A Two-Dimensional Audio Scaling Enhancement to an Internet Videoconferencing System*

Peter Nee Kevin Jeffay Michele Clark

University of North Carolina at Chapel Hill
Department of Computer Science
Chapel Hill, NC 27599-3175
{nee,jeffay,clark}@cs.unc.edu

Gunner Danneels

Intel Corporation
Intel Architecture Labs
Hillsboro, OR
Gunner_Danneels@ccm.jf.intel.com

http://www.cs.unc.edu/Research/Dirt

ABSTRACT: *Without widely deployed mechanisms for resource reservation, adaptive, best-effort techniques are the only methods of congestion control available to interactive, real-time applications. Here we discuss media scaling techniques for the audio component of videoconferencing systems and describe an experimental version of the ProShare system that incorporates two of these techniques. We also report the results of experiments using these techniques to adapt to daytime Internet traffic conditions.*

1. INTRODUCTION

This paper presents the results of experiments comparing two schemes for best-effort, adaptive congestion control for videoconferencing. In particular, these two techniques used two different approaches to the adaptation of the audio component of a videoconferencing system. The first of these two schemes used conventional bit-rate scaling to adapt the audio stream. The second scheme employed a two-dimensional scaling scheme to adjust the audio packet-rate as well as the audio bit-rate. Two-dimensional scaling schemes were proposed and shown to be effective in campus-size internetworks in [8, 9]. Further work incorporating a two-dimensional scaling scheme into an experimental version of the Intel ProShare™ videoconferencing system also showed some benefit using a lower bit-rate codec over longer, Internet paths [7]. The system described here focused specifically on audio adaptation which was not part of this previous ProShare implementation.

It is well known that the quality of the audio stream of a videoconference is the single largest factor in user perception of overall conference quality. Indeed, if a videoconferencing system cannot deliver low latency, low loss audio over the current network conditions, communication with the system becomes impossible. Our goals in applying two-dimensional scaling to the audio component of our ProShare system were to demonstrate an adaptive scheme that could first, deliver higher quality audio in commonly arising network conditions, and second, provide feasible audio over a wider range of network conditions.

Unfortunately, our experiments failed to show a clear advantage for the two-dimensional scheme in either of these areas. However, this work is still of interest as it highlights the issues concerning two-dimensional scaling, and serves as a first step toward a more effective implementation of two-dimensional scaling.

The next section provides some background on media scaling in general and two-dimensional scaling in particular. Section 3 describes our incorporation of audio scaling schemes into the ProShare framework. Sections 4 and 5, respectively, describe our experimental method and the results of our experiments. In section 6 we discuss the results, drawing conclusions and identifying potential future improvements.

2. BACKGROUND AND RELATED WORK

Distributed, real-time multimedia applications such as videoconferencing must provide a smooth, low loss flow of media units to a human participant in the face of the latency, variation in latency (delay-jitter), and loss introduced by congestion in the network. Current internetworks are primarily composed of best-effort components such as shared media LANs, wide-area telecommunication links, and routers and switches that use only simple methods of packet scheduling (*e.g.* FIFO queueing and packet queue tail drop for buffer shortages). Although most such networks provide a sufficient quality of service when lightly loaded, contention for resources leads to congestion which results in increased latency, increased delay-jitter, and greater packet loss for traffic through a constrained element of the network. Thus, the success of distributed real-time multimedia applications hinges on a successful method of congestion control.

One strategy for congestion control for real-time multimedia traffic is resource reservation. Resource reservation schemes allocate network resources to a specific application or class of applications [2, 6, 10, 11]. This approach can provide guaranteed quality of service to applications that make reservations, but only if support for the reservation mechanism is

*Supported by grants from the Intel Corporation, the National Science Foundation (grants IRIS-9508514 & CCR-9510156), and the Advanced Research Projects Agency (grant 96-06580).

In: *Proceedings of the International Workshop on Audio-Visual Services over Packet Networks*, Aberdeen, Scotland, UK, September 1997, pages 201-206.

widely deployed and network resources can be allocated along the entire path from sender to receiver.

Another strategy for congestion control is adaptive media scaling. This approach adjusts the attributes of the transmitted media streams to the current network conditions [1, 3, 4, 6, 8], for example, reducing the video frame rate in response to packet loss. This approach is well-suited to networks containing best-effort components, as well as the presumably more affordable best-effort service classes of most proposed resource reservation schemes. In addition, media such as audio and video are inherently scaleable and thus able to convey usable information over a wide range of quality/resource consumption tradeoffs. It is therefore useful to study adaptive media scaling techniques to optimize them for efficient use in best-effort internetworks.

The motivation for two-dimensional scaling schemes incorporating both bit-rate and packet-rate adaptations, is a classification of the resource constraints the underlie network congestion into one of two types. Some resources are consumed by the transmission of bits through the network. These include link bandwidth and packet movement time in router memory. If a network path is unable to sustain the current bit-rate of the aggregate traffic, queues begin to grow and congestion arises. We classify a network path congested for this reason as *capacity constrained*. In contrast, some network resources are consumed by packet transmission, regardless of packet size (and thus regardless of bit-rate). These include media access time for shared media LANs such as a shared Ethernet segment, and packet processing overhead at router CPUs. If a network is unable to sustain the current packet-rate of the aggregate traffic, queues again grow and the symptoms of congestion appear. We classify a network congested for this reason as *access constrained*. Note that an access constraint can possibly be relieved by simply repackaging the traffic, without a reduction of bit-rate, so that the same data is sent in fewer packets. Such a repackaging may introduce additional latency, but if the repackaging relieves congestion, total end-to-end latency may in fact be reduced.

We characterize the media streams produced by a videoconferencing system as operating points in a bit-rate × packet-rate plane. Each point in this plane represents the raw bit-rate of the system as well as the way the media units are partitioned into network packets. An adaptation is a move from one operating point to another. We can classify an adaptation scheme as one-dimensional or two-dimensional by the operating point set it uses.

One-dimensional schemes use operating points that are roughly collinear in the bit-rate × packet-rate plane. This include schemes with points on a vertical line (*e.g.*, audio bit-rate scaling that maintains a constant packet-rate), a horizontal line (*e.g.*, an audio aggregation scheme that does not change the sample rate, but rather the number of samples per packet), or a diagonal line that reduces both bit-rate and packet-rate in a roughly proportional fashion (*e.g.*, temporal video scaling, which reduces the video frame-rate and thereby both video bit-rate and video packet-rate). One-dimensional schemes can use a very simple probe and retreat algorithm to move through the operating point set in response to feedback.

Two-dimensional schemes make use of points that cover an area of the bit-rate × packet-rate plane, not just a line. This allows greater flexibility in adapting the packet-rate to access constraints and the bit-rate to capacity constraints. For example, a one-dimensional scheme that uses temporal video scaling to adapt to a capacity constraint will reduce video packet-rate as well, whereas a two-dimensional scheme might only reduce the bits per frame and maintain a higher frame-rate and bit-rate. In addition, a scheme that uses bit-rate scaling only will be unable to adapt to an access constraint. A two-dimensional scheme could reduce packet-rate to meet this constraint without affecting bit-rate.

With the additional adaptation options available to a two-dimensional scheme comes additional complexity in finding the best operating point. Both access and capacity constraints result in the same end-to-end symptoms of congestion: increased latency, delay-jitter, and/or packet loss. The optimal adaptation choice cannot be empirically determined at each step. However, a relatively simple heuristic approach (described in §3.3) has been used successfully [7-9].

3. SYSTEM DESCRIPTION

The system built for these experiments was based on two components. The ProShare videoconferencing system provided a video codec, call setup, and UDP based transport. A commercially available sound card capturing PCM audio was used to simulate an audio codec supporting audio scaling.

3.1 Video Component

Our experimental system was based on ProShare 1.8. This version of ProShare implements roughly constant bit-rate, point-to-point videoconferences. Before the start of a conference, the user may choose from one low bit-rate operating point suitable for ISDN or LAN operation, or one of two higher bit-rate operating points suitable only for LAN operation.

Our experimental system uses an internal interface to the video codec to allow on-the-fly adjustments to the bit-rate and frame-rate of the video. In the work presented here, video bit-rate and frame-rate were adjusted proportionally, so as to always generate frames of about 2,100 bytes. This allowed the implementation of a simple one-dimensional scaling scheme for

Table 1: Audio simulation parameters.

Simulated Sample Rate (kHz)	Record Buffer Size (bytes)	Transmit Buffer Size (bytes)
22	256	1024
16.5	256	768
16.5	384	1152
11	256	512
11	384	768
11	512	1024
8.25	256	384
8.25	384	536
8.25	512	768
5.5	256	256
5.5	384	384
5.5	512	512

the video component of the conferences run for these experiments.

3.2 Audio Component

ProShare's built in audio operates at a fixed rate of ten 200 byte frames per second. Since our goal was to compare two different systems for adapting audio, we instead used a separate sound card to generate a simulation of a very flexible audio codec. The goal was not to generate playable audio (in fact, the audio generated this way was unplayable), but rather to simulate the relevant characteristics of an abstract codec with the desired attributes.

Although our abstract audio codec is capable of operating at a variety of bit-rate and buffer size combinations, in a practical system the bit-rate variability might best be implemented by switching between various compression schemes. For our purposes, it was sufficient to model a PCM system with a selection of sampling frequencies. However, even this approach had two problems that lead to the further use of simulation in our experimental system.

First, even operating in full duplex mode, our sound card was incapable of supporting different sample frequencies for input and output. Although the two systems coordinated their operating point selection, the sound card could not support the transitional periods that arose during each change in sampling frequency. In essence, this meant that our system generated audio data that could not be successfully played out on our hardware. It was this limitation that lead us to simulate an ideal codec that generated the desired network traffic, without trying to produce playable audio. Although playable audio is a useful check on the successful operation of an experimental system, experience has shown us that measurements of the transmission quality of the audio stream are sufficient to judge the feasibility of the stream, and thus suitable to judge the success of properly simulated adaptations.

The second limitation was that changing the sampling frequency required closing and re-opening the card. This caused enough of a fluctuation in card output that it was decided to operate the card at one fixed sample frequency. To simulate operation at higher sample frequencies, two notions of buffer size were introduced. The record buffer size was varied to simulate variation in sample frequency. For example, at 5.5 kHz sample frequency filling 256 byte buffers generates interrupts at the same rate as 22 kHz sample frequency filling 1024 byte buffers. To complete the simulation of a 22 kHz, 1024 byte buffer operating point, it is still necessary to actually transmit 1024 bytes. Therefore, we introduce the concept of a transmit buffer size, distinct from the record buffer size used to obtain interrupts from the sound card at the correct rate. Table 1 summarizes the record buffer sizes and transmit buffer sizes used to simulate the complete set of operating points for our abstract audio codec. The card operates at a constant sample rate of 5.5 kHz.

The simulated operating points are plotted in the bit-rate × packet-rate plane in Figure 1. Note that they form a rectangular grid with some additional points at the upper right.

All of these audio operating points are available to the two-dimensional adaptive scheme, which seeks to adjust both the packet-rate and the bit-rate of the outgoing audio stream.

The one-dimensional scheme uses only a subset of the available audio operating points, because it seeks only to adjust the bit-rate of the outgoing audio stream. Thus, any of the three columns of the audio operating point grid would be a suitable subset for the

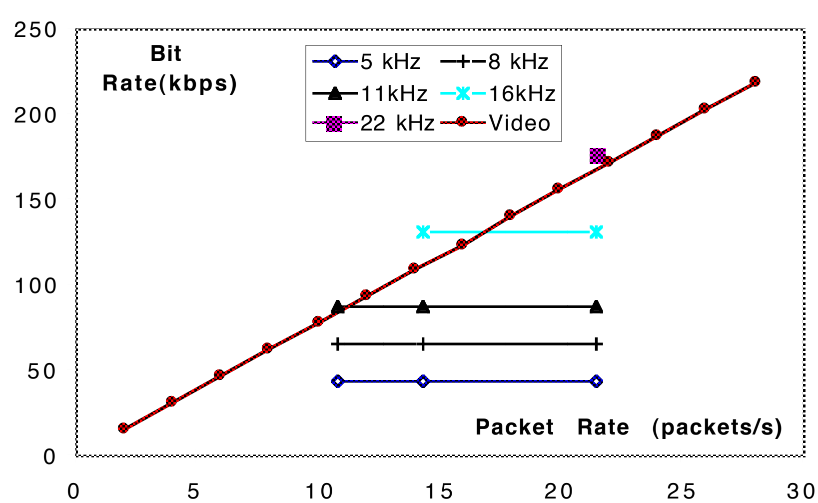


Figure 1: Experimental system operating points.

one-dimensional scheme. We chose the far right column, which has the lowest latency (highest packet-rate) of the three columns, because choosing to minimize latency is a reasonable design choice.

3.3 One-Dimensional Adaptation Algorithm

Once per second, feedback from the receiver is used to classify the network as either congested or uncongested. The network is classified as congested if more than two packets are lost, or if network latency has increased fifty percent over a moving average of the previous five uncongested latency measurements.

If the network is classified as congested, the algorithm retreats for both video and audio streams. For video at 12 or 14 frames per second a retreat is a reduction of two frames per second. For video between 7 and 10 frames per second, a retreat is a reduction of one frame per second. Six is the lowest frame-rate permitted. For audio, a retreat is a reduction to next lower simulated sampling rate.

If the network has been classified as uncongested for four consecutive feedback messages, the algorithm will probe to see if a better quality operating point can be maintained. A probe entails increasing both the current video frame rate (by one or two frames per second in the inverse of the retreat adaptations) and the current audio sampling rate to the next higher level.

3.4 Two-Dimensional Adaptation Algorithm

The algorithm used for our two-dimensional scheme is based on the one-dimensional scheme described above. In particular, the feedback messages, classification of the network, decision to retreat or probe, and the video adaptation are handled the same way. The two-dimensional scheme described here differs only in the way audio adaptation is done.

To control audio adaptation, a “recent success” history mechanism is used. For example, if the last instance of congestion was relieved by a reduction in bit-rate, it is assumed that the current bottleneck is a capacity constraint. A bit-rate reduction is attempted first if congestion appears again, and if a probe is necessary, an increase in packet-rate, rather than an increase in bit-rate will be attempted. Similarly, relief of congestion by a packet-rate reduction leads to initial retreats in the packet-rate dimension and initial probes in the bit-rate dimension. If a retreat fails to relieve congestion, an adaptation in the other dimension is attempted. Thus, repeated indications of congestion lead to a “stairstep” retreat to the origin. If a probe encounters congestion, it is undone, and after a four second interval of no congestion, a probe in the other dimension is attempted. This allows the algorithm to work its way up and to the right to find the best quality bit-rate and packet-rate combination sustainable in the current network conditions.

4. EXPERIMENTAL METHOD

To compare our two adaptive schemes we performed a five minute run of each scheme one after the other. For each pair of runs a coin toss was used to randomly order the two schemes. These runs were performed during the busiest part of the day, 10:00 am to 4:00 pm EDT, and repeated over many days.

To route the traffic over lengthy Internet paths, reflectors were set up at remote sites. Conference packets traveled from the sender to the reflector(s) and then back to the receiver. Up to four runs were performed a day, with a space of at least an hour and a half between each run. Two reflector configurations were used, each giving a total path length of roughly twenty-eight hops. The first configuration used a single reflector at the University of Washington. The second configuration used two reflectors one at the University of Virginia, the other at Duke University. These two reflectors were chained together so that all conference packets went through both reflectors.

For each reflector configuration, up to four runs were performed in a day. At least an hour and a half interval was required between runs using the same reflector configuration.

Because ProShare is a bi-directional conferencing system, the adaptations of the two systems were coordinated with a master/slave arrangement. Both systems monitored the quality of the conference data from the other system and reported feedback as described above. The master system ran the adaptive heuristic based on the slave system’s feedback and adjusted the current operating point as described above. The master system then reported its current operating point to the slave system and the slave system changed its current operating point to match.

5. EXPERIMENTAL RESULTS

We compared one minute averages of corresponding conference minutes for each of four criteria. These comparisons are in Figure 2 with the results for the Duke/UVa/UNC configuration on the left, and the results for the UW/UNC configuration on the right. The bars in each chart are ordered chronologically. Each group of five bars are minutes of the same run. Some correlation between minutes of the same run can be seen.

The first chart in each column shows the absolute difference in audio bit-rate throughput. A positive number is a higher bit-rate delivered by the two-dimensional system, a negative is a higher bit-rate delivered by the one-dimensional system. For both configurations, there is not a dramatic difference, although the one-dimensional system delivered a higher bit-rate slightly more often.

The second comparison is of video throughput, as measured in delivered video frame rate. Again, the absolute difference between the two-dimensional and one-dimensional rates are presented. Again, there is no dramatic difference, although the two-dimensional scheme delivered more video frames slightly more often.

The third comparison is a feasibility test, the percentage of packets lost in the network. These are presented separately for the two schemes in the third and fourth charts in each column. Taking an average percentage loss of three percent as an indication of infeasibility, there are no dramatic differences in feasibility between the two schemes, although the two configurations do differ. The Duke/UVa/UNC configuration was almost always able to deliver a feasible conference, the UW/UNC configuration had roughly twenty-five out a hundred infeasible minutes for both schemes.

The final comparison is a difference plot for video latency. Neither scheme demonstrates a significant advantage here.

Because we compared conference quality measurements gathered a few minutes apart, no conclusions can be drawn from individual comparisons. Shifts in network conditions dominate many individual measurements. Instead, we must examine the trends over a large number of runs to identify differences that stem from the capabilities of the two schemes. Examination of our results shows only a slight, but consistent video throughput advantage for the two-dimensional scheme, and a slight, but consistent audio throughput advantage for the one-dimensional scheme.

6. SUMMARY AND CONCLUSIONS

In this paper, we presented an explanation of two-dimensional scaling techniques and applied them to the audio component of an experimental videoconferencing system. We presented the results of experiments comparing the two-dimensional scheme to a more conventional media scaling scheme over lengthy Internet paths. Past experiments with higher bit-rate videoconferencing systems and shorter Internet paths have shown a clear advantage for two-dimensional techniques. We therefore conclude that scaling those benefits to the extreme environments reflected in this work remains a challenge.

It is quite possible that enhancements to the system described here could address this scalability issue. For example, adding a two-dimensional video component, or fine tuning the heuristic for congestion detection and adaptation control may extend the benefits of two-dimensional scaling to this larger scale. In any case, since the two-dimensional scaling was not disadvantaged at this scale, a system that uses two-dimensional scaling will have a net advantage in conference quality over a variety of internetwork path lengths and conditions.

7. ACKNOWLEDGEMENTS

We are indebted to Jorg Liebherr at the University of Virginia, Jeff Chase at Duke University, and Brian Bershad at the University of Washington for use of facilities to perform the experiments reported herein.

8. REFERENCES

[1] Bolot, J., Turletti, T., *A Rate Control Mechanism for Packet Video in the Internet*, Proc. IEEE INFOCOMM ‘94, Toronto, Canada, June 1994, pp. 1216-1223.

[2] Braden, R., D. Clark, and S. Shenker, “Integrated Services in the Internet Architecture: an Overview,” IETF RFC-1633, July 1994.

[3] Chakrabarti, S., Wang, R., *Adaptive Control for Packet Video*, Proc. IEEE International Conference on Multimedia Computing and Systems 1994, Boston, MA, May 1994, pp. 56-62.

[4] Delgrossi, L., et al., *Media Scaling for Audiovisual Communication with the Heidelberg Transport System*, Proc. ACM Multimedia ‘93, Anaheim, CA, Aug 1993, pp. 99-104.

[5] Ferrari, D., Banjea, A., and Zhang, H., *Network Support for Multimedia: A Discussion of the Tenet Approach,* Computer Networks and ISDN Systems, Vol. 26, No. 10 (July 1994), pp. 1267-1280.

[6] Hoffman, Don, Spear, M., Fernando, Gerard, *Network Support for Dynamically Scaled Multimedia Streams*, Network and Operating System Support for Digital Audio and Video, Proc., D. Shepard, *et al* (Ed.), Lecture Notes in Computer Science, Vol. 846, Springer-Verlag, Lancaster, UK, November 1993, pp. 240-251.

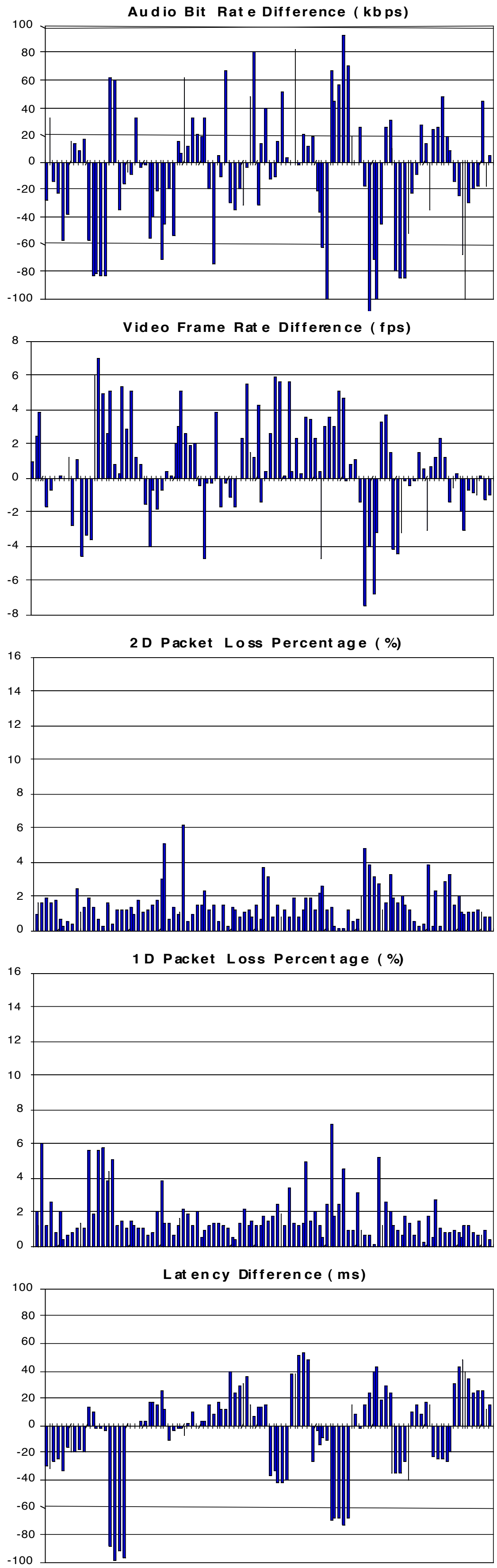
[7] Nee, P., Jeffay, K., Danneels, G., *The Performance of Two-Dimensional Medua Scaling for Internet Videoconferencing,* Proc. 7th Intl. Workshop on Network and Operating System Support for Digital Audio and Video, St. Louis, MO, May 1997, pp. 237-248.

[8] Talley, T.M., Jeffay, K., *Two-Dimensional Scaling Techniques For Adaptive, Rate-Based Transmission Control of Live Audio and Video Streams*, Proc. Second ACM Intl. Conference on Multimedia, San Francisco, CA, October 1994, pp. 247-254.

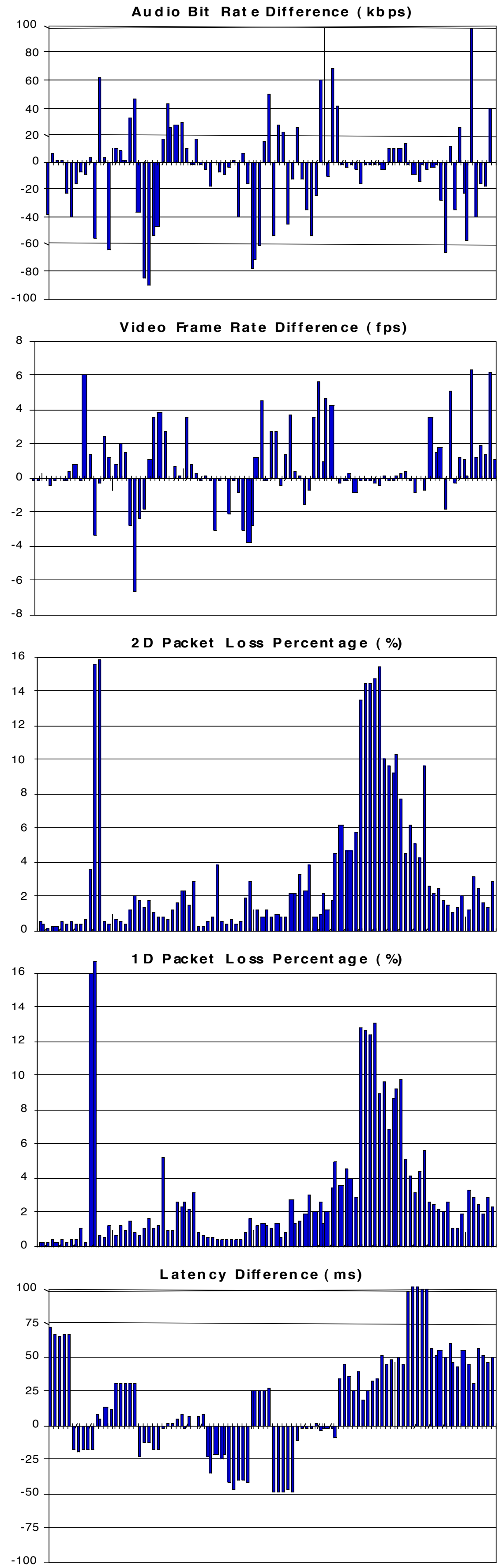
[9] Talley, T.M., Jeffay, K., *A General Framework for Continuous Media Transmission Control*, Proc. 21st IEEE Local Computer Networks Conference, Minneapolis, MN, October 1996, pp. 374-383.

[10] Topolcic, C. (Ed.), *Experimental Internet Stream Protocol, Version 2 (ST-II)*, RFC 1190, IEN-119, CIP Working Group, October 1990.

[11] Zhang, L., et al., *RSVP: A New Resource ReSerVation Protocol*, *IEEE Network*, Vol. 5, No. 5 (September 1993), pp. 8-18.



a) Duke/UVa/UNC experiments.



b) UW/UNC experiments.

Figure 2: Comparison of one- and two-dimensional media scaling results for two Internet paths.