The Effects of Jitter on the Perceptual Quality of Video

Mark Claypool
Jonathan Tanner
{claypool,jtanner}@cs.wpi.edu
Computer Science Department
Worcester Polytechnic Institute, USA *

Abstract

Today's powerful computers and networks present the opportunity for video across the Internet right to the desktop. However, Internet video often suffers from packet loss and jitter, degrading the user's perceived quality of the video. Understanding the effects of delay, loss and jitter on media quality is critical for choosing delay buffer sizes and packet repair techniques. While the effects of packet loss on perceptual quality are well-understood, to date there have not been careful user studies measuring the impact of jitter on perceptual quality. The major contributions of this work are carefully designed experiments that measure and compare the impact of both jitter and packet loss on perceptual quality of packet video. We find that jitter degrades perceptual quality nearly as much as does packet loss, and that perceptual quality degrades sharply even with low levels of jitter or packet loss as compared to perceptual quality for perfect video.

1 Introduction

The power of today's computers and the connectivity of today's networks present the opportunity of packet video over the Internet and to the desktop. Desktop video over a local area network is often of acceptable quality, providing the user with a smooth stream of video frames. Video on the Internet, however, is usually quite poor, having a jerky appearance from jitter and corrupt or missing frames from lost packets [CR99].

In the absence of jitter and packet loss, video frames can be played as they are received, resulting in a smooth playout, as depicted in Figure 1-top. However, in the presence of jitter, interarrival times will vary, as depicted in Figure 1-middle. In Figure 1-middle, the third frame arrives late at r*2*. In this scenario, the user would see the frozen image of the most recently delivered frame (frame two) until the tardy frame (frame three) arrived. The tardy frame (frame three) would then be played only briefly in order to preserve the timing for the subsequent frame (frame four).

*This paper appears in proceedings of *ACM Multimedia* as a poster paper, October 30 - November 5, 1999 in Orlando, FL, USA.

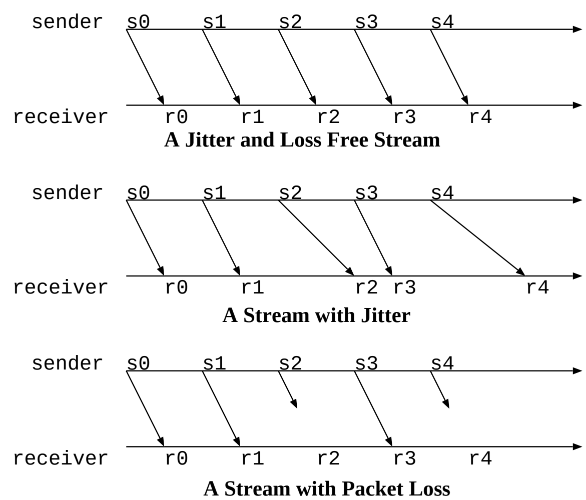

Figure 1: The above figures model packet video between sender and receiver. Each s*i* is the time at which the sender transmits video frame *i*. Each r*i* is the time at which the receiver receives frame *i*.

In the presence of packet loss, some frames will not even arrive at the receiver, as depicted in Figure 1-bottom. In Figure 1-bottom the third and fifth frames do not arrive at the receiver. In the case of the loss of frame three, the viewer would see a frozen image of the most recently delivered frame (frame two), and the video stream would then jump to the next frame that arrived (frame four).

Methods to ameliorate the effects of jitter have been explored by many researchers [SJ95, RKTS94, Fer92] through delay buffering. Delay buffering can compensate for jitter at the expense of latency. Transmitted frames are buffered in memory by the receiver, allowing each frame to be played out with a constant latency, achieving a steadier stream. However, the added latency from buffering can be disturbing especially for interactive applications. Ideally, the buffer size is chosen to give the best *perceptual quality*[1] by understanding the tradeoff between decreased jitter and increased latency. However, to the best of our knowledge, current delay buffering algorithms do not choose their buffer sizes with any consideration of the impact of residual jitter on perceptual quality. Our detailed study of the impact of jitter on perceptual quality can be used to develop more effective delay buffering algorithms.

[1] We use the term *perceptual quality* to mean a quantitative measure of video quality from the user perspective.

Methods to repair lost packets have been explored by many researchers, and include using interleaving or redundancy [PHH98] and retransmission [CPW98, PVM98, XMZY97]. Unfortunately, these repair techniques increase jitter. For example, retransmitting a lost packet at least doubles the delay for the re-sent packet, resulting in a jitter 'spike'. If lost packets are found to degrade perceptual quality more than the extra jitter added by a repair technique, then the lost packets should be repaired. If, however, jitter decreases perceptual quality more than does packet loss, then packets should not be repaired. To the best of our knowledge, current packet repair techniques do not consider the impact of increased jitter in evaluating their effectiveness. A careful understanding of the effects of jitter and packet loss on perceptual quality can be used to select repair techniques that achieve better perceptual quality.

The effects of delay on a user's perception of video quality is well-understood and well-researched [IKK93, Roy94, DCJ93]. Similarly, there is a clear relationship between packet loss and multimedia quality deterioration [MS94, SW93, HSK98, PHH98, GKL+98]. The effects of framerate on user understanding of video clips has been explored [GT98]. There has also been research done in classifying perceptual quality assessment techniques [WS98]. However, to the best of our knowledge, there has been no measurement of the impact of jitter on the quality of packet video from the user perspective. Similarly, there have not been any careful studies of the relative impact of jitter as compared to packet loss on a user's perspective of video quality. The primary goal of this work is thus to address these two problems.

An additional parameter that may affect perceptual quality is the temporal aspect of a video. A video which has very few differences between most frames, such as a "talking head" in a videoconference or a news broadcast, has a low temporal aspect. Videos with low temporal aspect may not degrade much under jitter or packet loss since viewers may not notice delayed or missing frames. A video which has many differences between frames, such as a sporting event or a music video, has a high temporal aspect. Videos with high temporal aspect may be more sensitive to jitter or packet loss since viewers will be deprived of more information when packets are lost or delayed.

More details of our ongoing work are available in [CT99] and [Tan99]. In this paper, we present:

- a controlled user study which quantitatively evaluates the effects of jitter on perceptual quality,
- a direct comparison of the effects of packet loss on perceptual quality as compared to the effects of jitter on perceptual quality,
- an analysis of the effects of the temporal aspect of a video on perceptual quality in the presence of jitter and packet loss, and
- an experimental model for conducting user studies that measure perceptual quality of video.

The rest of this document is laid out as follows: Section 2 describes experiments that measure perceptual quality with packet loss and with jitter; Section 3 analyzes the results of the experiments; and Section 4 summarizes our conclusions.

2 Experiments

In an effort to measure the effects of jitter and packet loss on perceptual quality, we have designed and conducted experiments to evaluate the same.

To evaluate the perceptual quality of video, we use a *quality opinion score* in which test subjects are asked for an explicit rating after watching a video clip [WS98]. For our quality opinion scores, test subjects entered their evaluation by means of a slider with values ranging from 1 to 1000. The slider had labels on the ends only, indicating *worst* (1) and *best* (1000).

We chose five video clips from each of five different temporal categories: Animated (The Simpsons), Information (Home Shopping Channel), News (CNN), Situation Comedies (News Radio and Married with Children) and Sports (Soccer, Hockey and Jet Skis). These categories were meant to provide video clips ranging from low temporal aspect (Information) to high temporal aspect (Sports). The video clips were captured using a VCR recording in SP mode and an ATI All-In-Wonder video capture card. MainActor video software was used to encode the captured video into MPEG-1 format. The clips were each about 60 seconds long and watching the entire set of clips took each test subject about 1/2 hour.

We induced residual jitter and packet loss into the video clips based on Internet traces from previous work that sent simulated video from locations across the United States and New Zealand to Worcester, Massachusetts [GBC98]. We selected a "low" jitter trace and scaled it three-fold to create a "high" jitter trace. The same traces were then used for inducing both loss and jitter in order to allow for a direct comparison of perceptual quality for the same "amount" of jitter and packet loss. Thus, we had two levels of jitter and two levels of packet loss, resulting in five different quality levels for each video clip: perfect, low jitter, low loss, high jitter or high loss. The low loss trace had about 8% packet loss and the high loss trace had about 22% packet loss.

In addition to jitter and packet loss, a perceptual quality opinion score may be affected by the frame size, color depth or frame rate. As our focus was on the effects of jitter and packet loss on perceptual quality, we kept these other parameters constant. Specifically, we used a `320x240` video frame, true color (24 bits per pixel) and 30 frames per second. In addition, although video is usually accompanied by an audio track, we wanted to isolate the effects of packet loss and jitter on video only, and so all video clips were played without their audio accompaniment.

We had 41 people participate in the study. Before the tests began, we gathered profile information on each person, including age, computer experience and experience with Internet video. Most of our subjects were young students (20-25 years old) having moderate to extensive computer experience, although a few that were much older and worked in non-computer related professions. Most had occasionally watched video on the Internet, but none had frequently done so.

All the test subjects were evenly distributed into one of five test groups. The quality level for each clip was different for each group, as depicted in Figure 2, ensuring an equal number of subjects saw each video clip at each of the five quality levels.

Prior to beginning the test clips, each test subject was

	ski	hockey	cnn	shopping	
Group 1	low jitter	high jitter	low loss	high loss	...
Group 2	high loss	low loss	high jitter	perfect	...
Group 3	low loss	perfect	low jitter	high jitter	...
Group 4	high jitter	high loss	perfect	low loss	...
Group 5	perfect	low loss	high loss	low jitter	...

Figure 2: Test Groups. The test subjects were divided evenly into among five groups, depicted above. The quality levels were then varied for each of the movie clips such that each group saw each clip at one of the five tested quality levels.

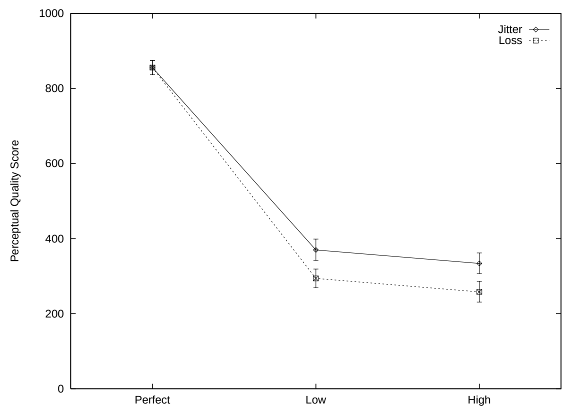

Figure 3: Perceptual Quality versus Jitter or Packet Loss. There are five quality levels depicted: perfect, low loss, low jitter, high loss and high jitter. The horizontal axis indicates perfect, low or high. The vertical axis indicates the average perceptual quality opinion score given for all test subjects. All points are shown with 95% confidence intervals.

primed with a medium temporal aspect video clip at perfect quality in order to set the expectations of all test subjects to the same level before the actual recording of perceptual quality scores. The procedure was then: play a clip, have the user record a quality opinion score, repeat for each of the 25 clips.

3 Analysis

In this section, we analyze the results of the experiments described in Section 2. We examine the effects of jitter and packet loss on perceptual quality and the influence of temporal aspect of video on perceptual quality.

3.1 Perceptual Quality versus Jitter or Packet Loss

Figure 3 depicts the quality opinion scores for the five quality levels tested: perfect, low jitter, low loss, high jitter and high loss. Each point is shown with a 95% confidence interval. The average "perfect" quality scores are around 850 instead of the maximum 1000, possibly because of the small window size used in the study or because the lack of audio caused the test subjects to feel there was room for improvement in the video clips.

Both low jitter and low packet loss severely degrade perceptual quality versus perfect video. Perceptual quality drops by over 50% to less than 400 in the presence of jitter or packet loss. However, moving from low to high loss or

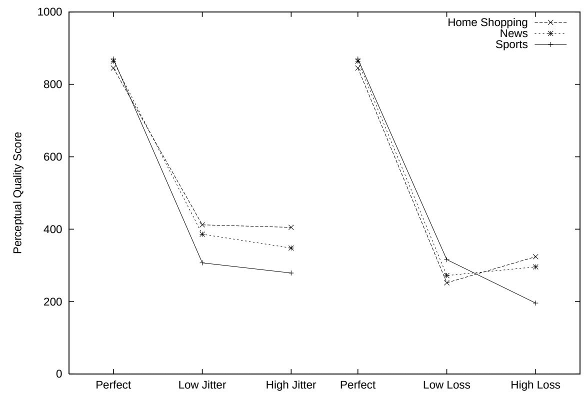

Figure 4: Perceptual Quality and Temporal Aspect. There are three classes of videos shown: Home Shopping (low temporal), CNN News (medium temporal) and Sports (high temporal). The vertical axis is the perceptual quality score given by the user. The data points on the left side of the graph are perceptual quality scores with jitter. The data points on the right are perceptual quality scores with packet loss.

jitter does *not* result in a significant decrease in perceptual quality despite there being a three-fold increase in the magnitude of jitter or packet loss.

Most importantly, there is only a small difference between perceptual quality under equal amounts of jitter and packet loss. Perceptual quality differs by only about 10% between jitter and packet loss under both low and high. From this we conclude that jitter can be nearly as important as packet loss in influencing perceptual quality.

3.2 Perceptual Quality and Temporal Aspect

Figure 4 depicts how the temporal aspect of video influences the impact of jitter and packet loss on perceptual quality. For this analysis, we chose low temporal (Information, via the Home Shopping clips), medium temporal (News, via the CNN clips) and high temporal (via the Sports clips) video. For video clips with jitter, the low temporal aspect clips suffered from slightly less perceptual quality degradation than the high temporal aspect clips. In the data points on the right in which the clips were induced with packet loss, the effects on temporal aspect is less clear, as the lines overlap between low and high data loss. We conclude that temporal aspect is only marginally significant in determining perceptual quality under jitter, and may not be significant in determining perceptual quality under packet loss.

4 Conclusions

The power of today's computers and the connectivity of today's networks present the opportunity of packet video over the Internet and to the desktop. However, packet video over the Internet often suffers from packet loss and jitter, making a clear understanding of the effects of packet loss and jitter on video quality as seen from the user's perspective crucial in order to focus research efforts on the bottleneck in video quality as well as influence playback and packet repair strategies. To date, there has been a solid understanding of the effects of packet loss on video quality, and there has been research into recovering from or ameliorating the effects of both packet loss and jitter. However, to the best of our knowledge, there have not been any user studies that measure the effects of jitter on the perceptual quality of packet video.

In this work, we present carefully designed and executed experiments that measure the effects of jitter and packet loss on the perceptual quality of video. Our experiments incorporated over 40 users who viewed 25 video clips at 5 different quality levels, comprising over 15 hours of user testing while watching over 1000 video clips.

We find jitter can degrade video quality nearly as much as packet loss. Moreover, the presence of even low amounts of jitter or packet loss results in a severe degradation in perceptual quality, while higher amounts of jitter and packet loss do not degrade perceptual quality a proportional amount. Video with low temporal aspect does not suffer in perceptual quality quite as much in the presence of jitter as does high temporal aspect video, but the difference is quite small. These results can be used to enhance delay buffering algorithms in order to achieve better perceptual quality and for better decisions on how to repair lost packets.

Our results indicate a steep drop in perceptual quality between perfect video and our "low" levels of jitter and packet loss. Additional user studies are needed to pinpoint exactly where below our "low" threshold perceptual quality starts to degrade. Such future experiments may best be done using forced choice comparison rather than quality opinion scores[WS98].

The impact of delay, the third fundamental component in perceptual quality, has not been directly compared to either packet loss or to jitter. Future work involves careful user studies following the experimental model described in Section 2 to determine the relative importance to perceptual quality of delay, compared to jitter and packet loss. Moreover, there may be effects that influence quality when there is a combination of delay, packet loss and jitter that do not show up when measuring them separately. Careful user studies are needed to determine perceptual quality of packet video under these combined conditions.

References

[CPW98] S. Cen, C. Pu, and J. Walpole. Flow and congestion control for internet streaming applications. In *Multimedia Computing and Networking*, 1998.

[CR99] M. Claypool and J. Riedl. End-to-end quality in multimedia applications. *Chapter 40 in Handbook on Multimedia Computing*, 1999.

[CT99] M. Claypool and J. Tanner. The effects of jitter on the perceptual quality of video. Technical Report WPI-CS-TR-99-02, Worcester Polytechnic Institute, January 1999.

[DCJ93] S. Dimolitsas, F. Corcoran, and J. Phipps Jr. Impact of transmission delay on ISDN videotelephony. In *Globecom '93*, pages 376 – 379, Houston, TX, November 1993.

[Fer92] D. Ferrari. Delay jitter control scheme for packet-switching internetworks. *Computer Communications*, 15(6):367 – 373, July 1992.

[GBC98] J. Gerek, W. Buchanan, and M. Claypool. MMLIB – a library for end-to-end simulation of multimedia over a WAN. Technical Report MQP MLC-ML98, Worcester Polytechnic Institute, May 1998.

[GKL+98] S. Gringeri, G. Khasnabish, A. Lewis, K. Shuaib, R. Egorov, and B. Basch. Transmission of MPEG-2 video streams over ATM. *IEEE Multimedia*, 5(1):58 – 71, Jan-Mar 1998.

[GT98] G. Ghineas and J. Thomas. QoS impact on user perception and understanding of multimedia video clips. In *ACM Multimedia*, Bristol, UK, September 1998.

[HSK98] V. Hardman, M. A. Sasse, and I. Kouvelas. Successful multi-party audio communication over the Internet. *Communications of the ACM*, 41(5):74 – 80, 1998.

[IKK93] S. Iai, T. Kurita, and N. Kitawaki. Quality requirements for multimedia communcation services and terminals – interaction of speech and video delays. In *Globecom '93*, pages 394 – 398, Houston, TX, November 1993.

[MS94] M. Massimino and T. Sheridan. Teleoperator performance with varying force and visual feedback. In *Human Factors*, March 1994.

[PHH98] C. Perkins, O. Hodson, and V. Hardman. A survey of packet-loss recovery techniques for streaming audio. *IEEE Network Magazine*, Sep/Oct 1998.

[PVM98] M. Podolsky, M. Vetterli, and S. McCanne. Limited retransmission of real-time layered multimedia. In *IEEE Workshop on Multimedia Signal Processing*, December 1998.

[RKTS94] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *IEEE Computer and Communications Societies on Networking for Global Communciation*, pages 680–688, Los Alamitos, CA, June 1994. IEEE Computer Society Press.

[Roy94] R. Roy. Networking contraints in multimedia conferencing and the role of ATM networks. *AT&T Technical Journal*, July/August 1994.

[SJ95] D. Stone and K. Jeffay. An empirical study of delay jitter management policies. *ACM Multimedia Systems*, 2(6):267 – 279, January 1995.

[SW93] M. Swartz and D. Wallace. Effects of frame rate and resolution reduction on human performance. In *IS&T's 46th Annual Conference*, Munich, Germany, 1993.

[Tan99] J. Tanner. Java jitters – the effects of Java on jitter in a video stream. Master's thesis, Worcester Polytechnic Institute, May 1999.

[WS98] A. Watson and M. Sasse. Measuring perceived quality of speech and video in multimedia conferencing applications. In *ACM Multimedia*, Bristol, UK, September 1998.

[XMZY97] X. Xu, A. Myers, H. Zhang, and R. Yavatkar. Resilient multicast support for continuous-media applications. In *Intl Workshop on Network and OS Support for Digital Audio and Video (NOSSDAV97)*, 1997.