Visibility of individual packet loss on H.264 encoded video stream – A user study on the impact of packet loss on perceived video quality

Mu Mu[*a], Roswitha Gostner[a], Andreas Mauthe[a], Gareth Tyson[a], Francisco Garcia[b]

[a] Computing Department, Lancaster University, United Kingdom
[b] Agilent Laboratories, Edinburgh, United Kingdom

ABSTRACT

Assessing video content transmitted over networked content infrastructures becomes a fundamental requirement for service providers. Previous research has shown that there is no direct correlation between traditional network QoS and user perceived video quality. This paper presents a study investigating the impact of individual packet loss on four types of H.264 main-profile encoded video streams. Four artifact factors to model the degree of artifacts in video frames are defined. Further, the visibility of artifacts considering the video content characteristics, encoding scheme and error concealment is investigated in conjunction with a user study. The individual and joint impacts of artifact factors are explored on the test video sequences. From the results of user tests, the artifact factor-based assessment method shows superiority over PSNR-based and network QoS based quality assessment.

Keywords: Quality of experience, Quality of Service, packet loss impact, user study

1. INTRODUCTION

Various types of video content are being delivered over packet-based content distribution networks. Quality degradation of packet based delivery is due to network impairments and a lack of hard QoS provisions. This can result in the loss or deterioration of packets in encoded video streams during transmission. The absence of this data within specific frames results in artifacts (distortions within the video frames) which will be visible by end users. Packet loss rate (PLR) is a common network QoS parameter to evaluate the quality of video delivery. PLR-based assessment is based on the assumption that video packets have equal significance on perceptual video quality. However, the impact of the loss of a specific packet depends on a number of factors related to the encoding, packet transmission scheme and decoding. Thus, for example, the importance of adjacent packets within the same stream can vary considerably. In order to establish the impact of the loss of a specific packet on the perceived quality, individual packet loss must be analysed separately.

In this paper, the perceptual impact of individual packet loss on H.264 main-profile encoded video streams is investigated. The research is divided into two phases so the results can contribute to a flexible assessment model. Methodology for estimating the artifacts in the video frame by analyzing delivery information, video codec and video content characteristics is studied in the first phase. Four artifact factors to model the degree of artifacts are also defined. In the second phase, nature and the visibility of artifacts are studied in conjunction with subjective user test to build the correlation between video distortion and user perception. Hypothesis on artifact factors of each test sequences are presented and tentative estimations on the joint impact (user perceived quality) of artifacts is also performed to evaluate the artifact factor-based quality assessment. From the results of user tests, the artifact factor-based assessment method shows superiority over PSNR-based and network QoS based quality assessment. Hypothesis on artifacts factors are also verified with the user test results. The presented study is part of a wider research programme aimed at establishing a link between network impairments and user perceived quality. This aims to result in the creation of a network level Quality of Experience (QoE) service that helps to maintain an acceptable (user perceived) quality level in the face of packet loss [1].

The rest of the paper is structured as follows: Section 2 presents the background issues of video content delivery over

[*] m.mu@comp.lancs.ac.uk; phone +44 (0) 1524 510383; fax +44 (0) 1524 510492; www.comp.lancs.ac.uk

packet networks, as well as related work. In section 3 the visibility of artifacts is investigated in describes the impact of different video codec schemes, video content characteristics and user preference on the visibility of packet loss. User tests, test results and discussion are presented in Section 4, 5 and 6 respectively. Section 7 concludes the paper.

2. BACKGROUND AND RELATED WORK

2.1 Video Codecs and Communication Issues

This section discusses the different issues influencing the quality of video when transmitted over packet switched networks. Main factors are the codec and the kind of network impairments that can occur.

2.1.1 H.264 Codec, Error Control and Error Concealment

With the increasing demand for different qualities of video transmitted over the Internet, a video coding standard, taking into account the specific requirements of emerging high quality video applications was designed and approved by the ITU-T (as recommendation H.264) and ISO/IEC (as the international standard MPEG-4 part 10 Advanced Video Coding (AVC)). Within the H.264 standard, only the decoder process is standardized with syntaxes and parameters. Subsets of syntaxes and parameters are defined as *profiles*. It is guaranteed that all the decoders which conform to a certain profile will produce similar output when given an encoded bitstream that also conforms to the same profile. Four profiles are defined in the first version of the standard: *Baseline*, *Main, Extended* and *High* Profile. The Baseline Profile is applicable to delay-sensitive conversational services. The Main Profile is designed for digital storage media and television broadcasting. The Extended Profile aims at multimedia services over the Internet. The High Profile is defined in the fidelity range extensions for applications such as content contribution, content distribution, and studio editing [2]. In this paper, the discussion and experiments are based on the Main profile of H.264 codec.

Compression techniques are conventionally used to reduce the redundancy. This is usually done by eliminating bit level redundancy and perception redundancy in both the spatial dimension (intra-frame) and the temporal dimension (inter-frame). The video frame can be encoded to either I-frame, P-frame or B-frame. All of the macroblocks in I-frame are coded using intra prediction. I-frame is immune to artifacts from previous and subsequent frames. Macroblocks of the P-frame and B-frame can also be encoded with inter prediction. In P-frame, only *one* motion-compensated prediction to the previous frame is permitted. The content of P-frame can also be used for the inter prediction of the other slices. Artifacts may be adopted by the P-frame from previous key frames and propagated to the following frames. *Two* motion-compensated predictions are allowed for the B-frame. In previous standards (e.g. MPEG 2), two motion-compensations are defined as bidirectional prediction where one prediction is derived from subsequence prediction signals and another is derived from previous picture. H.264 supports not only the forward/backward prediction pair, but also forward/forward and backward/backward pairs [3]. Two forward references can be beneficial for motion compensated prediction of a region just before scene change, and two backward references just after scene change[2]. However, the error robustness of the video is then considerably reduced with the inter-frame or intra-frame prediction. For example, individual packet loss may propagate to adjacent macroblocks in current, previous or following frames. Error control and error concealment techniques are introduced to minimize the impact of data absence on user perception.

Slicing is a common error control method. Several macroblocks constitute a slice and the spatial prediction is allowed only for macroblocks within the same slice. Error propagation is then terminated at the end of each slice. The slices are encapsulated into packets in a way that each packet will carry an integer number of slices to guarantee the effectiveness of the slicing mechanism.

Error concealment reconstructs the lost data by predicting from known knowledge of the video in the spatial or temporal domain. The error concealment is out of the scope of the H.264 standard. Applications may choose to employ certain concealment algorithms for the best results. There are two recommended non-normative error concealment schemes: intra-frame interpolation and inter-frame interpolation. For intra-frame interpolation, the content of missing macroblocks is deduced from the pixel values of spatially adjacent macroblocks as a weighted sum of the closest boundary pixels of the selected adjacent macroblocks. The weight associated with each boundary pixel is relative to the inverse distance between the pixel to be concealed and the boundary pixel [4]. For inter-frame interpolation, if the motion of the target area is relatively small, lost macroblocks are replaced by content from the reference frame. If the motion is high, the motion-vector of the missing macroblock is concealed from motion-vector of the neighbouring blocks. More details about these two schemes can be found in [5] and [6].

2.1.2 IP Video Delivery – Network Impairment

The lack of network resources brings three major impairments to video content during delivery: network delay, jitter and packet loss. Depending on the application characteristics, the packets that exceed a certain delay threshold will be recognized by end applications as lost packets. Variations of network delay are known as jitter. Modern user devices are usually equipped with the de-jitter buffer to smooth transmission delay while introducing buffer delay. With buffering mechanisms, jitter will either be transferred to buffer delay or packet loss when the buffer is overwhelmed. Packet loss is mainly caused by congestion as a result of insufficient or non-optimal usage of network resource. For a connection-less transport protocol such as UDP, the loss may directly affect application performance. The effect of packet loss is usually expressed by packet loss rate (PLR). PLR is defined as the ratio of lost packets during transmission to the total number of transmitted packets.

It can be concluded from the discussion above that network impairments result in either playout delay or packet loss effects. The impact of playout delay is decided by application properties. In this paper, the packet loss impact on streaming video is investigated. Due to the nature of modern video codecs, data losses in certain parts of the data stream are more visible to end user than loss in other parts. Moreover, the visibility of packet loss is also greatly dependent on the video content characteristics.

2.2 Related Work

A number of researchers have addressed issues related to the relationship between data loss and user perception [3-7]. Lopez *et al* [7] investigated the combined analytical correlation between user experience to all the relative parameters by using a "black box" system with a H.264 encoded sequence and subjective experiments. Though, the impact of packet loss cannot only be judged by the loss rate. Boyce *et al* [8] studied the different packet loss patterns and their impact, i.e. absolute packet loss, conditional packet loss and packet loss rate over time. They use the frame error rate to quantify the loss effect. Their work suggests that even a small packet loss rate may translate in to much higher frame error rates. For example, a 3% packet loss percentage could translate into a 30% frame error [8].

Verscheure *et al* [9] studied the impact of bit rate, data loss and their combined impact on MPEG-2 video quality with a spatio-temporal model of human vision [10, 11]. The authors explored how the video quality behaves according to the average encoding bit rate with respect to spatio-temporal complexity. Their research shows that the video quality remains constant until the PLR reaches certain value and that the higher the average bit rate, the lower the PLR after which the video quality drops. This is due to the constant packet size with which higher bit rates are achieved by more packet units. However, we believe that the bit-rate has no direct link to user experience concerning packet loss because video content is represented by a series of frames. Bit-rate only decides the balance between the frame rate and the quality of each frame.

Most of the previous studies on the impact of packet loss on user perceived video quality aimed to find a correlation between network impairments and user experience. However, the results which reflect directly to network QoS are only suitable for certain sets of configurations (e.g. codec, bitrates or application type). When certain configuration is changed, the entire research which usually includes costly and time-consuming user test must be repeated.

In order to overcome the disadvantages of previous studies, in this paper, research of video assessment is divided into two phases. In the first phase, we investigate the methodology for estimating the artifacts in the video frame by analyzing packet delivery information, video codec and video content characteristics. In the second phase, characteristics and the visibility of artifacts are studied with user tests so to explore the correlation between artifacts and user perception. The two-phase study is still codec dependent but highly scalable. For different application scenarios, the artifact estimation in the first phase can be conveniently modified regarding the specific codec, delivery parameter sets etc. As long as the visibility of artifacts has been modelled, no more costly and time-consuming subjective tests will be required to build the model from sketch. In Section 3.1, artifact factors are defined to model the visual artifacts. The codec and characteristics of video content which affect the degree of artifact factors of a packet loss are discussed in Section 3.2.

3. PERCEIVED QUALITY: INFLUENCING FACTORS

3.1 Artifact factors

Previous research into the visibility of artifacts is based on a single threshold method. Participants were asked whether or not they observed artifacts and then the visibility was quantified by the percentage of positive responses from participants [12]. This single threshold method is suitable for investigating the limit of human perception for artifacts. However, the viewer's opinion on the degree of deterioration is not considered. Investigating the degree of deterioration is particularly critical for video content which is encoded with modern codecs. The advanced features of modern codecs, such as error control and concealment mechanisms, decide that the visible artifacts deteriorate user experiment differently. In order to quantify the deterioration degree of artifact, the visibility of the artifact in a video stream is modeled using four artifact factors (Figure 1): Spatial Inconsistency (SPIC), Spatial Extent (SPXT), Spatial Priority (SPPR) and Temporal Duration (TPDR) as listed in Table 1.

Table 1 Artifact factors

SPIC	Spatial Inconsistency
SPXT	Spatial Extent
SPPR	Spatial Priority
TPDR	Temporal Duration

The SPIC represents the discontinuity between artifacts and their neighbouring content. In signal processing, the SPIC is measured by the contrast at the edge of macroblocks. However, it is believed that video content may hide or magnify the spatial discontinuity effect. Due to the nature of the H.264 codec, video artifacts are in the units of fixed-sized macroblocks. Individual packet loss may affect one or more of these macroblocks directly. The number of consecutive macroblocks impaired by an individual packet loss then defines the SPXT. One important issue is that the viewers' attention is often on particular areas of the video. The SPPR is defined as the priority of the region where artifacts are located. SPPR is high when artifact occurs in the region of interest (ROI) of the user's attention. The further the artifact away from the ROI is, the lower the SPPR will be. The propagation of the artifact not only increases the possibility of it being observed by the viewer but also worsen the user experience. TPDR is defined as the number of the frames that are affected by a packet loss.

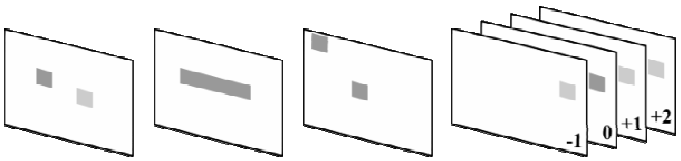

Figure 1 Artifact factors (a) SPIC (b) SPXT (c) SPPR (d) TPDR

Both content-dependent and content-independent issues affect the four artifact factors. SPXT and TPDR are related to the codec, while SPIC and SPPR rely on characteristics of the video. In the next section, codecs and characteristics of video content are discussed.

3.2 Codec and Characteristics of Video Content

Video codec, characteristics of video content and priority of user's attention are the main Influencing factors which affect the degree of artifact factors of a packet loss.

The SPIC is highly dependent on the frame type, motion and complexity of the video. As it is presented in Section 2.1, when the data of an I-frame is lost, the intra-frame interpolation algorithm in the decoder conceals the loss of data by prediction from neighbouring macro blocks. If the I-frame is complex, the concealment algorithm is unlikely to achieve high performance with satisfactory results because the texture of the missing macro block is replaced by prediction from the spatial content at the edge of the macroblock. Hence the SPIC is high in I-frames for the video with high complexity and low for the video with simple content. Data loss in the P-frame and B-frame is concealed with inter-frame interpolation by content from the reference frame or from the motion-vector of the neighbouring blocks depending on the

motion of the video. The SPIC is proportional to the level of the motion of video content for artifacts in the B-frames and P-frames. Furthermore, the SPIC is also affected by the nature of the motion in the video content. Football and Betes contain a large amount of irregular motion (motion-vector varies between frames) so the performance of inter-frame interpolation will be lower which results in high SPIC. In contrast, the SPIC is relatively low for loss in P- and B-frame in videos with well-regulated motion even if the motion level is high.

Artifacts can propagate to previous and/or following frames as a result of inter-frame prediction of video codec. The TPDR is decided by the dependency of the deteriorated video frames. TPDR is proportional to the number of frames that the target frame is depended on.

As it is described in Section 2.1, error propagation within the frame will be terminated at the end of each slice. The SPXT is determined by the slice length of the video codec. For instance, if the slice mode is set as *fixed number of macroblock per slice* then the SPXT is decided by the number of macroblock in slices. If the slice mode is *fixed number of bytes per slice* then the SPXT is decided by the entropy of the frame.

The viewer's attention on different parts of the frames is different. The SPPR is high when artifacts occur in the region of interest (ROI) of the user's attention. The further the artifact is away from the ROI, the lower the SPPR will be. However, when watching video content with an equal interest level throughout the frame, users are likely to scan different areas of the frame in order to find interesting objects. In this case, the SPPR effect would be relatively low to the visibility of artifacts. For instance, an artifact with a high SPIC but low SPPR may not be detected by the participants during the user test.

In the next section, the subjective testbed and test plan are designed to study the correlation between artifacts and user perceptions.

4. ASSESSING THE IMPACT OF INDIVIDUAL PACKET LOSS ON THE PERCEIVED QUALITY: USER EXPERIMENTS

The aim of the study is to investigate the visibility of artifacts, considering video content characteristics, the encoding scheme and the decoder's error concealment. A *testbed* and *test plan* have been developed following relative recommendations and video test standards [13-15].

4.1 Video sequences

Table 2 Test sequences

Motion \ Complexity	High	Low
High	Football	Betes
Low	Autumn	Susie

We selected four videos (named Susie, Betes, Football and Autumn) from the test sequences published by the Video Quality Experts Group [13] reflecting two content characteristics (motion and complexity) (see Table 2). Football has a high level of texture and complexity in the sequence as several subjects move quickly to follow the ball. We assume that the football in the video is the user's region of interest. In contrast, Susie has a low level of motion and texture because there are only a few head movements in the foreground. Further, both the foreground and background are simple. We assume the face to be the region of interest. The cartoon sequence (Betes) and the Autumn are in the middle of our classification scheme of content characteristics. For instances, Betes has a high level of motion and a low level of texture with the centre of the screen as region of interest. In contrast, Autumn has a low level of motion with only the waterfall in the middle moving while the high amount of autumn trees forms a high level of texture.

Figure 2 outlines the motion variance over time and Figure 3 shows the different texture complexity for each video. None of the videos keep their characteristics within the whole playing time e.g. there are obvious head movements around 5.5 seconds in Susie. We avoid performing experiments on the part of video which are controversial to the global characteristics of the video sequences to achieve a higher level of homogeneity. For instance, in the Susie video (low

motion and low complexity), we don't introduce packet loss between the fifth and sixth second of video sequence which contain quick head movement.

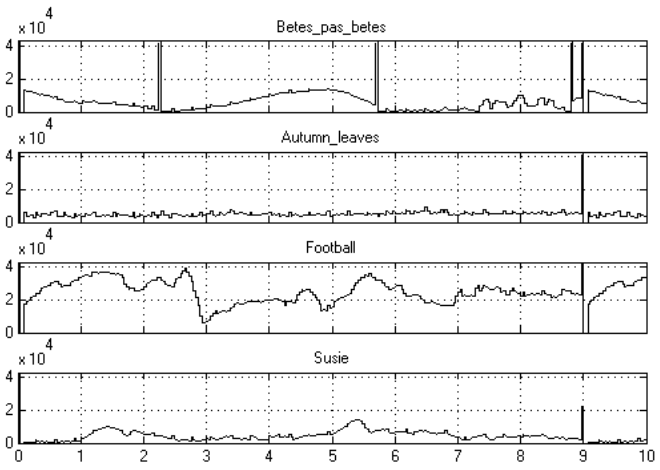

Figure 2 Motion of the test sequences

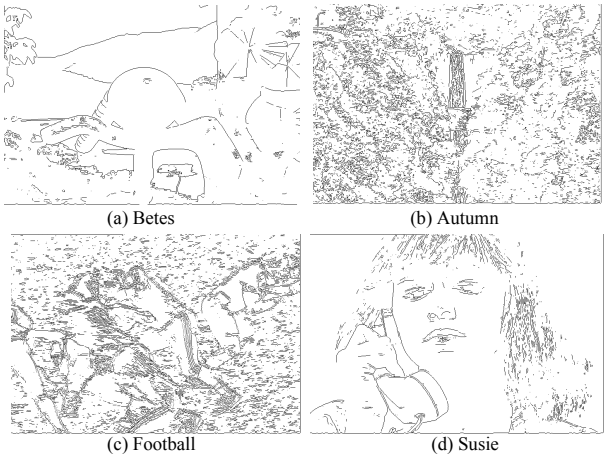

Figure 3 Texture of the test sequences

4.2 Design

The study is within a subject design with three independent variables: type of packet loss (with three levels, I-frame, P-frame, B-frame), content characteristic (the four levels as described in previous section) and the region of interest (packet loss within and outside of ROI). These variables result in 3x4x2=24 different test conditions. One test block consists of 24 trials. Latin squares were used to randomize video sequences and counterbalance the order of presentation.

4.3 Apparatus

All uncompressed sequences are encoded by H.264/AVC JM Reference Software [16] with the main profile of H.264 codec. Key parameters of the encoder are list in Table 3:

Table 3 Encoding parameters

Parameter	Value	Parameter	Value
Framerate	30	NumberReferenceFrames	2
IDR Period	3	NumberBFrames	2
Frameskip	2	Slice mode	45 macroblocks per slice (which covers the width of the frame)

We used an RTP loss simulator to remove content from a certain packet to simulate the three types of packet loss during transmission (I, P and B-frames). The SPXT is constant for all artifacts because the length of the slice is fixed in the encoding parameter. According to the encoding scheme of the test sequences, GoP (Group of Pictures) length is fixed at 9. In the loss simulator, only the P-frame that follows I-frame is selected for the loss simulation on P-frame type, hence, the frame dependency of I-, P- and B-frame are settled. TPDR is identical between each type of frame. Both reference (with no packet loss) and deteriorated (with packet loss) videos are decoded to the YUV uncompressed video sequence to assess the packet loss impact. Basic intra-frame and inter-frame interpolation are applied to the video for the integrity of the video on screen.

A Power PC (Intel® Core™2 Duo Processor E6700 2.66GHz, 4.0GB 800MHz DDR2 SDRAM Memory) with a 17-inch Dell Monitor on the resolution of 1024*768 was running the trials. The distance between the eyes of participants and the screen was six time the height of the video sequence, as recommended with ITU [14]. Each sequence has 720 pixels per horizontal line and 486 active lines per frame. The video is centred on the screen. The videos did not use the full screen size; hence the videos were played against a black monitor background.

4.4 Participants

24 participants (12 of them are male and the other 12 are female) took part in the study where the average age was 24.4. The participants include 12 undergraduates, 6 PhD students, 2 master students and 2 university employees while none of them was part of our research group. Regarding their computer skill level, 11 rated their expertise as "medium", 8 as "high" and 4 as "expert", and one as "poor". 23 out of the 24 participants have watched online videos before, 14, of whom, watch videos on a daily basis, 6 on a weekly and 3 on a monthly basis.

4.5 Test Procedure

First, the investigator introduces a test video trial. A trial consists of watching the non-impaired video (for reference), followed by the same video including a single packet loss impairment. Only for introducing the test trial, the investigator points out the impaired frames to reduce the impact of the irrelevant artifacts based on compression. We used the Double Stimulus Impairment Scale (DSIS) [15] questionnaire with a five level scale (1="very annoying", 2="annoying", 3="slightly annoying", 4="perceptible but not annoying" and 5="imperceptible"), as recommended by [17] for evaluating visible artifacts caused by transmission errors. After the introduction trial, each participant watches first a complete test block (consisting of 24 trials) followed by a break during which we conducted a structured interview about their background and finished with a second test block which also consists 24 trials. On average, a participant needed between 30 and 45 minutes to complete the experiment.

5. HYPOTHESIS AND EXPERIMENT RESULTS

5.1 Hypothesis and MOS estimation

The mean PSNR values of deteriorated video are presented in Table 3. The mean PSNR of video sequence is calculated by averaging the PSNR of all the frames. The PSNR of the uncorrupted video frames are bounded by 100 dB instead of infinite so to model the quality of video over all frames. Values greater than 37 are mapped MOS of 5 according to the PSNR to MOS mapping table (Table 4) which is recommended in [18].

Table 4 PSNR to MOS map [18]

PSNR [dB]	MOS
>37	5 (Excellent)
31-37	4 (Good)
25-31	3 (Fair)
20-25	2 (Poor)
<20	1 (Bad)

The SPIC, TPDR and SPPR of artifacts in each test video sequences are estimated from the discussion in Section 3.2 as in Table 5. SPXT is not listed because it is constant for the experiment due to the slice mode of the codec (Table 3).

The perceived video quality (MOS) is then estimated from the joint deterioration effect of artifact factors. The estimation is done by first defining a worse case (e.g. I-frame-Autumn-ROI scores high on all the factors) or best case (e.g. B-frame-Susie-outside ROI) and scored the rest of the video referring to the selected reference case considering artifact factors. This estimation method is preliminary but can indicate the feasibility of the proposed artifact factor based method by comparing the relativity between estimation results and the relativity between test results. The user test results will verify the hypothesis and improve the artifact factor based quality assessment.

5.2 User test results

In the result section, the impacts of the SPPR, TPDR and SPIC on the user test results are presented. Figure 4 shows the joint impacts of ROI and frame types on video content. Figure 5 present the ROI impacts on different video types. Figure 6 shows the average MOS of all the video sequences regardless of the ROI.

Table 5 Expected MOS of test video sequences

Packet Loss Type in Video Sequence	Average PSNR	MOS estimated from PSNR	SPIC	TPDR	SPPR	MOS estimated from artifact factors
I-frame -Betes (ROI)	44.1	5	Low	High	High	4
I-frame -Betes (outside ROI)	54.9	5	Low	High	Low	5
P-frame-Betes (ROI)	49.2	5	High	Medium	High	2
P-frame-Betes (outside ROI)	52.5	5	High	Medium	Low	3
B-frame-Betes (ROI)	53.9	5	High	Low	High	3
B-frame-Betes (outside ROI)	65.3	5	High	Low	Low	4
I-frame-Autumn (ROI)	55.6	5	High	High	High	1
I-frame-Autumn (outside ROI)	55.1	5	High	High	Low	2
P-frame-Autumn (ROI)	66.3	5	Low	Medium	High	4
P-frame-Autumn (outside ROI)	72.1	5	Low	Medium	Low	5
B-frame-Autumn (ROI)	76.3	5	Low	Low	High	5
B-frame-Autumn (outside ROI)	80.0	5	Low	Low	Low	5
I-frame-Football (ROI)	53.8	5	High	High	High	2
I-frame-Football (outside ROI)	54.3	5	High	High	Low	3
P-frame-Football (ROI)	48.2	5	High	Medium	High	1
P-frame-Football (outside ROI)	51.5	5	High	Medium	Low	2
B-frame-Football (ROI)	59.5	5	High	Low	High	3
B-frame-Football (outside ROI)	62.2	5	High	Low	Low	4
I-frame-Susie (ROI)	58.7	5	Low	High	High	2
I-frame-Susie (outside ROI)	57.5	5	Low	High	Low	3
P-frame-Susie (ROI)	66.8	5	Low	Medium	High	3
P-frame-Susie (outside ROI)	66.1	5	Low	Medium	Low	4
B-frame-Susie (ROI)	75.2	5	Low	Low	High	5
B-frame-Susie (outside ROI)	71.3	5	Low	Low	Low	5

5.2.1 Impact of SPPR

Figure 4 shows the average user scoring divided for the two independent variables: region of interest and frame types. On average, participants scored I-frame packet loss lower (=lower perceived quality) within the ROI (M=3.75, SD=1.07) than packet loss outside the ROI (M=4.20, SD=0.85). For the two other frame type, participants scored results similarly high; for instance, P-frame ROI (M=4.22 SD=0.94) outside the ROI (M=4.40, SD=0.81) and B-frame packet loss outside the ROI (M=4.64, SD=0.60) and inside the ROI M=4.65, SD=0.58). Moreover, the difference between ROI and outside ROI is only great (0.46) for Susie (ROI: M=4.14, SD=0.92, outside ROI: M=4.60, SD=0.60) (Figure 5).

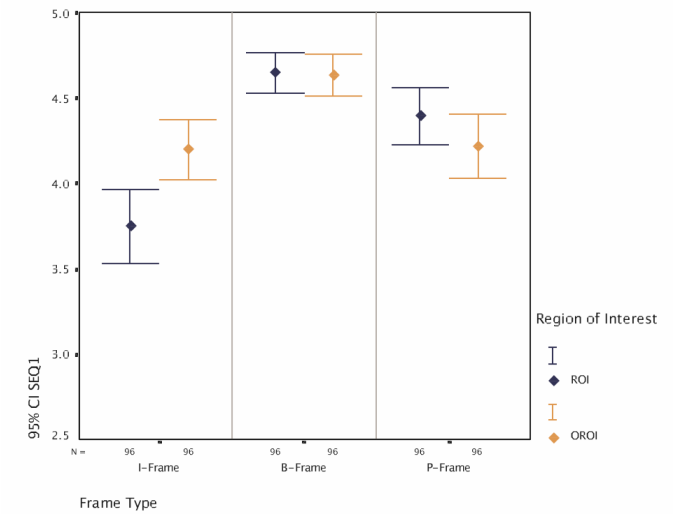

Figure 4 Frame type versus ROI

5.2.2 Impact of TPDR

We study the TPDR (number of frames affected by packet loss) by comparing the MOS of B-frames and P-frames. This is because that the artifacts of these frame types are both results of inter-frame interpolation concealment (to limit the joint impact from the SPIC). Figure 4 clearly shows that participants rated artifacts for B-frame packet loss higher (better video quality) than for P-frame packet loss for both the ROI and outside the ROI.

Looking at the results with respect to the different content characteristics, Figure 6 clearly shows that movies with a low level of movement (Susie, Autumn) are less sensitive to both B-frame and P-frame packet loss, in fact, for both frame types, participants rated these two movie sequences fairly high (for B-frame Susie has M=4.60, SD=0.64, Autumn M=4.52 SD=0.65 and for P-frame Susie has M=4.54, SD=0.62, Autumn has M=4.60, SD=0.68). In contrast, the difference between B-frame rating and P-frame rating is larger when the content type has a high level of movement. Participant's rating for B-frame loss (Betes M=4.77, SD=0.42 and Football M= 4.67, SD=0.60) is on average higher than for P-frame loss, (Betes M=4.21, SD=0.92 and Football M=3.88, SD=1.06) as outlined in Figure 6. These results clearly indicate that P-frame packet loss in content with high movement is more perceptible by the user than for video sequences with low motion level.

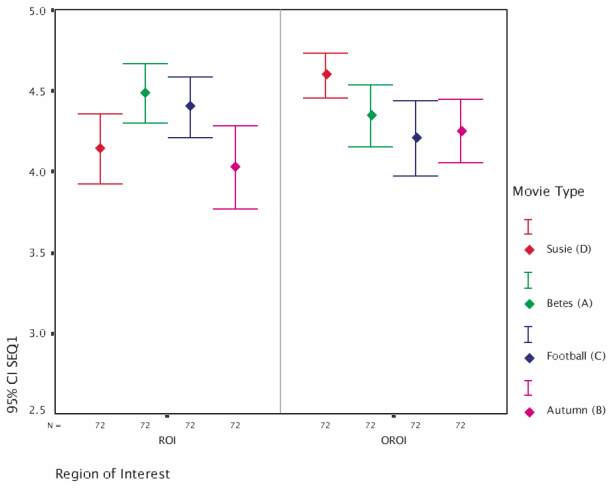

Figure 5 ROI versus Video characteristics

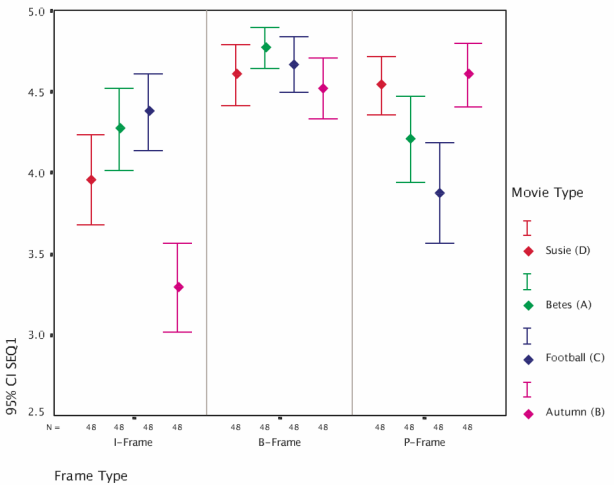

Figure 6 Frame type versus video characteristics

5.2.3 Impact of SPIC

The SPIC represents the discontinuity between artifacts and the neighbouring contents. Video sequences with a low level on motion (such as Susie and Autumn) show a low rating for I-frame packet loss and a high rating for P-frame loss (see Figure 6 with Susie M= 3.96 SD=0.97 and Autumn M=3.29 and SD=0.94 for I-frame and Susie M=4.54, SD=0.62 and Autumn M=4.60 and SD=0.968 for P-frame). In contrast, video sequences with a high level of motion (Football, Betes) received a lower scoring for I-frame packet loss than for P-frame packet loss, see Figure 6 with Football M=4.38 SD=0.82 and Betes M=4.27 and SD=0.87 for I-frame and Football M=3.88 SD=1.06 and Betes M=4.21 SD=0.92 for P-frame. Figure 6 shows clearly that for B-frame artifacts, all scoring is above 4.5 regardless of the content characteristics of the different movies (Susie M=4.60, SD=0.64; Betes M=4.77 SD=0.42, Football M=4.67 SD=0.60, Autumn M=4.52, SD=0.65). Although the SPIC is high for B-frames, the number of frames affected by packet loss is so low that the impact of SPIC is diluted and results in a low visibility for all studied content characteristics.

6. DISCUSSION

6.1 PSNR-based MOS estimation

As discussed in Section 5.1, PSNR values are high for all the test videos with single packet loss artifacts. With the PSNR-MOS mapping scheme [18], all videos score 5 by MOS estimation from PSNR. The subjective user test results have a wide range from 2.96 to 4.75 (Figure 7), which indicates that one individual packet loss is capable of deteriorating video quality down to "slightly annoying" and different types of packet loss have different significance on the perceived video quality.

6.2 Artifact factor-based MOS estimation

The MOS from the user tests are mostly higher than the MOS estimation according to the artifact factors in Table 5. Most of the MOS are above 4.0 in the results, which imply that artifacts in the test videos are detected by the participants but the quality of the impaired video sequences is still considered acceptable. From the results of the questionnaires, it can be concluded that the scale of MOS is largely affected by user preferences. During our test, participants scored under the *free* video on demand scenario so that unless the artifacts obscure the key parts of the video frame beyond recognition, participants will always give a score above 3.0 (as the artifacts are not annoying). However, several participants declared that if the scenario was to be changed to *pre-paid* video streaming services, most of the scores will be downgraded by one or two levels (to "slightly annoying" or "annoying").

Results presented in the Section 5 have shown the impact of artifact factors (SPIC, SPPR and TPDR) on user test results. In the following, we compare the user test results to the artifact factor-based MOS assessment method.

In Section 5.1, it was estimated that MOS will be higher for outside the ROI than ROI, but this is only significant for video content with clear ROI area. As can be seen from Figure 4, for video sequence with lower motion, artifacts within the ROI (high SPPR) are more visible than the ones out of the ROI (low SPPR). In contrast, the SPPR has no obvious impact on videos with high motion. Figure 4 shows that the effect of the SPPR on user perceived quality is correlated with the MOS of the artifact. For I-frame loss, the difference between the ROI and outside ROI is almost 1.0 (1 level of 5). When the scores of the ROI and outside the ROI both increase for P-frame loss, the difference between them is reduced. For B-frame loss, the MOS is particularly high which results in unnoticeable difference between ROI and outside ROI scores.

The impact of the TPDR can be studied by comparing the MOS of B-frames and P-frames because artifacts in these two types of frames are both results of inter-frame interpolation concealment. The TPDR significantly increases the deterioration level in P-frames in comparison with B-frames for Betes and Football. The MOS of B-frame and P-frame loss are similar for Susie and Autumn because the videos with low motion level are less sensitive to the length of the B-/P-frame artifacts as is discussed in Section 5.1. Figure 4 also reveals the TPDR effects. For both the ROI and outside-ROI circumstances, the artifacts in P-frames appear in several consecutive frames leading to more damage to the user perception than the artifacts in B-frames.

According to the hypothesis in Section 5.1, for packet loss on I-frames, the SPIC is in proportion to the complexity of video content. In contrast, the SPIC is proportional to the level of motion of video content within B-frames and P-frames. The SPIC is higher for video containing higher levels of irregular motion (motion-vector varies between frames) such as

Football and the Betes for B- and P-frame loss. The test results (Figure 6) show that the SPIC of I-frame artifacts is low for videos with low complexity (Betes and Susie) and high for videos with high complexity (Autumn). The MOS of I-frame artifacts for Football is relatively higher than what is expected. It is believed that the high MOS is a result of the special characteristics of the Football video. The Football scene is different from other video content by its irregular and unpredictable high complexity motion within the frames. Even when the SPIC and TPDR is high for I-frame artifacts in Football, it is hard for the viewers to recognise the I-frame artifacts from the neighbouring content (which is also proven by the results of the questionnaire section in the user tests). It also can be seen from Figure 6 that MOS are relatively high for all the B-frames artifacts because the TPDR of B-frame is so low that the impact from the SPIC are diluted. Even though SPIC is believed to be high in B-frame for the videos with high level of motion, TPDR still intensively reduced the visibility of the artifacts. The SPIC is high for P-frames in videos with high motion levels which can be approved by comparing I-frame artifacts and P-frame artifacts of Betes and Football. Although the P-frame artifacts have lower TPDR than the I-frame artifacts, the MOS of P-frame artifacts in Betes and Football are still lower than I-frame loss due to the high SPIC level. The results prove our previous hypothesis, that the SPIC is high for I-frame artifacts in the videos with high complexity and for B, P-frame artifacts in the videos with high motion. Moreover, within videos with low motional level (Susie and Autumn), the B-frame artifacts and P-frame artifacts are both less visible because the SPIC is considerably low due to high performance of inter-frame interpolation's on low motional videos.

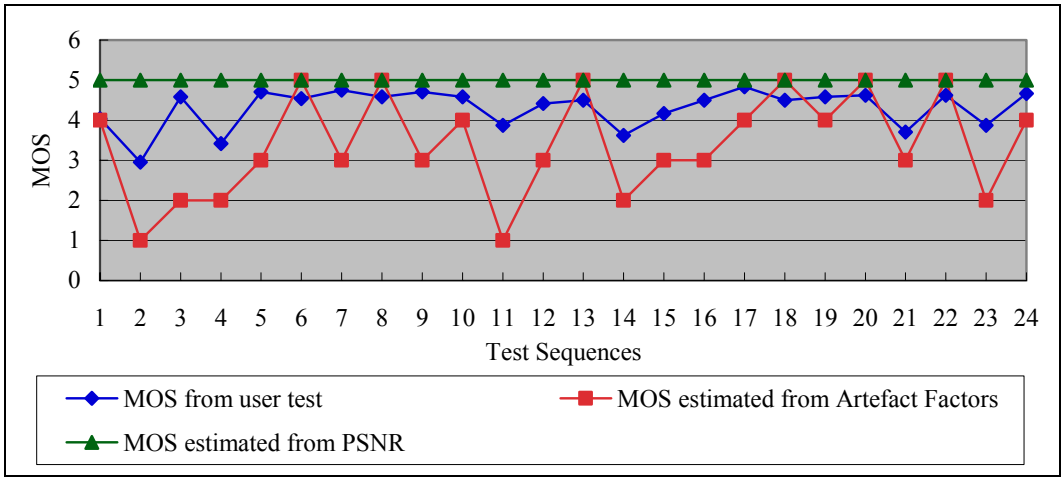

Figure 7 Comparison between MOS from user test results, MOS estimated from artifact factors and MOS estimated from PSNR.

Figure 7 compares between MOS from user test results, MOS estimated from artifact factors and MOS estimated from PSNR on all the 24 corrupted test sequences. As a whole, it is concluded from the previous discussion and from Figure 7 that video content characteristics, encoding/decoding configurations and frame types do have great impact on visibility of packet loss on the testing video sequences. Furthermore, traditional network QoS and PSNR-MOS mapping method cannot represent the user opinion in our experiment while artifact factor-based assessment method follows the trend of user test results except for few special cases.

7. CONCLUSION

In this paper, the impact of individual packet loss on four videos representing different content types encoded with H.264 main-profile video streams has been studied. Four factors (SPIC, TPDR, SPPR and SPXT) are defined to model the level of artifacts in video frames. Video content characteristics, the encoding scheme and the error concealment, which affect the visibility of artifacts, are then investigated in terms of how they contribute to the four artifact factors. To establish a correlation between artifact factors and user experiences, user tests have been conducted and user opinions on deteriorated videos have been collected. The test results verified the expected user perceived quality scores, which are estimated from the joint impacts of artifact factors. PSNR-based MOS estimation was also performed to compare with the proposed artifact factor-based method. The test results proved that traditional network QoS parameters such as packet loss rate are not able to perform accurate quality assessment in certain scenarios as different individual packet loss has a

significantly different impact on user perceived video quality. Our results also show that PSNR-MOS mapping cannot represent the user opinion in our experiment while artifact factor-based method follows the trend of user test results.

The artifact factors were estimated in a three level manner (high, medium or low). In our future work, image/video signal processing will be employed to quantify artifact factors within the video frame in order to establish a mathematical artifact factor model through statistical analysis. With this artifact factor model, higher performance can be achieved on perceived video quality assessment. Furthermore, the SPXT which is one of the four artifact factors is fixed in our experiment. The impact of a variable SPXT will be explored in future study. An accumulation function will also be designed to combine the impacts of individual packet losses within the video content to assess video content over long time period. The research results from this paper and our future works will contribute to a non-reference or reduced-reference objective video quality assessment service in the video content distribution network.

8. ACKNOWLEDGEMENT

The work presented in this paper is supported by the European Commission, under the Grant No.FP6-0384239 (Network of Excellence CONTENT) and Agilent Laboratories UK.

9. REFERENCES

[1] Romaniak, P., Mu, M., Mauthe, A., D'Antonio, S., and Leszczuk, M., "Framework for the Integrated Video Quality Assessment," in *18th ITC Specialist Seminar on Quality of Experience*, Blekinge Institute of Technology, Karlskrona, Sweden, 2008.

[2] Kwon, S.-k., Tamhankar, A., and Rao, K. R., "Overview of H.264/MPEG-4 part 10," *Journal of Visual Communication and Image Representation,* vol. 17, 2006.

[3] Flierl, M. and Girod, B., "Generalized B pictures and the draft H.264/AVC video-compression standard," *IEEE Transactions on Circuits and Systems for Video Technology, ,* vol. 13, pp. 587-597, July 2003.

[4] Stockhammer, T., Hannuksela, M. M., and Wiegand, T., "H.264/AVC in Wireless Environments," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,* 2003.

[5] Hannuksela, M. M., "Non-normative error concealment algorithms," *ITU-T VCEG-N62,* 2001.

[6] Ye-Kui, W., Hannuksela M., M., Varsa, V., Hourunranta, A., and Gabbouj, M., "The Error Concealment Feature In The H.26L Test Model " in *Proc. ICIP*, 2002.

[7] Lopez, D., Gonzalez, F., Bellido, L., and Alonso, A., "Adaptive multimedia streaming over IP based on customer oriented metrics," in *2006 International Symposium on Computer Networks*, 2006.

[8] Boyce, J. M. and Gaglianello, R. D., "Packet loss effects on MPEG video sent over the public Internet," in *Proceedings of the sixth ACM international conference on Multimedia* Bristol, United Kingdom ACM Press, 1998.

[9] Verscheure, O., Frossard, P., and Hamdi, M., "User-oriented QoS Analysis in MPEG-2 Video Delivery," *Real-Time Imaging,* 1999.

[10] Van den Branden Lambrecht, C. J. and Verscheure, O., "Perceptual quality measure using a spatiotemporal model of the human visual system," *Digital Video Compression: Algorithms and Technologies,* 1996.

[11] Van den Branden Lambrecht, C. J., "A working spatio-temporal model of the human visual system forimage restoration and quality assessment applications," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1996.

[12] Reibman, A. R., Kanumuri, S., Vaishampayan, V., and Cosman, P. C., "Visibility of individual packet losses in MPEG-2 video," in *IEEE ICIP*, 2004.

[13] "Video Quality Experts Group," http://www.its.bldrdoc.gov/vqeg.

[14] "Subjective assessment of standard definition digital television (SDTV) systems," *ITU-R BT.1129-2,* 1998.

[15] "Methodology for the subjective assessment of the quality of television pictures," *ITU Recommendation BT.500-11*.

[16] "H.264/AVC JM Reference Software," http://iphome.hhi.de/suehring/tml/.

[17] Winkler, S., *Digital Video Quality: Vision Models and Metrics*: Wiley, 2005.

[18] Ohm, J.-R., "Bildsignalverarbeitung fuer multimedia-systeme," 1999.