WPI-CS-TR-99-02 January1999

The Effects of Jitter on the Perceptual Quality of Video

by

Mark Claypool and Jonathan Tanner

Computer Science
Technical Report
Series

WORCESTER POLYTECHNIC INSTITUTE

Computer Science Department
100 Institute Road, Worcester, Massachusetts 01609-2280

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

January 19, 1999

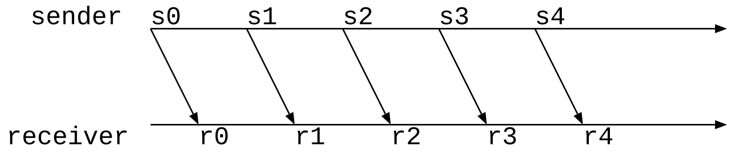
Abstract

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. The effects of packet loss on perceptual quality are well-understood, but 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.

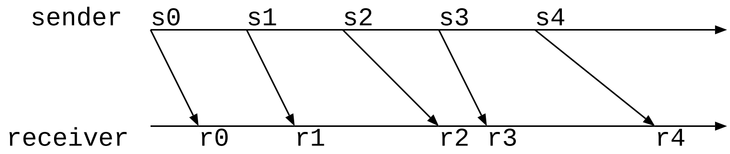
1 Introduction

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. Video on the desktop and over a local area network are 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, long delays and missing frames. Although we often think of a video as a continuous stream of data, computer systems handle video in discrete events. An event may be that of receiving a new video frame or of displaying an uncompressed frame on the screen. The quantity and timing of these events give us measures that affect video quality. There are three measures that determine quality for most video applications: delay, packet loss and jitter [CR98].

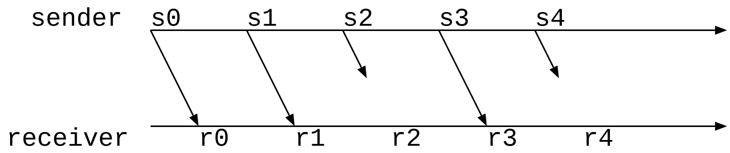
In the absence of jitter and packet loss, video frames can be played as they are received, resulting in a smooth playout, as depicted in Figure 1-top. However, in the presence of jitter, interarrival



A Jitter and Loss Free Stream



A Stream with Jitter



A Stream with Packet Loss

Figure 1: The above figures model packet video between a sender and receiver. Each si is the time at which the send process initiates the transmission of a video frame i. Each ri is the time at which the receiving process plays frame i.

times will vary, as depicted in Figure 1-middle. In Figure 1-middle the third frame arrives late at r*2*. The user in this scenario would see the frozen image of the most recently delivered frame, i.e. frame two until the tardy frame, frame three, arrives. The tardy frame (frame three) would then be played only briefly in order to preserve the timing for the subsequent frame. 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. For example for the loss of the frame three, the viewer would see the frozen image of the most recently delivered frame, i.e., frame two, and the video stream would then jump to the next frame that arrived, i.e., frame four.

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 [AFKN95, MS94, SW93, HSK98, PHH98, GKL+98]. Methods to ameliorate the effects of jitter have been explored by many researchers [SJ95, RKTS94, Fer92]. 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. We use the term *perceptual quality* to mean a quantitative measure of video quality from the user perspective.

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 the "talking heads" in videoconferences or news broadcasts, 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 clips of sporting events or music videos, on the other hand 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.

A clear evaluation of the impact of jitter as compared to packet loss on perceptual quality can help to decide on the retransmission strategy in the presence of packet loss. If lost, video packets may be retransmitted, but at the cost of roughly doubling the delay for the re-sent packets, thereby increasing jitter. Other techniques to repair packet loss, such as interleaving or redundancy, also result in an increase in jitter [PHH98]. If lost packets are found to degrade quality more than the extra jitter they add, then lost packets should be retransmitted. If, however, the degradation in quality from the added jitter decreases perceptual quality more than that caused by the lost packets, then retransmissions should not be done. A careful understanding of the effects of jitter and packet loss on perceptual quality can be used to decide when to do retransmission of lost packets.

To summarize, in this work 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 and,
- an analysis of the effects of the temporal aspect of a video on the perceptual quality in the presence of jitter and packet loss.

In addition in this work we also show:

- a new means of quantifying perceptual quality for future user studies and,
- an experimental model for conducting user studies that measure perceptual quality of packet video

The rest of this document is laid out as follows: Section 2 describes our experiments that measure perceptual quality with packet loss and with jitter; Section 3 analyzes the results of the experiments; Section 4 summarizes our conclusions; and Section 5 lists possible future work.

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. There are four basic methods to measure perceptual quality [WS98, GT98]:

- *Forced Choice* in which test subjects are presented the same clip under two different quality levels. Forced choice evaluations are useful for determining precise boundaries of human perception between quality levels, but are not as effective when the general whereabouts of such boundaries is not known, as in our case for perception of jitter.
- *Content Query* in which test subjects are asked questions about the content of video clips after watching them. Content query questions are often difficult to choose properly since "trick" questions can unduly influence the test results.
- *Effectiveness Study* in which different quality videos are used for real life tasks and the task effectiveness, then, becomes the measure of quality. Effectiveness studies generally use real-users and real-applications, making their results directly applicable, but they tend to be lengthy and expensive for studies such as ours.
- *Quality Opinion Score* in which test subjects are asked for an opinion score after watching a video clip. We use this method for our experiments since it does not have the drawbacks of the other three methods and has been used quite successfully in the form of Mean Opinion Score (MOS) testing for audio quality.

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 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 from 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 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 Polytechnic Institute [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 5 different quality levels for each video clip: perfect, low jitter, low loss, high jitter or high

	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.

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 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. During each clip, the test subjects were asked to indicate by clicking the mouse whenever they observed a quality degradation event.

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, the influence of temporal aspect of video on perceptual quality, and the effectiveness of recording quality degradation events in determining 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 window size used in the study or 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 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 jitter does *not* result in a significant decrease in perceptual quality despite there being a three-fold magnitude increase in the magnitude of jitter and 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 is nearly as important as packet loss in influencing perceptual quality in some cases.

3.2 Perceptual Quality and Temporal Aspect

Figure 4 depicts how temporal aspect of video influences the effects of jitter and packet loss. 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 lower the 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.

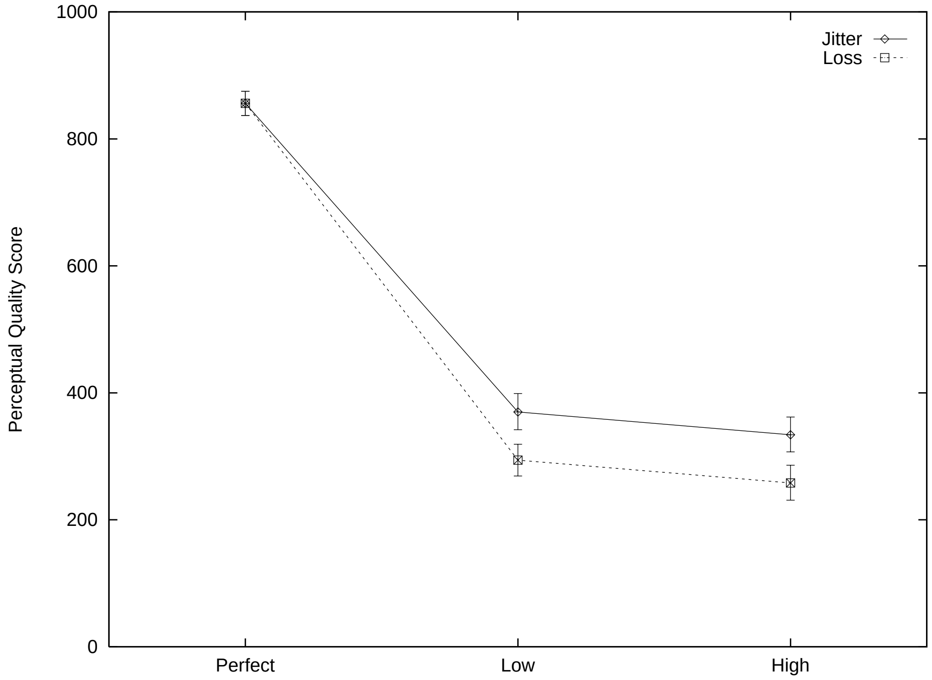


Figure 3: Perceptual Quality versus Jitter or Packet Loss. There are five quality levels depicted: perfect, low loss, low jitter, high loss and high jitter. The horizontal axis indicates perfect, low or high. The vertical axis indicates the average perceptual quality opinion score given for all test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

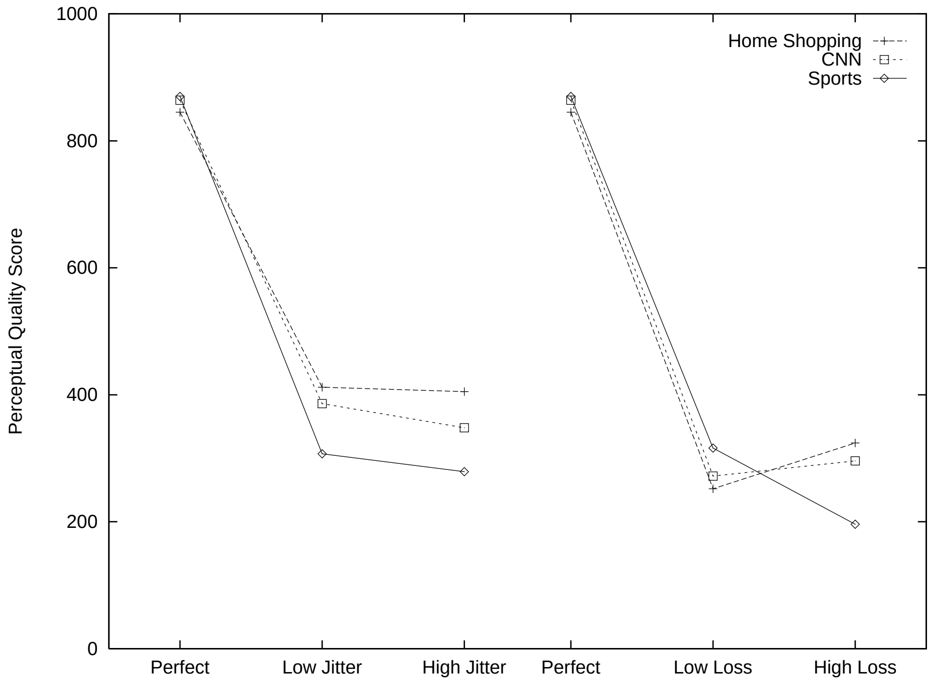


Figure 4: Perceptual Quality and Temporal Aspect. There are three classes of videos shown: Home Shopping (low temporal), CNN News (medium temporal) and Sports (high temporal). The vertical axis is the perceptual quality score given by the user. The points on the left are perceptual quality scores with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

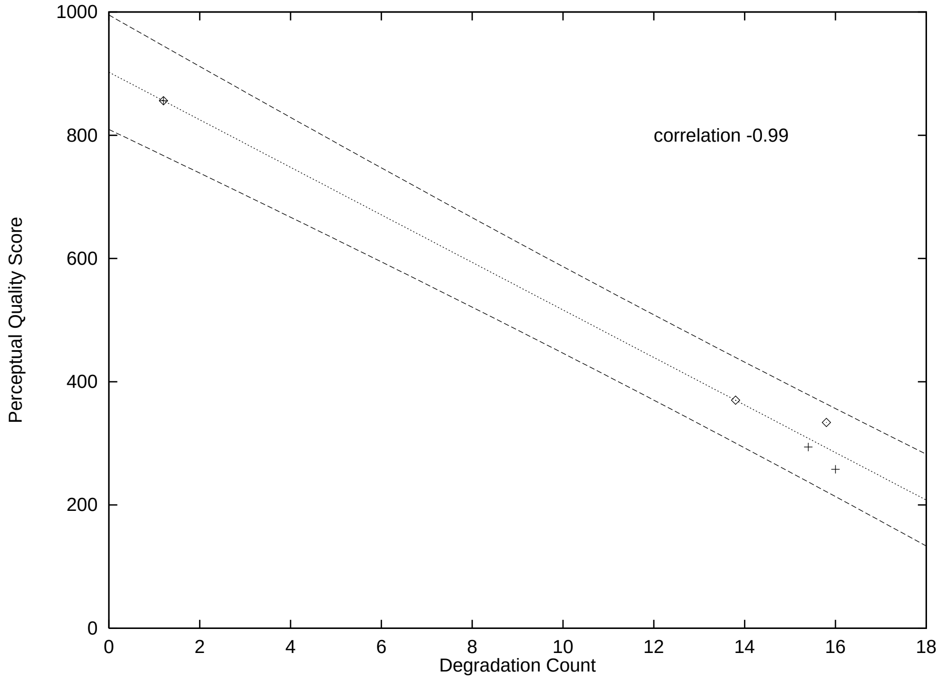


Figure 5: Average Perceptual Quality versus Average Degradation Count. The horizontal axis is the average number of degradation events show. The vertical axis is the average perceptual quality value recorded. The five points are each of: perfect, low jitter, low loss, high jitter and high loss. The middle line is the least squares line fit between the points. The two curves around the line are 95% confidence intervals. The Pearson's correlation coefficient is -0.99.

3.3 Perceptual Quality and Degradation Count

Lastly, we examine how the recording of degradation events corresponds with perceptual quality. We compute the correlation between average number of degradation events and average perceptual quality. Figure 5 depicts these results. The Pearson's correlation coefficient between the average number of degradation events recorded and average perceptual quality is a very strong -0.99, suggesting that the presences of degradation events is a very good indication of a low quality value.

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 retransmission 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 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.

In summary, 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.
- There is a strong correlation between the average number of quality degradation events and the average perceptual quality values recorded. This suggests that the number of degradation events observed may be a good indicator of how much a user will like a video clip.

5 Future Work

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.

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 or jitter that do not show up when measuring them separately. Careful user studies are needed to determine perceptual quality of packet video under these combination conditions.

References

[AFKN95] Ronnie T. Apteker, James A. Fisher, Valentin S. Kisimov, and Hanoch Neishlos. Video acceptability and frame rate. *IEEE Multimedia*, pages 32 – 40, Fall 1995.

[CR98] Mark Claypool and John Riedl. End-to-end quality in multimedia applications. In *Chapter 40 in Handbook on Multimedia Computing*, CRC Press, Boca Raton, FL, 1998.

[DCJ93] Spiros Dimolitsas, Franklin L. Corcoran, and John G. Phipps Jr. Impact of transmission delay on ISDN videotelephony. In *Proceedings of Globecom '93 – IEEE Telecommunications Conference*, pages 376 – 379, Houston, TX, November 1993.

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

[GBC98] Jason Gerek, William Buchanan, and Mark 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] Steven Gringeri, Ghumip Khasnabish, Arianne Lewis, Khaled Shuaib, Roman Egorov, and Bert Basch. Transmission of MPEG-2 vidoe streams over ATM. *IEEE Multimedia*, 5(1):58 – 71, Jan-Mar 1998.

[GT98] G. Ghineas and J.P. Thomas. QoS impact on user perception and understanding of multimedia video clips. In *Proceedings of ACM Multimedia*, Bristol, United Kingdom, 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] Satoru Iai, Takaaki Kurita, and Nobuhiko Kitawaki. Quality requirements for multimedia communcation services and terminals – interaction of speech and video delays. In *Proceedings of Globecom '93 – IEEE Telecommunications Conference*, pages 394 – 398, Houston, TX, November 1993.

[MS94] Michael J. Massimino and Thomas B. Sheridan. Teleoperator performance with varying force and visual feedback. In *Human Factors*, pages 145 – 157, March 1994.

[PHH98] Colin Perkins, Orion Hodson, and Vicky Hardman. A survey of packet-loss recovery techniques for streaming audio. *IEEE Network Magazine*, September/October 1998.

[RKTS94] Ramachandran Ramjee, Jim Kurose, Don Towsley, and Henning Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *Proceedings of the 13th Annual Joint Conference of the IEEE Computer and Communications Societies on Networking for Global Communciation. Volume 2*, pages 680–688, Los Alamitos, CA, USA, June 1994. IEEE Computer Society Press.

[Roy94] Radhika 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] Merryanna Swartz and Daniel Wallace. Effects of frame rate and resolution reduction on human performance. In *Proceedings of IS&T's 46th Annual Conference*, Munich, Germany, 1993.

[WS98] A. Watson and M. Sasse. Measuring perceived quality of speech and video in multimedia conferencing applications. In *Proceedings of ACM Multimedia*, pages 55 – 60, Bristol, United Kingdom, September 1998.