User perception of adapting video quality
Nicola Cranley*, Philip Perry, Liam Murphy

Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland

Received 9 November 2004; received in revised form 18 August 2005; accepted 20 December 2005
Available online 7 February 2006
Communicated by J. Taylor

Abstract

In general, video quality adaptation and video quality evaluation are distinct activities. Most adaptive delivery mechanisms for streaming multimedia content do not explicitly consider user-perceived quality when making adaptation decisions. Equally, video quality evaluation techniques are not designed to evaluate instantaneous quality where the quality is changing over time. We propose that an Optimal Adaptation Trajectory (OAT) through the set of possible encoding exists, and that it indicates how to adapt encoding quality in response to changes in network conditions in order to maximize user-perceived quality. The subjective and objective tests carried out to find such trajectories for a number of different MPEG-4 video clips are described. Experimental subjective testing results are presented that demonstrate the dynamic nature of user perception with adapting multimedia. The results demonstrate that adaptation using the OAT out-performs conventional adaptation strategies in which only a single aspect of the video quality is adapted. In contrast, the OAT provides a mechanism to adapt multiple aspects of the video quality thereby giving better user-perceived quality in both the short and long term.

Keywords: Video quality; Evaluation/methodology; Perception; Subjective and objective quality; Adaptation

1. Introduction

Best-effort IP networks are unreliable and unpredictable, particularly in wireless networks. There can be many factors that affect the quality of service (QoS). QoS is a general term that measures the performance of a transmission system via parameters that reflect its transmission quality, such as delay, loss and jitter. Congested network conditions result in lost video packets, which, as a consequence, produces poor quality video. Further, there are strict delay constraints imposed by streamed multimedia traffic. If a video packet does not arrive before its playout time, the packet is effectively lost. Packet losses have a particularly devastating effect on the smooth continuous playout of a video sequence due to inter-frame dependencies. A slightly degraded quality but uncorrupted video stream is less irritating to the user than a corrupted stream. However, rapidly fluctuating quality should also be avoided as the Human Vision System (HVS) adapts to a specific quality after a few seconds and it becomes annoying if the viewer has to adjust to a varying quality over short time scales (Ghinea et al., 1999). Controlled video quality adaptation is needed to reduce the negative effects of congestion on the stream whilst providing the highest possible level of service and quality.

One possible approach to the problem of network congestion and resulting packet loss and delay is to use feedback mechanisms to adapt the output bit rate of the encoders, which, in turn adapts the video quality, based on implicit or explicit information received about the state of the network. Several bit rate control mechanisms based on feedback have been presented in the last few years. As the Real-Time Control Protocol (RTCP) provides network-level QoS monitoring and congestion control information such as packet loss, round trip delay, and jitter. Many applications use RTCP to provide control mechanisms for transmission of video over IP networks (Schulzrinne et al., 1996). However, the network-level QoS parameters provided by RTCP are
not video content-based and it is difficult to gauge the quality of the received video stream from this feedback.

In the past few years, there has been much work on video quality adaptation and video quality evaluation. In general, video quality adaptation indicates how the bit rate of the video should be adjusted in response to changing network conditions. However, this is not addressed in terms of video quality, as for a given bit rate budget there are many ways in which the video quality can be adapted. Video quality evaluation measures the quality of video as perceived by the users, but is not designed for adaptive video transmissions.

Previously we have proposed that there is an optimal way in which multimedia transmissions should be adapted in response to network conditions to maximize the user-perceived quality. Our work is based on the hypothesis that within the set of different ways to achieve a target bit rate, there exists an encoding that maximizes the user-perceived quality. Extensive subjective testing suggests that an Optimal Adaptation Trajectory (OAT) in the space of possible encoding does exist and that it is related to the content type. We propose that knowledge of these OATs can be used as part of an adaptation strategy, which aims to maximize the user-perceived quality of the delivered multimedia content. A key issue addressed in this paper is how to adapt in order to maximize the user-perceived quality. This raises the question of how user-perceived quality can be assessed in practice. In this paper, we explore the possibility of applying objective metrics to discover the OAT and compare it to those found through extensive subjective testing. Then we evaluate the user perception of adapting video quality. We compare the perception of adapting video quality where the video is adapted within a two-dimensional adaptation space given by the OAT, with the more common approach of video adaptation used by streaming service providers where the video is adapted in a single dimension of adaptation space. Furthermore, we present some results that show how user perception varies in three different scenarios, that is when the quality is being increased, when the quality is being decreased and when the video quality is adapted in a manner that mimics the behaviour of a typical adaptive streaming system.

The rest of the paper is organized as follows: in Section 2, we introduce the concept of OATs and describe a subjective methodology for finding the OAT for a given video content type, and show some results from our subjective testing to illustrate this methodology. In Section 3, we investigate objective metrics and their suitability for finding the OAT comparing the results of objective measurements with subjective test results. In Section 4, we demonstrate how two-dimensional adaptation using the OAT out-performs conventional adaptation techniques. Finally, in Section 5 we present our conclusions and some directions for future work.

2. Optimum adaptation trajectories

The authors have proposed that there is an optimal way in which multimedia transmissions should be adapted in response to network conditions to maximize the user-perceived quality (Cranley et al., 2003). This is based on the hypothesis that within the set of different ways to achieve a target bit rate, there exists an encoding configuration that maximizes the user-perceived quality. If a particular multimedia file has $n$ independent parameters that define the encoding configuration, then there exists an adaptation space with $n$ dimensions where each dimension of adaptation space represents an independent encoding configuration. Adaptation space consists of all possible dimensions of adaptation for the content, which can be implemented as part of an adaptive streaming server or adaptive encoder.

When adapting the transmission from some point within that space to meet a new target bit rate, the adaptive server should select the encoding configuration that maximizes the user-perceived quality for that given bit rate. When the transmission is adjusted across its full range, the locus of these selected encoding configurations should yield an OAT within that adaptation space.

This approach is applicable to any type of multimedia content. The work presented here focuses for concreteness on the adaptation of MPEG-4 video streams (Pereira and Ebrahimi, 2002) within a two-dimensional adaptation space defined by frame rate and spatial resolution. These encoding variables were chosen as they most closely map to the spatial and temporal complexities of the video content. The example shown in Fig. 1(a) indicates that, when degrading the quality from an encoding configuration of 25fps and 100% resolution or $[25_{FP}, 100_{R}]$, there are a number of possibilities such as reducing the frame rate only, $[X_{FP}, 100_{R}]$, reducing the resolution only, $[25_{FP}, Y_{R}]$ or reducing a combination of both parameters, $[U_{FP}, V_{R}]$. Each of these possibilities lies within a zone of equal average bit rate (EABR). The clips falling within a particular zone of EABR have different, but similar bit rates. For example, the bit rates corresponding to the encoding points $[17_{FP}, 100_{R}]$, $[25_{FP}, 79_{R}]$ and $[25_{FP}, 63_{R}]$ were 85, 88, and 82 kbps, respectively. To compare clips of exactly the same bit rate would require a target bit rate to be specified, and then the encoder would use proprietary means to achieve this bit rate by compromising the quality of the encoding in an unknown manner. Using zones of EABR effectively quantizes the bit rate of different video sequences with different encoding configurations. The boundaries of these zones of EABR are represented as linear contours for simplicity, since their actual shape is irrelevant for this scheme.

The OAT indicates how the quality should be adapted (upgraded or downgraded) so as to maximize the user-perceived quality. The OAT may be dependent on the characteristics of the content. There is a content space in which all types of video content exist in terms of spatial and temporal complexity (or detail and action). Every type
of video content within this space can be expanded to an adaptation space as shown in Fig. 1(b).

2.1. OAT discovery by subjective testing

User perception of video quality may vary with the content type; for example, viewers may perceive action clips differently from slow moving clips. Thus, there may exist a different OAT for different types of content based on their spatial and temporal characteristics. In order to characterize content in terms of its spatial and temporal complexity, a spatial–temporal grid was constructed. The spatial and temporal perceptual information of the content was determined using the metrics spatial information (SI) and temporal information (TI) (ITU-T Recommendation P.910, 1999; Cranley et al., 2004a, b). Eight different content types were sampled covering as much of the spatial–temporal grid as possible. Each of the sampled content taken from this grid was then encoded to form an adaptation space.

The test sequences were acquired from the Video Quality Experts Group (VQEG) in YUV format (VQEG, 2004). These were then converted to a reference test sequence with MPEG-4 encoding, using the most accurate best quality compression at 25 fps and QCIF (176 × 144) display size. This reference test sequence was then used to obtain samples of the adaptation space with various combinations of spatial resolution and frame rate. During the preparation of the test sequences for the subjective testing, the encoding method used was the “most accurate”; no target bit rate was specified, and the encoder followed the supplied encoding parameters as closely as possible regardless of the resulting bit rate.

The subjective tests were conducted in two phases. Phase 1 considered 4 test sequences, one taken from each quadrant of the SI–TI grid. To facilitate subjective testing and reduce the number of test cases, adaptation space was sampled using a logarithmic scale to reflect Weber’s Law of Just Noticeable Difference (Efford, 2000; Levine, 2000). The frame rates tested were {5, 7, 11, 17, 25} fps and the spatial resolutions were \{100, 79, 63, 50, 40\} R. One hundred and twenty participants were tested during Phase 1. Phase 2 considered 4 different test sequences with similar SI–TI values to those used for Phase 1. However, this time, the adaptation space was sampled using a linear scale. The frame rates tested were \{5, 10, 15, 20, 25\} fps and the spatial resolutions were \{100, 85, 70, 55, 40\} R. Forty participants were tested during Phase 2. The main objective of having two different test phases was to verify and validate the results from Phase 1. In addition, using different encoding scales, it could be ascertained that the OAT was similar in shape regardless of whether a linear or logarithmic scale was used, and regardless of the encoding points tested.

There are a number of different subjective test methodologies that can be used for testing user perception of video quality. These include the absolute category rating (ACR), degraded category rating (DCR) and pair comparison (PC) (ITU-T Recommendation P.910, 1999; ITU-T Recommendation J.143, 2000). To find the OAT we employed the Forced Choice method in which the participant is presented with a number of spatial or temporal alternatives in each trial. The advantage of using the forced choice method is that the bias is binary, which simplifies the grading procedure and allows for reliability, verification and validation of the results. In contrast, the other test methodologies typically use a 5-point grading scale and as a consequence there is more ambiguity in the grading procedure as different participants may interpret the grading scale in different ways (Watson and Sasse, 1996, 1998, Bouch et al., 2001).

Our subjective test consisted of the participant watching every combination of pairs of clips from each EABR zone and making a forced choice of the better quality or preferred encoding configuration. By using zones of EABR, the bit rate of different test sequences with different encoding configurations is effectively quantized which in turn dramatically reduces the number of test cases. To further reduce the number of test cases, it was decided that comparing video clips in a zone of EABR with equal frame rates but different resolutions was redundant as it was considered counter-intuitive that a user would prefer a clip with a lower spatial resolution when all other encoding parameters are the same. For example, comparing the encoding configurations \[25_{\text{FPS}}, 79_{\text{R}}\] and \[25_{\text{FPS}}, 63_{\text{R}}\], it is expected that the user would prefer the video clip with the encoding configuration \[25_{\text{FPS}}, 79_{\text{R}}\]. Similarly, comparing
video clips with equal resolution but differing frame rates were also removed. Intra-reliability and inter-reliability of a participant were factored into our test procedure by including repetition of the same test sequence presentation.

The diagram in Fig. 2 shows the subjective test results obtained for a particular content type. The diagram consists of a grid of circular encoding points where the frame rate is on the x-axis and the resolution is on the y-axis. The diagonal thin grey lines denote the zones of EABR, ranging from 100 to 25 kbps. The encoding points marked with a percentage preference value are those points that were tested within a zone of EABR. For example, in EABR-100 kbps, there were two encoding configurations tested, [17FPS, 100R] and [25FPS, 79R]. Seventy percent of the subjects preferred encoding configuration [17FPS, 100R] while the remaining 30% preferred encoding configuration [25FPS, 79R]. However, in the leftmost zone of EABR the preferred encoding configuration is [5FPS, 63R]. In this zone of EABR there are three encoding configurations but since the frame rate is the same, the preferred encoding configuration is that with the highest resolution [5FPS, 63R].

The Path of Maximum Preference is the path through the zones of EABR joining the encoding configurations with the maximum user preference or the highest percentage preference. Weighted points were then used to obtain the Optimal Adaptation Perception Points (OAP). The weighted points were interpolated as the sum of the product of preference with encoding configuration. For example in the EABR-100 kbps, 70% of subjects preferred encoding [17FPS, 100R] and 30% preferred encoding point [25FPS, 79R]. The weighted vector of these two encoding configurations is [70%(17FPS) + 30%(25FPS), 70%(100R) + 30%(79R)] which equals OAP point [19.4FPS, 93.7R]. The Weighted Path of Preference is the path joining the OAPs. There are two possible paths, which can be used to represent the OAT, the path of maximum user preference and the weighted path of preference. It seems likely that by using the weighted path of preference, the system can satisfy more users by providing a smooth graceful quality adaptation trajectory. Using the same subjective testing methodology, the OAPs in each zone of EABR were compared against the maximum preferred encoding and all other encoding configurations in the zone of EABR. In all cases, the interpolated OAP did not have a statistically significant preference from the maximum preferred encoding indicating that this simple weighted vector approach is acceptable. It was also observed that there was a higher incidence of forced choices when the maximum preferred encoding and the OAP were close together.

Fig. 3 shows the paths of maximum preference and weighted paths of preference for the four content types used during phase one of testing. The content type C1: Football is a fast action highly detailed segment from an American football match. The content type C2: Canoe is a fast action but less detailed clip of a man in a canoeing competition. The content type C3: Susie is a low action low detailed clip of a woman listening to someone on the phone. The content type C4: WashDC is a slow moving highly detailed panoramic shot of the Washington DC skyline. It can be clearly seen from the paths of maximum user preference that when there is high action (C1 and C2), the resolution is less dominant regardless of whether the clip has high spatial characteristics or not. This implies that the user is more sensitive to continuous motion when there is high temporal information in the video content. Intuitively this makes sense as, when there is high action in a scene, often the scene changes are too fast for the user to be able to assimilate the scene detail. Conversely, when the scene has low temporal requirements (C3 and C4), the

![Fig. 2. Subjective test results for content type, C3.](image)

![Fig. 3. Path of maximum user preference and weighted path of preference for four different content types.](image)
resolution becomes more dominant regardless of the spatial characteristics.

3. OAT discovery using objective metrics

In this section, the possibility of discovering the OAT using objective metrics is investigated. Several objective metrics of video quality have been proposed, but they are limited and not satisfactory in quantifying human perception (Winkler, 2000; Yu and Wu, 2000; Masry and Hemami, 2002). It is well known that a commonly used metric, the signal-to-noise ratio (SNR), is not correlated with human vision (van den Branden Lambrecht and Verscheure, 1996; Verscheure et al., 1998). The most commonly used objective metric of video quality assessment is the peak signal-to-noise ratio (PSNR), which has been widely used in many applications and adaptation algorithms (Kim et al., 2003; Wang et al., 2003) to assess video quality. The advantage of PSNR is that it is very easy to compute using the mean square error (MSE) of pixel values of luminance for frames from the degraded and reference clips. However, PSNR does not match well to the characteristics of the HVS. The main problem with using PSNR values as a quality assessment method is that even though two images may be different, the visibility of this difference is not considered. The PSNR metric does not take into consideration any details of the HVS such as its ability to “mask” errors that are not significant to the human comprehension of the image. For example, consider an image where the pixel values have been altered slightly over the entire image and another image where there is a concentrated alteration in a small part of the image: both may have the same MSE value, but they appear to be very different to the user.

The ITU-T has recently accepted the video quality metric (VQM) as a recommended objective VQM that correlates adequately to human perception (ITS, 2002; ITU-T Recommendation J.149, 2004; ITU-T Recommendation J.148, 2003). However, in recommendation ITU-T Recommendation J.144 (2001) the ITU state that subjective testing is required to complement objective metrics but also, more importantly, that objective metrics are not a direct replacement for subjective testing. The VQM from National Telecommunications and Information Administration (NTIA) has a freely available software implementation, which performs the sophisticated VQM analysis and calculations (Wolf and Pinson, 2002).

The VQM suite of metrics consists of objective metrics, each designed for a different target application. In this section we test the VQM, VQMPSNR, VQMGeneral, VQMDeveloper and VQMV-Conf. The methodology for discovering the OAT using objective metrics compares a degraded version of each clip with the reference version of that clip. The reference clip corresponds to the highest-quality clip at 25 fps and 100% spatial resolution. Each of the degraded clips corresponded to a point in adaptation space. The encoding configuration of the clip within each zone of EABR with the highest-quality value was chosen as the encoding configuration that should be adopted when making a quality adaptation. The Path of Maximum Objective Quality is the path through the encoding configurations in each zone of EABR that yields the highest quality as measured by the objective quality metric.

The adaptation paths determined using the VQM are shown in Fig. 4. The VQM have a range 0–1, where 0 is the best quality, that is, with no impairment. For convenience the VQM quality is converted into a percentage, 0–100%, where 100% is the highest quality. It can be seen that the proposed path tends to maintain the frame rate in preference to the spatial resolution. As lower bit rates are required, the measured path through the discrete space moves quite erratically and it is not clear that this path could be mapped to a sensible path through a continuous adaptation space.

This behaviour may be due to the temporal alignment constraints during frame-by-frame analysis of clips with differing frame rates. Frames are aligned either by matching frames by their play-out time or by applying a best-match frame algorithm during the calibration phase of the analysis. In addition, most objective metrics operate by comparing pixel values in consecutive frames for a particular time segment of the clip. If the reference clip and the degraded clip have different frame rates, this analysis results in a lower-quality measurement. Thus, the path of maximum objective quality will have a tendency to maintain the frame rate.

The results from the subjective tests on same test clip are shown in Fig. 5. Comparing the path of maximum preference discovered by subjective testing with the path of maximum objective quality discovered using objective

![Fig. 4. Paths of maximum objective quality using the VQM for content type, C1: (a) VQM dev, (b) VQM general, (c) VQM PSNR and (d) VQM V-conf.](image-url)
metrics indicates that objective metrics do not correspond to how users perceive adapting video. The correlation between the path of maximum quality using the VQM and the path of maximum preference obtained by subjective testing are shown in the legend of Fig. 5. The poor correlation is due to the fact that these metrics are designed for comparing clips with impairments incurred in the transmission system, rather than evaluating the quality of video sequence where the quality is adapting in the dimensions of frame rate and/or resolution with the goal to yield an adaptation strategy. The results demonstrate that there is a significant difference between the adaptation trajectories yielded using objective metrics and subjective testing techniques.

4. OATs in practice

Streaming multimedia over best-effort networks and wireless networks is becoming an increasingly important source of revenue. A content provider is unlikely to have the resources to provide real-time adaptive encoding for each unicast request and as such reserves this for “live” adaptive multicast sessions only. Typically, pre-encoded content is transmitted by unicast streams where the client chooses the connection that most closely matches their requirements. For such unicast sessions, the adaptive streaming server can employ several techniques to adapt the pre-encoded content to match the clients’ resources. In such adaptive streaming systems, two techniques that are most commonly used are frame dropping (Krasic et al., 2003) and stream switching (Conklin et al., 2001). The OAT shows how to adaptively stream the video in order to maximize the user’s perceived quality in a two-dimensional adaptation space defined by frame rate and resolution (Fig. 6). The OAT has been used as a basis of a new perception-based adaptation algorithm (Cranley et al., 2004a). Adaptive frame rate is achieved by frame dropping and adapting spatial resolution is achieved using track or stream switching. In this section, we shall demonstrate that two-dimensional adaptation using the OAT out-performs adaptation in a single dimension defined by resolution (by track switching) and frame rate (by frame dropping). Of particular interest is the performance of the OAT over other schemes when the available bit rate is very low and the resulting video quality must be reduced to a low quality in order to achieve this low bit rate.

All video adaptation algorithms exhibit some form of increase and decrease behaviour. When there is no congestion, the server increases its transmission rate and when congestion occurs the server decreases its transmission rate. Such rate changes can take place in either an additive, proportional or multiplicative fashion. For example,

- **AIMD**: Consists of an additive increase (AI) and multiplicative decrease (MD).
- **AIAD**: Consists of an AI and additive decrease (AD).

Several variants such as additive increase/proportional decrease (AIPD) (Venkitaraman et al., 1999), multiplicative increase/multiplicative decrease (MIMD) (Turletti and Huitema, 1996) have been proposed in literature. Most adaptation algorithms have some form of AI and AD behaviour whilst a majority of these adaptation algorithms are AIMD (Feamster et al., 2001). Thus, these three popular adaptation schemes were investigated to determine the usefulness of the OAT in the context of a typical multimedia adaptation system. The first assesses user perception when quality is adapted up in an AI manner whilst the second assesses perception when quality is degraded down in an AD manner. The third assesses quality adapting in an AIMD manner.

4.1. Test methodology

The Forced Choice methodology is suitable for clips lasting not longer than 15 s (Watson and Sasse, 1996;
de Ridder and Hamberg, 1997). For video clips lasting longer than this duration, recency and forgiveness effects can influence the participant, which can have a significant impact when the participant must grade the overall quality of a video sequence. For example, the participant may forget and/or forgive random appearances of content-dependent artifacts when they are making their overall grade of the video sequence. To test clips of a longer duration a different test methodology to the forced choice method needs to be applied to overcome the forgiveness and recency effects and to ensure the participant can make an accurate judgment.

The single-stimulus continuous quality evaluation (SSCQE) methodology is intended for the presentation of sequences lasting several minutes (ITU-R Recommendation BT.500-7, 1999). Continuous evaluation is performed using a slider scale on the PC screen to record the participants’ responses without introducing too much interference or distraction, and provides a trace of the overall quality of the sequence (Alpert and Evain, 1997; Pinson and Wolf, 2003). A reference clip was played out at the beginning of the test so that the participants were aware of the highest-quality sequence. The three varying quality sequences were then presented in random order to each participant in turn. As each sequence was played out the participant continuously rated the quality of the sequence using the slider. When the slider is moved, the quality grade of the slider is captured and related to the play-out time of the media. A total of 27 participants were tested for this experiment. During each test, the media time and grade were written to a file. For each clip, the results were aggregated for all participants. The mean opinion score (MOS) and standard deviation are calculated at each media time instant. In this case, each media time instant corresponds to one second of media. The MOS and standard deviation is calculated for each clip segment.

4.2 Test sequence preparation

The test sequence chosen for this experiment contains a wide range of spatial and temporal complexity. The test sequence contains periods of high temporal complexity that are generally bursty containing many scene changes. In this test sequence, periods of high temporal complexity are generally followed by periods of relatively low temporal complexity but high spatial complexity consisting of detailed scenes such as facial close-ups and panoramic views. This test sequence contains a broad diversity of complexity and is typical of film trailers. The test sequence was divided into segments of 15 s duration, and each segment was encoded at various combinations of spatial resolution and frame rate. These video segments were then pieced together seamlessly to produce three bit rate-varying versions of the test sequence. It was necessary to control and align the times of adaptation events for each of the test sequences used. In the work presented here, we have assumed that some mechanism is implemented that informs the streaming server of the required transmission bit rate.

4.3 Results

Three different adaptation scenarios were tested. In the first scenario the quality is progressively adapted down from the best quality to lowest quality whilst in the second the quality is adapted up from the lowest quality to the best quality. These two tests are complementary and are designed to assess the symmetry of perception, that is, whether participants perceive quality increases and quality decreases in the same way. In the third scenario the quality is varied in an AIMD fashion. This test is designed to mimic the behaviour of a typical adaptive streaming server.

4.3.1 AD quality adaptation

In this test, the quality of the clip degrades from the best quality to the worst quality. Fig. 7(a) shows the bit rate decreasing as the quality degrades. Fig. 7(b) shows the encoding configuration of frame rate and resolution for each segment as the quality is adapting down in either the frame rate dimension only or the resolution dimension only or using the OAT adapting down in both the frame rate and resolution dimensions. Through time interval 0–45 s the resolution and frame rate dimensions are perceived the same (Fig. 7(c)). In time interval 45–60 s, there appears to be imperceptibility when the resolution is decreased from 80% to 70%. Using the OAT there is a smooth decrease in the MOS scores, which outperforms both one-dimensional adaptation of frame rate and resolution. During time interval 45–60 s, there is high action in the content, which may explain the sharp decrease in the MOS scores for adapting the frame rate only. When there is high action, participants prefer smooth continuous motion. Further, when there is high action content, reductions in spatial resolution cannot be perceived as clearly as there is too much happening in the video clip for the detail to be perceived properly. Fig. 7(d) shows a close up view of MOS scores during the lowest quality level in time interval 70–90 s, the frame rate is perceived worst of all whilst the resolution performs very well. This may be due to the fact that the bit rate for the resolution is significantly greater than the two other methods.

4.3.2 AI quality adaptation

In this test, the quality of the clip upgrades from the worst quality to the best quality. Fig. 8(b) shows the encoding configuration of frame rate and resolution as the quality is adapting up in either the frame rate dimension only or the resolution dimension only or using the OAT adapting down in both the frame rate and resolution dimensions. During this experiment, the slider is placed at the highest-quality value on the rating scale when the clip begins. It can be seen that it took participants several seconds to react to the quality level and adjust the slider to the appropriate value (Fig. 8(c)). At low quality,
Fig. 7. Time series during additive decrease (AD) in quality: (a) segment average bit rate variations over time, (b) video-encoding parameter variations over time, (c) MOS scores over time and (d) MOS scores during period of lowest-quality adaptation in Segment 6.

Fig. 8. Time series during additive increase (AI) in quality: (a) segment average bit rate variations over time, (b) video encoding parameter variations over time, (c) MOS Scores over time and (d) MOS Scores during period of lowest-quality adaptation in Segment 2.
participants perceive adaptation using the OAT better than one-dimensional adaptation. The quality is slowly increasing however participants do not seem to notice the quality increasing nor do they perceive it significantly differently indicating that participants are more aware of quality when the quality is low (Fig. 8(d)).

4.3.3. AIMD quality adaptation

This section presents the results for AIMD adaptation. The behaviour of this is modelled as a typical AIMD process as might be expected from a TCP-friendly rate control mechanism. The same bit rate variation patterns were obtained in these three sequences by adapting quality in the frame rate dimension only, the spatial resolution dimension only, or both frame rate and spatial resolution dimensions, as shown in Fig. 9(a). The traces in Figs. 9(b) show the encoding configuration of frame rate or resolution for each segment as the quality was adapted in either the frame rate dimension only, or the resolution dimension only, or using the OAT adapting in both the frame rate and resolution dimensions.

In Fig. 9(a), it can be seen that although the first bit-rate reduction occurs at time 15 s, it is not fully perceived until time 28 s because there is a time delay for participants to react to the quality adaptation. At time interval 70–90 s, a larger drop in bit rate occurs resulting in the lowest-quality level that might reflect a mobile user entering a building. The MOS scores for adapting only the frame rate and spatial resolution are quick to reflect this drop. However, using the OAT, it takes participants much longer to perceive this drop in quality. This is a high action part of the sequence and so the reduced frame rate is perceived more severely. The standard deviation of MOS scores using the OAT was much less than that for adapting frame rate only or spatial resolution only.

4.4. Discussion

From the experiments reported here, it appears that if a user’s average bit rate changes from being quite near their maximum to near the minimum that they can tolerate, then a one-dimensional adaptation policy will cause the perceived quality to degrade quite severely. Using the two-dimensional adaptation strategy given by the OAT allows the bit rate to be dropped quite dramatically but maintain substantially better user-perceived quality. Table 1 below summarizes the MOS percentage scores for each of the experiments conducted.

In addition to the greater bit rate adaptation range achieved using the OAT, adaptation using the two-dimensional OAT out-performs one-dimensional adaptation and reduces the variance of perception. From the various experiments conducted, participants perceived adapting frame rate the worst, followed by resolution.
and the OAT best of all. We observed that there is a time delay of several seconds for participants to react to quality adaptations. We also observed that quality perception is asymmetrical when adapting the quality down and adapting quality up, which is users, are more critical of degradations in quality and less rewarding of increased quality. Both these observations were reported in (Pinson and Wolf, 2003). Perception is strongly dependent on the spatio-temporal characteristics of the content. Given this understanding of user perception, adaptation algorithms should consider the contents characteristics when making adaptation decisions. Also, frequent quality adaptation should be avoided to allow the users to become familiar with the video quality. In the experiments, the globally averaged OAT was used, but the OAT can be dynamic if the contents spatial and temporal characteristics are known at a given instant, thus making it more flexible to adapt according to the contents’ characteristics and maximize user-perceived quality. It is expected that a dynamic OAT that adapted on the changing complexity of the content would yield even higher MOS scores.

5. Conclusions

A majority of video adaptation algorithms indicate how the bit rate of the videostream should be adjusted in response to changing network conditions. However, this is not addressed in terms of video quality, as for a given bit rate budget there are many ways in which the video quality can be adapted nor does it take into consideration the user-perceived quality. The problem this paper addresses is how to adapt video quality in terms of video encoding parameters and user-perceived quality for streamed video over best-effort IP networks. With a better understanding of user perception, this knowledge can be integrated into video quality adaptation techniques. A key issue addressed by this paper is how to adapt in order to maximize the resulting user-perceived quality.

We have shown that when adapting the quality of a stream in response to network conditions, there is an Optimal Adaptation Trajectory (OAT) that maximizes the user-perceived quality. More specifically, within the set of different ways to achieve a target bit rate, there exists an encoding that maximizes the user-perceived quality. We have described a subjective methodology to discover the OATs through subjective testing, and applied it to finding OATs for various MPEG-4 video clips. Further it was shown that by using a two-dimensional adaptation strategy given by the OAT allows the bit rate to be dropped quite dramatically but maintain substantially better user-perceived quality over one-dimensional adaptation strategies. Adaptation in two-dimensions using the OAT out-performs one-dimensional adaptation where only the frame rate or spatial resolution are adapted and reduces the variance of perception.

Objective metrics were investigated to determine whether they yielded an OAT that correlated to that discovered using subjective testing. The results show that there is a significant difference between the adaptation trajectories yielded using objective metrics and subjective testing techniques. This suggests that measuring quality and adapting quality based on this measurement are different tasks.

Currently work is in progress to assess the possibility of using and/or modifying existing objective metrics in order to mimic the OATs found by subjective methods and enable the development of a dynamic OAT. Some preliminary results have been presented in Cranley et al., 2004b. This involves a greater analysis of the relationship between content characteristics and the corresponding OAT to determine the sensitivity of an OAT to the particular video being transmitted.

Acknowledgements

The support of the Research Innovation Fund and Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

References
