1335
			832*480	37	1383
	Racehorses	30fps	416*240	21	1375
			832*480	29	1488
	Flowervase	30fps	416*240	11	1305
			832*480	22	1420
	BlowingBubbles	50fps	416*240	23	1460
	BasketballPass	50fps	416*240	23	1458
	BasketballDrillText	50fps	832*480	29	1440
High	Keiba	30fps	416*240	13	2231
			832*480	33	2225
	Racehorses	30fps	416*240	17	2254
			832*480	27	2025
	Flowervase	30fps	416*240	10	1557
			832*480	19	2080
	BlowingBubbles	50fps	416*240	21	1993
	BasketballPass	50fps	416*240	21	2111
	BasketballDrillText	50fps	832*480	26	2153

Three bandwidth testing points were used, which were chosen to represent typical mobile broadband speeds in the UK as reported by the official regulator [14]. These speeds were 500 Kbps (low), 1400 Kbps (average) and 2200 Kbps (high). A complete list of the testing points used is given in TABLE II, which also shows the QP used in the HM8.0 [5] encoder to match, as closely as possible, the required bandwidth for each testing point. Sequences used are shown in Fig. 2.



Fig. 2. Example frames from each of the video sequences used.

Each test sequence was encoded using the HM8.0 encoder in both *Random-Access* and *Intra-Only* encoder configurations. Test sequences were encapsulated for transmission using the HEVStream framework [7] and transmitted over the hybrid wired and wireless networking testbed shown in Fig. 3. Each Network Abstraction Layer (NAL) unit of an HEVC stream was encapsulated in a single Real-time Transport Protocol (RTP) packet using the RTP payload format for HEVC specified by Schierl *et al*. [15]. The Maximum Transmission Unit (MTU) matching option in the HM8.0 encoder was used to encode NAL units at a size capable of being carried in an IPv6 packet payload, which has a maximum capacity of 1460 bytes.

Network impairment, in the form of random packet loss, was introduced at the Wide Area Network (WAN) emulation router in the testbed. This was achieved using the NETEM and IPROUTE2 components of the Linux Kernel. Packet loss ratios of 1%, 3% and 5% were used as testing points, sequences were also transmitted without any packet loss to provide a datum for comparison. Packets from the video stream were received at the client, decapsulated using the HEVStream framework [7] and passed to a modified HM8.0 [5] decoder. The HEVStream framework [7] and the priority weighting proposal [16] for HEVC both employed a modified decoder in order to overcome issues of decoder robustness to packet loss. This the same modified decoderwas used in this work. Particularly in respect of *Random-Access* sequences

where the HM8.0 decoder would fail when presented with non-conforming sequences with unmet dependencies within the Reference Picture Set (RPS). In addition to decoder modifications to improve robustness, further modifications were added to provide a level of error concealment for missing NAL units. Two methods of error concealment were used. In the case where the first NAL unit of a picture was lost, the entire missing picture was recreated by copying from the nearest available picture in the Decoded Picture Buffer (DPB) of the HM8.0 decoder.

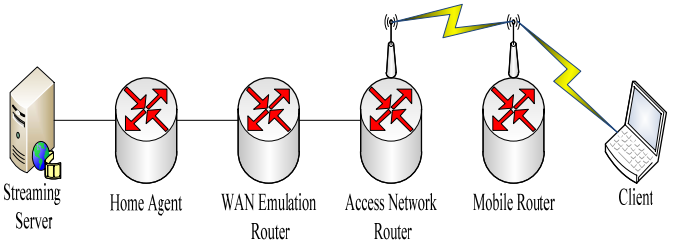


Fig. 3. The topology of the physical testbed environment deployed to obtain the impaired video sequences used in subjective testing.

This was determined for *Intra-Only* sequences as the immediately preceding picture in the DPB, while for *Random-Access* sequences the RPS (of the current picture as determined from the last recorded Sequence Parameter Set's control NAL unit) was searched to determine the nearest available picture. In other cases where NAL units other than the first NAL unit were lost, the co-located area from the nearest picture (identified in the same manner as for the first case) is copied to replace the missing area in the current picture.

Subjective evaluations were supplemented by the use of two well-known objective measurements of video quality. The PSNR metric is calculated from the Mean Square Error (MSE). Given an image I with dimensions $m \times n$ and its noisy approximation K, MSE is defined as shown in (1) and PSNR is derived from (2).

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (1)$$

$$PSNR = 20 \cdot log_{10}(MAX_I) - 10 \cdot log_{10}(MSE) \quad (2)$$

The other objective metric employed was the Video Quality Metric (VQM) defined in ITU-R J.144 [17]. VQM is intended to provide a closer approximation of human perceptual quality than alternative metrics such as PSNR.

IV. Results

The subjective evaluation experiments primarily investigated human perception of the effects of packet loss on HEVC encoded bitstreams; however, analysis of the impact of error concealment method, bandwidth, spatial resolution and video content on the perception of quality is also provided. A summary of the results for each of the test sequences is given in TABLE III. The Mean Opinion Score (MOS) was calculated by the testing software. The mean MOS from all individual test cases for each sequence are shown together with the standard deviation and the number of individual comparisons for each testing point.

Statistical analysis of the observed scores was conducted using statistical tools to identify the impact and significance of each of the above factors. Subjective scores were screened to remove outliers (see Section V.F).

TABLE III
Mean Opinion Scores For Each Sequence

Class	Sequence	Spatial Resolution	Mean MOS	Std. Dev.	*N*
Low	Keiba	416*240	3.64	1.21	95
		832*480	3.41	1.18	97
	Racehorses	416*240	3.64	1.23	84
		832*480	3.53	1.36	88
	Flowervase	416*240	3.61	1.21	95
		832*480	3.43	1.24	96
	BlowingBubbles	416*240	3.42	1.30	96
	BasketballPass	416*240	3.33	1.18	96
	BasketballDrillText	832*480	3.33	1.18	98
Med.	Keiba	416*240	3.79	1.09	98
		832*480	3.44	1.13	98
	Racehorses	416*240	3.93	1.16	95
		832*480	3.79	1.26	96
	Flowervase	416*240	3.78	1.09	97
		832*480	3.44	1.25	96
	BlowingBubbles	416*240	3.53	1.16	97
	BasketballPass	416*240	3.51	1.20	100
	BasketballDrillText	832*480	3.41	1.18	99
High	Keiba	416*240	3.70	1.18	94
		832*480	3.47	1.21	96
	Racehorses	416*240	3.95	1.20	86
		832*480	3.76	1.24	89
	Flowervase	416*240	3.70	1.18	94
		832*480	3.48	1.22	96
	BlowingBubbles	416*240	3.58	1.14	97
	BasketballPass	416*240	3.50	1.19	98
	BasketballDrillText	832*480	3.43	1.20	97

A. Packet Loss

Random packet loss can have a significant detrimental effect on the quality of video streams. Given the higher compression ratio of HEVC over previous encoding methods, it is important to establish the tolerance of HEVC encoded bitsreams to random packet loss.

Table IV provides a summary of the MOS for each packet loss scenario investigated. The results are given as the mean MOS of all experiments for each different packet loss category. A minimum of 608 measurements (N) were made in each category. The mean MOS achieved when no packet loss occurred was measured as 4.4013 (from a maximum 5) indicating that the consensus amongst users was that the difference between the original video and the reconstructed video was either "imperceptible" or "perceptible but not annoying". These results are broadly similar to those obtained in other loss-free studies [9] conducting subjective analysis of HEVC encoded content in support the standardisation effort. The standard deviation in MOS when no packet loss occurred was 0.63671 with a minimum assessed score of 3 (slightly

annoying) and a maximum assessed score of 5 (imperceptible). The median MOS score, for the 0% packet loss case, was 4 (perceptible but not annoying) with a variance of 0.85. 8% of scores in this packet loss category were given as 3 (slightly annoying). The range of cases at each level of satisfaction is given in the boxplot in Fig. 4. A detailed analysis of scores in the slightly annoying category for 0% packet loss indicated that these were primarily attributed to comparisons at the low bandwidth testing point (@500 Kbps) where the level of quantisation error required to meet the bandwidth constraint introduced 'blurring' effects [18] observable by the test subjects. Although clear guidance and training was provided, a number of subjects appeared to be influenced by this.

The introduction of a 1% rate of random packet loss reduced the mean MOS to 4.2531 with most test subjects still describing the impaired video as either "imperceptible" or "perceptible but not annoying". The number of tests rating the impaired video as "slightly annoying" rose from 8% in the 0% packet loss case to 15.7%. 5.1% of tests scored the video quality at 2 (annoying). Although the median score for the 1% packet loss case remained at 4 (perceptible but not annoying), the level of variability in test results rose with the standard deviation in MOS from 0.63671 in the 0% packet loss scenario to 0.79459 in the 1% loss case.

Similarly variance from the median rose from 0.85 to 1.4 indicating a divergence of opinion on the quality of the impaired video. Given that both the 0% and 1% packet loss cases have the same median MOS score of 4 (perceptible but not annoying), although with an increased variability in the distribution of scores, it can be concluded that the random loss of 1% of packets is tolerable to users without significantly reducing their QoE. In the 1% loss case an analysis of low scores (annoying) determined that two factors (low bandwidth and error concealment method) were common factors.

TABLE IV
MEAN OPINION SCORES BY PACKET LOSS RATIO

PLR	N	Mean	Std. Dev	Std. Error	Min.	Max.
0%	608	4.4013	0.63671	.02582	3.00	5.00
1%	648	4.2531	0.79459	.03121	2.00	5.00
3%	675	2.9822	1.10957	.04271	1.00	5.00
5%	637	2.7190	1.14424	.04534	1.00	5.00
∑	2568	3.5736	1.20609	.02380	1.00	5.00

When a random packet loss rate of 3% was introduced, the mean MOS dropped significantly to 2.9822 and the degree of variability, as expressed by the standard deviation, increased to 1.10957. The median of the normal distribution fell to 3 (slightly annoying) and the variance from the median rose to 2.48 indicating a significant variability in the MOS scores. Over 65% of the test results at 3% PLR scored the video quality as 3 (annoying) or worse.

The 5% packet loss case further reduced the mean MOS to 2.7190, but perhaps more significantly, increased the variability in MOS scores. The standard deviation further increased to 1.14424 and the variance from the median, which remained at 3 (annoying) increased to 2.62. 19% of all test results scored the video quality as 1 (very annoying).The plot of mean MOS scores shown in Fig 5 clearly shows the significant reduction in subjective quality assessment that took place between 1% packet loss and 3% packet loss.

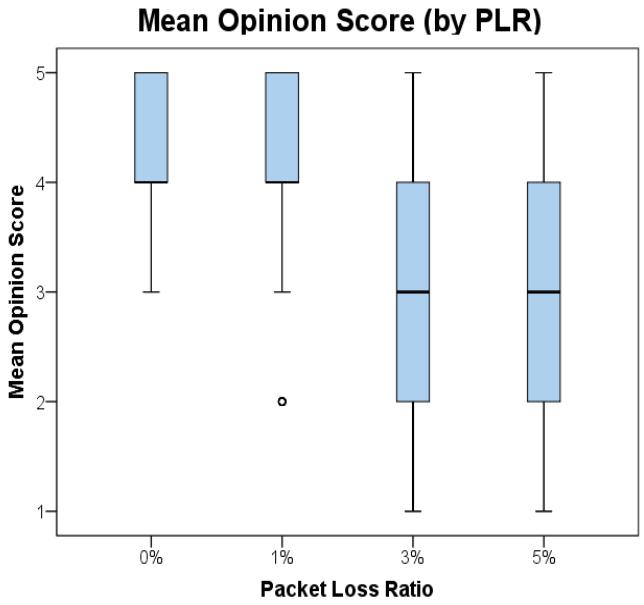


Fig. 4. Box plot showing the range of MOS scores for each packet loss category. 17 cases (<2%) were recorded at 2 for the 1% PLR scenario.

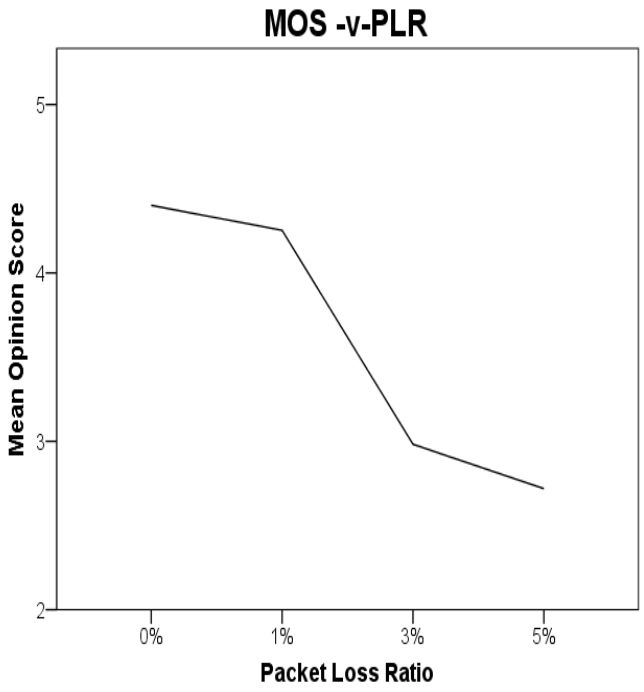


Fig. 5. Mean of all MOS scores for each packet loss category, showing the mean MOS reduction between 0% PLR and 5% PLR.

These observations on the subjective opinions of the effects of packet loss on video quality suggest that a loss rate of 1% is tolerable to users while a loss rate of 3% is intolerable to most users who found the quality to be annoying (or worse).

Consequently results of this evaluation suggest that packet loss rate greater than 3% in HEVC streaming reduces the viewer's QoE beyond what they would reasonably accept as a level of service provision.

Since the 2% packet loss case (or other intermediate step) was not evaluated in this study, the conclusion drawn is that user tolerance of packet loss deteriorates significantly in the interval between 1% and 3% PLR and that video quality is deemed unacceptable beyond 3% PLR.

B. Per Sequence

A detailed breakdown of the mean MOS scores for each of the test sequence/encoder configurations used is given in Table III and a summary of the mean MOS scores and standard deviation for each of the test sequences used is also shown in Table V. In respect of the *BlowingBubbles*, *BasktballPass* and *BasketballDrill* sequences; the number of samples examined was approximately half of those for the other three video sequences as only one spatial resolution was used in these cases. It can be seen that these sequences, which also have a higher (50 fps) frame rates, tended to score lower than the overall mean.

The *Racehorses* sequence exhibited a slightly higher-than-average MOS compared with all other sequences, exceeding the mean MOS by approximately 0.2. This can be attributed to the fact that fewer *Racehorses* sequences that had been subjected to 3% or 5% packet loss were presented to the viewers than was the case for other sequences. This anomaly was a result of randomised selection of reference/impaired sequence pairs (from the available pool of prepared samples) by the testing station software.

The standard deviation did not vary significantly between sequences, indicating that although the assessed quality varied, the level of agreement on MOS between subjects remained consistent. Normal distributions of MOS for each sequence are given in Fig 6. The median was 4 (perceptible but not annoying) for all sequences with variance from the median being consistent in the range from 2.84 and 3.08 for every sequence.

TABLE V
OVERALL MEAN MOS SCORES FOR EACH SEQUENCE

Sequence Name	Mean	N	Std. Deviation
Keiba	3.5761	578	1.17244
Racehorses	3.7714	538	1.24389
Flowervase	3.5749	574	1.20375
BlowingBubbles	3.5103	290	1.19756
Basketballpass	3.4490	294	1.19260
Basketballdrilltext	3.3912	294	1.18583
∑	3.5736	2568	1.20609

Differences in MOS were observed between sequences. These can mainly be attributed to the different content types of the sequences used in the experiments as each sequence has different spatial-temporal characteristics and to the explanations given above.

Results for the 0% packet loss case were similar to those observed in other subjective evaluations [9] of HEVC where no transmission losses were considered. The results for 1%, 3% and 5% packet loss cases suggest that there is a complex relationship between content type/classification and packet loss. It can be seen from Fig 7 that, although the results for each sequence broadly follow the mean for all sequences shown in Fig 5, differences exist that highlight the contribution made by the type of content contained in a video sequence.

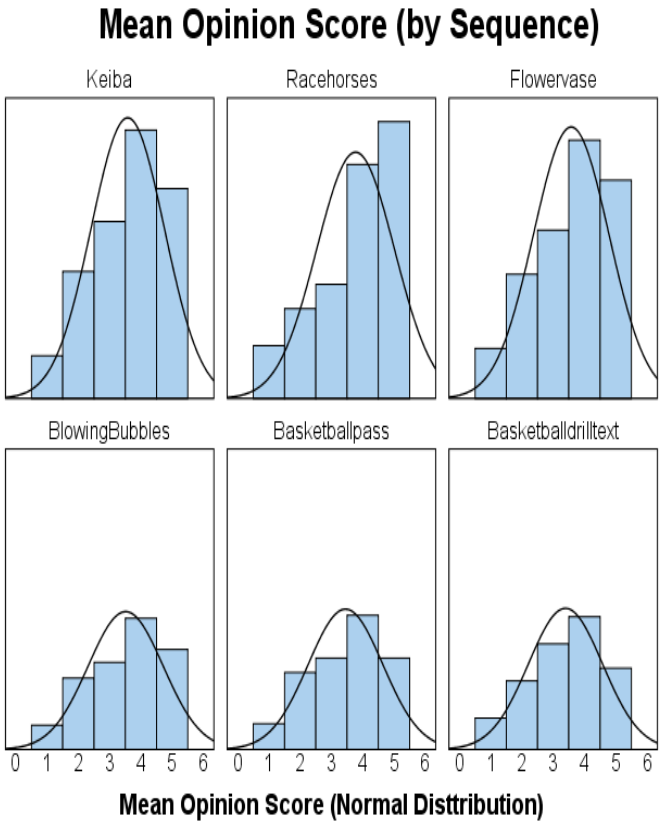


Fig. 6. Normal distribution of MOS scores for each video sequence.

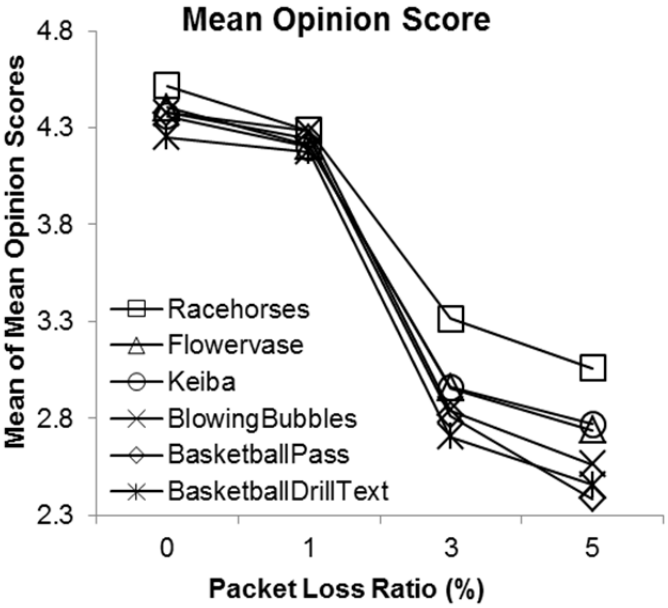


Fig. 7. Comparison of mean MOS scores for each video sequence highlighting the difference between each sequence across the range of packet loss testing points.

C. Bandwidth and Spatial Resolution

Other factors that may influence a viewer's perception of video quality include the bandwidth at which the sequence is encoded and the spatial resolution of the sequence. In order to meet a bandwidth constraint, a sequence may be encoded using a higher QP of the encoder, which introduces a greater degree of error while reducing the number of bits required for encoding. Reducing the QP introduces visual artefacts, most notably blurring. Table VI shows the mean MOS and standard deviation for each of the three bandwidth testing points. The mean MOS for the low bandwidth scenario was 3.4805 with a median of 3 (slightly annoying). The mean MOS increased to 3.6244 for the medium bandwidth scenario. This improvement is consistent with the QP change method of matching lower bandwidths used in this work. In the high bandwidth scenario, viewers scored the mean MOS as 3.6139, which was slightly below that for the medium bandwidth scenario.

TABLE VI
MEAN MOS SCORES BY BANDWIDTH

Bandwidth Range	Mean	N	Std. Deviation
Low	3.4805	845	1.23194
Medium	3.6244	876	1.17956
High	3.6139	847	1.20320
∑	3.5736	2568	1.20609

The median for both medium and high bandwidth cases was 4 (perceptible but not annoying). In the higher bandwidth scenario the encoder configuration resulted in more NAL units (and network packets) carrying a payload that only consisted of a small portion of a picture. Packet loss in these scenarios was more commonly concealed using the block copy method, which (see section VI. D) has been shown to be less effective. This provides an explanation for the small difference in mean MOS between the medium and high bandwidth scenarios. The standard deviation in MOS was similar for all three bandwidth scenarios, indicating a similar level of agreement on quality between viewers.

Spatial resolution was also shown to be a factor in the viewer's assessment of video quality. The mean MOS score for the higher resolution sequences was 0.14 less than that for the lower resolution sequences, with a marginally increased variability expressed in terms of the standard deviation in TABLE VII. This indicates that viewers were more sensitive to impairments in higher resolution sequences.

TABLE VII
MEAN MOS SCORES BY SPATIAL RESOLUTION

Spatial Resolution	Mean	N	Std. Deviation
416*240	3.6392	1422	1.18871
832*480	3.4921	1146	1.22294
Total	3.5736	2568	1.20609

As can be seen from Fig. 8, the normal distribution of MOS scores for each spatial resolution was broadly similar, with both having the same median. A larger number of comparisons were made at the lower resolution as can be observed in Fig. 8. As can be noticed from Fig. 8, the normal distribution of MOS scores for each spatial resolution was largely similar, with both having the same median.

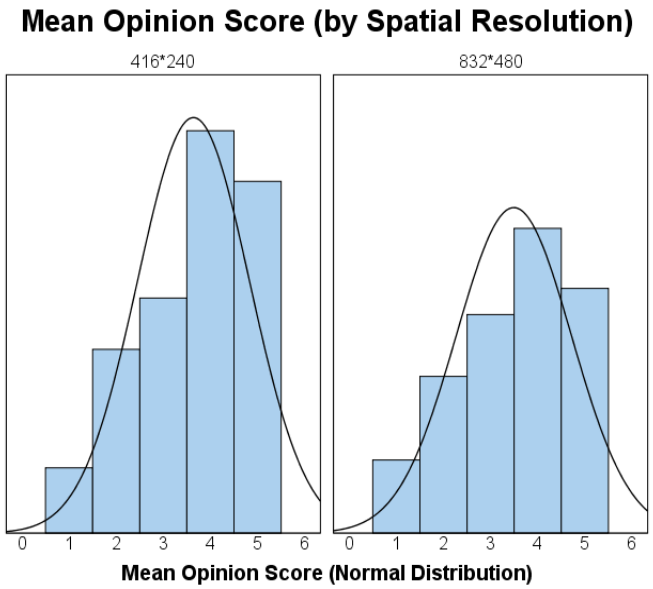


Fig. 8. Normal distribution of MOS scores by spatial resolution.

D. Error Concealment Method

The choice of an error concealment method, whilst primarily governed in this study by the type of NAL unit lost, was shown to have a significant impact on the MOS scores assessed by the participants. The method in which a full frame was copied resulted in the perception of the video sequence 'freezing', for a short period of time while the method where only part of the frame was copied resulted in the introduction of visual artefacts and blocking into the video stream.

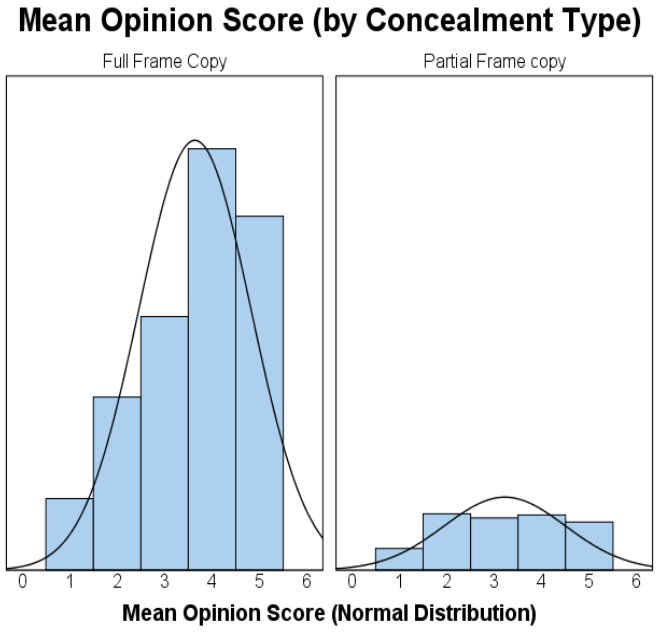


Fig. 9. Mean MOS scores for each error concealment method.

The normal distribution of MOS for each error concealment method is shown in Fig. 9. The median MOS score for the full frame method was 4 (perceptible but not annoying) whereas the median for the partial copy method was 3 (slightly annoying).

TABLE VIII
MEAN MOS SCORES BY CONCEALMENT METHOD

Concealment Method	Mean	N	Std. Deviation
Full Frame Copy	3.6382	2170	1.18174
Partial Frame copy	3.2211	398	1.27615
Total	3.5736	2568	1.20609

E. Comparison to Alternative Metrics

The results in TABLE VIII demonstrate that the viewers were more tolerant of the full frame copy method by a MOS of 0.42. It can also be seen that the level of agreement on quality between participants also reduced in the partial frame copy method as signified by the increased standard deviation. Fig. 10 shows the objectively measured PSNR scores taken as the mean for all comparisons. When compared with the subjectively assessed MOS scores in Fig. 5, it can be observed that PSNR-based assessment, although decreasing markedly with packet loss, does not identify the sharp decline in subjective MOS between 1% PLR and 3% PLR.

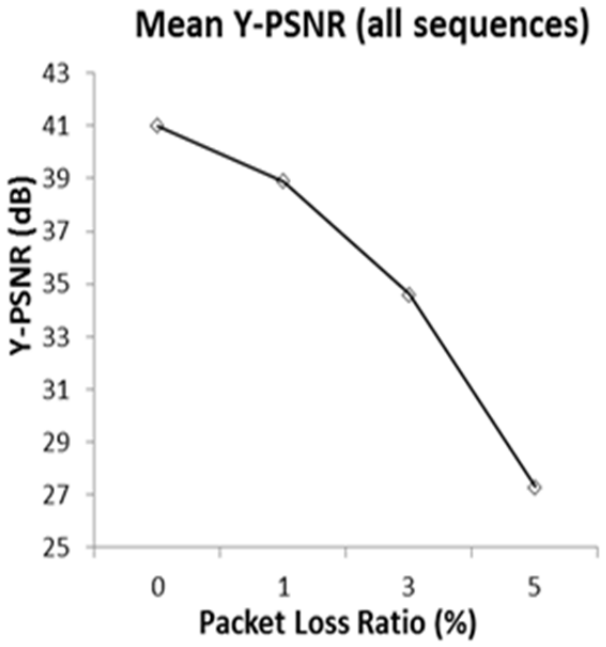


Fig. 10. Mean PSNR scores for each packet loss category.

The mean VQM scores for all comparisons are shown in Fig. 11. VQM outperforms PSNR by following the trend in subjective MOS more closely. In particular, a noticeable difference in the rate of change of VQM between 1% PLR and 3% PLR indicates that the VQM metric follows the subjective MOS score better than PSNR.

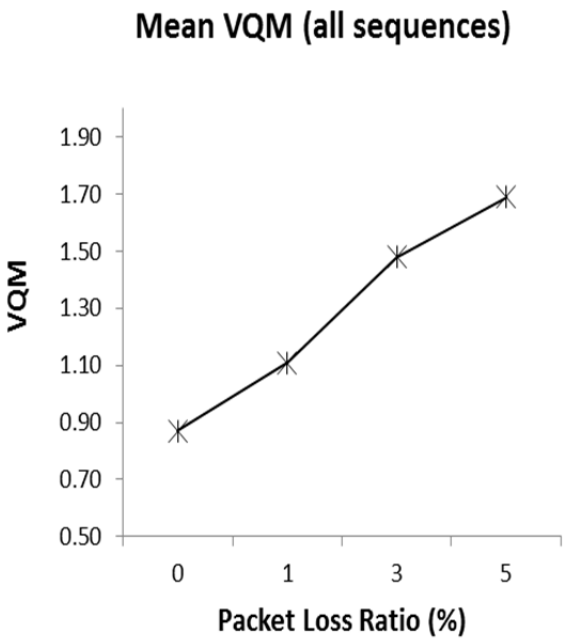


Fig. 11. Mean VQM scores for each packet loss category.

F. Consistency between Subjects

The mean scores recorded by each participant were examined to identify any anomalies. While there were several individual comparisons that fell out of the expected norms, the mean scores for all but one of the subjects fell within the range specified in ITU-R BT500-13 [12]. Of the 2731 individual assessments conducted, 163 (6%) were excluded. Fig. 12 shows the mean scores for each participant.

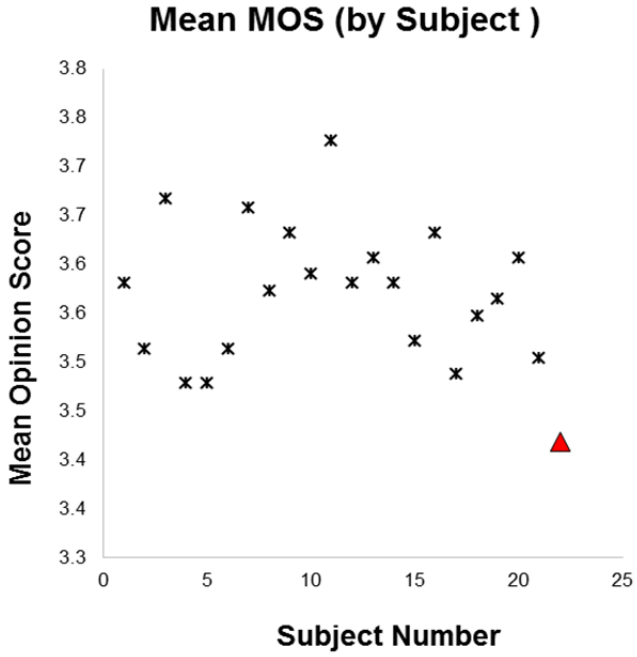


Fig. 12. A comparison on mean MOS scores recorded by each participant. The red triangle shows the excluded subject.

V. CONCLUSION

This paper presents the results of a subjective evaluation of the perceptual quality or QoE of video streaming, resulted from the delivery of HEVC encoded bitstreams over impaired network channels. The timeliness and importance of this work arises from the recent finalisation of the H.265 standard, which will herald the development of a new generation of consumer device video applications.

The results and observations made in this paper will act as drivers for the development of QoE driven streaming solutions for H.265 in challenging network scenarios. Future work in this area may take this study further by using the insights gained to design an H.265 decoder that is more robust to packet loss and incorporates error concealment methods. The outputs from this work may also be explored to help shape the design of reduced reference or no reference metrics for H.265.

REFERENCES

[1] B. Bross et al, "High efficiency video coding (HEVC) text specification draft 10", JCT-VC document JCTVC-L1003, Jan 2013.

[2] ITU, "High efficiency video coding", ITU-T Rec. H.265, April 2013.

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, July 2003.

[4] Cisco, Visual Networking Index Forecast 2011 – 2016.

[5] JCT-VC, "HEVC test model under consideration (TMuC) revision 8.0," JCT-VC Contribution HM 8.0, Sept. 2012.

[6] S. Wenger, "NAL unit loss software," JCT-VC Document, JCTVC-H0072, Feb. 2012.

[7] J. Nightingale, Q. Wang, and C. Grecos, "HEVStream: a framework for streaming and evaluation of high efficiency video coding (HEVC) content in loss-prone networks," *IEEE Trans. Consumer Electron.*, vol.58, no.2, pp.404-412, May 2012.

[8] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment", *Electronic Letters*, vol. 44, no. 13, pp. 800-801, June 2008.

[9] J-R. Ohm G. J. Sullivan, H. Schwarz, T. K. Tan and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol.22, no.12, pp.1669-1684, Dec. 2012.

[10] R. Garcia and H. Kalva, "Human mobile-device interaction on HEVC and H.264 subjective evaluation for video use in mobile environment," *Proc. of IEEE International Conference on Consumer Electronics (ICCE 2013),* Jan. 2013.

[11] J. Nightingale, Q. Wang, C. Grecos and S. Goma, "Subjective evaluation of the effects of packet loss on HEVC encoded video streams," *Proc. of 3rd International Conference on Consumer Electronics – Berlin (ICCE Berlin 2013)*, Sept. 2013.

[12] ITU, "Methodology for the subjective assessment of television pictures," ITU-R BT500-13, Jan. 2012.

[13] U. Reiter, J. Korhonen and J. You, "Comparing apples and oranges: assessment of the relative video quality in the presence of different types of distortion," *EURASIP Journal on Image and Video Processing*, 2011:8, doi:10.1186/1687-5281-2011-8, Sept. 2011.

[14] OFCOM, "Measuring mobile broadband in the UK", Regulator market report, Dec 2010.

[15] T. Schierl, S. Wenger, Y.-K, Wang and M. M. Hannuksela, "RTP payload format for high efficiency video coding," IETF Internet Draft, Sept. 2013.

[16] J. Nightingale, Q. Wang, and C. Grecos, "Priority-based methods for reducing the impact of packet loss on HEVC encoded video streams," *Proc. of SPIE Real Time Imaging 2013*, San Francisco, Jan. 2013.

[17] ITU, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference", ITU-R J.144, June 2003.

[18] C-M Fu et al, "Sample adaptive offset in the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.* , vol.22, no.12, pp.1755,1764, Dec. 2012

BIOGRAPHIES



James Nightingale (S'09-M'12) is a Postdoctoral Research Associate with Centre of the Audio-Visual Communications & Networks (AVCN) in the University of the West of Scotland (UWS), UK, working on the EPSRC funded project "Enabler for Next-Generation Mobile Video Applications" (EP/J014729/1). His research interests include mobile networks, multihoming and video streaming techniques. He received the BSc (Honours) degree in Computer Networks from UWS with First Class Honours and won the Best Honours Dissertation Prize. He received his PhD from UWS with Outstanding Progression Award**.**



Qi Wang (S'02-M'06) is a senior lecturer with the Centre of Audio-Visual Communications & Networks (AVCN) in the University of the West of Scotland (UWS), UK. He is the Principal Investigator of the UK EPSRC project "Enabler for Next-Generation Mobile Video Applications" (EP/J014729/1) and several industry-sponsored projects. His research interests include mobile/wireless networks, Internet protocols and visual applications, and cloud computing. He has published over 60 peer-reviewed papers in these areas. He was a winner of two IEEE ICCE Best Paper Awards (2012 and 2014), a finalist of the IEE 3G2003 Best Paper Award, and a recipient of a UWS 2013 STARS Award (Highly Commended) for "Outstanding Research & Knowledge Exchange" and an ORS Award by the Higher Education Funding Council for England (HEFCE). He is a regional editor for the international journal of Recent Patents on Telecommunications by Bentham Science, and on the technical programme committees of over 20 international conferences.



Christos Grecos (M'01-SM'06) is a Professor in Visual Communications Standards, and Head of School of Computing, the University of the West of Scotland (UWS), UK. He leads the Centre of Audio-Visual Communications and Networks (AVCN) with UWS, and his research interests include image/video compression standards, image/video processing and analysis, image/video networking and computer vision. He has published many research papers in top-tier international publications including a number of IEEE transactions on these topics. He is on the editorial board or served as guest editor for many international journals, and he has been invited to give talks in various international conferences. He has been the Principal Investigator for several national or international projects funded by UK EPSRC or EU. He received his PhD degree in Image/Video Coding Algorithms from the University of Glamorgan, UK.



Sergio Goma is a senior director at Qualcomm Inc. in San Diego, USA, where he leads the multimedia R&D standardization group for imaging, driving Qualcomm's vision on future imaging technologies. His research interests include computational photography and programmable hardware architectures. Before joining Qualcomm, Dr Goma developed the image processing solution present in AMD's Imageon series of chips. He received the Ph.D. in reliability and fault tolerance of computers and holds several US patents on image processing algorithms and architectures.