PERFORMANCE EVALUATION OF DESKTOP VIDEOCONFERENCING

Licha Mued, Benn Lines, Steven Furnell and Paul Reynolds

Network Research Group, Department of Communication and
Electronic Engineering, University of Plymouth, United Kingdom
Email: lmued@jack.see.plym.ac.uk, nrg@jack.see.plymouth.ac.uk

Abstract

This paper discusses the evaluation of multimedia conferencing systems with respect to audio and video quality. We investigated various aspects that affect the perceived audio and video quality in desktop videoconferencing (DVC), including network constraints (packet loss, delay jitter and delay), CODEC, and task performance. The paper outlines the benchmarking of Microsoft NetMeeting, using the subjective assessment method of Mean Opinion Score (MOS) and concludes that the various factors mentioned above have different effects and that video is the main determinant of the overall perceived audiovisual quality.

1. Introduction

The state of the art in audio and video assessment has generally focused on finding the point at which degradation is not discernable. Audio and video quality can be measured either subjectively or objectively. It is generally agreed that subjective methods are more reliable [1], but recent research findings suggest that the subjective method alone is inadequate to determine the audio and video quality in videoconferencing [2,3,4].

At present, despite the increased popularity of low cost DVC, it is often questioned whether the quality of the audio and video provided is adequate to the tasks that users wish to perform performance. Some findings suggest that the perceived quality of audio and video varies according to the task undertaken and user expectation also varies accordingly. In the meantime, the issue of determining multimedia conferencing quality has certain difficulties, as there is no recognized industry standard of what really determines audio and video quality. In addition, assessing the quality of multimedia over the Internet is further complicated due to its constantly changing and unpredictable nature [2]. Many research efforts are now being directed toward developing new approaches in assessing audio and video quality in IP multimedia [4,5,6].

The authors are conducting research into quality of service in IP videoconferencing scenarios. The research to date has investigated the current state of the art in desktop videoconferencing. This paper focuses upon benchmarking the performance of the popular Microsoft NetMeeting with respect to the related issues that affect the perceived audio and video quality, such as network congestions, computing resources, tasks performance, CODEC, and conferencing hardware. NetMeeting was selected over other existing IP telephony tools due to its readily available software and its highly demand in the current market. The associated study had three main aims:

(1) To investigate the performance of NetMeeting. This involved assessment of audio quality, video quality and combined audio and video quality under a real network environment, and also under assigned network constraints.

(2) To investigate the task performance effects on audio quality, video quality and the combined quality of audio and video under a real network environment, and also under assigned network constraints.

The purpose of the test in (1) and (2) was to quantify traffic related phenomena, such as packet-loss (i.e. the number of lost packets, reported at the total traffic), delay (i.e. the time passed between the sending of a packet and its arrival at the destination), and delay jitter (i.e. the variance of the delay value) and to present the data in the form, which could be related to perceived quality of service experienced using NetMeeting software.

(3) To investigate the impact of the two speech CODECs (PCM and G723.1) on perceived audio and video quality.

The method of assessment being used is an objective test method, called Mean Opinion Score (MOS) and is the standard recommended by the ITU-T (CCITT, 1984). The MOS is typically a 5-point rating scale, covering the options Excellent (5), Good (4), Fair (3), Poor (2) and Bad (1).

2. The experiment

2.1 The NetMeeting System

NetMeeting is a Windows application that allows real-time communications, offering audio, video, and data conferencing functionality [7]. In NetMeeting, like most videoconferencing systems, the audio and video signals are encoded and transmitted using two separate TCP/IP sessions, and then reassembled at the receiving end. It is designed for use on the Internet and other IP networks. Video transmission requires camera and video capture card. NetMeeting uses the H.263 standard for video compression and, for our experiment, we used the Quarter Common Interchange Format (QCIF-176x144) frame size, as Common Interchange Format (CIF-352x288) provided an almost static picture.

2.2 Method

Ten pairs of volunteers from the university were involved in the test. The subjects had very little experience (if any) of using the software. Previous research implies that informal communication tends to be representative of individuals who are familiar with each other [8]. Hence, to maximise task motivation and to ensure subjects were fully at ease with each other, subjects who were acquainted with one another were selected and they were allowed to select their own issues for discussion.

The tests can be classified as below:

(i) Stage 1; Passive Communication (i.e. viewing and listening to a 'talking head', without interaction), was to benchmark the quality of: (a) Audio only, (b) Video only and (c) Audio and video overall.

(ii) Stage 2; Interactive Communication (where a variety of contents were introduced for example, intensive informal communications, moving object and using the shared workspaces, for example; sending and receiving large text files), was to benchmark the effects of content on the quality of: (a) Audio only, (b) Video only and (c) Audio and video overall.

In this stage, Stages (1) and (2) above were repeated, but tested under various network constraints. Different aspects of network congestion under investigation were packet lost, delay and delay jitter. Stages (1) and (2) were also conducted using two different audio CODECs (µ-Law, PCM and G723.1 6400bit/s). The video quality was unchanged throughout the test.

(iii) Stage 3 was to evaluate the performance of NetMeeting in a real network environment. The performance of two configurations was compared, operating firstly with no network constraints (Config. 1 - no packet loss, delay or delay jitter) and then subsequently with impediments introduced (Config. 2 - 5% packet loss, 400msec delay and 20msec delay jitter).

2.3 Apparatus and procedure

The subjects were seated in the two separate rooms provided with 15 inch monitors. For the experiments, Pentium III 933 MHz systems with128MB (CDRAM) were used and a USB video blaster Webcam Plus (capable of video capture up to 30 frames per second @ 352x288 pixels and 15 frame per second @ 640x480 pixels), was mounted on each monitor. To send and receive audio, two identical Platronic PC headsets were used. The audio channel was a 64kb/s full duplex.

The audio and video were transmitted from one end to another using a testbed containing the NIST Network Emulation Tool (NISTNet) - a general-purpose tool for emulating performance dynamics in IP network [9]. Each subject was provided with a self-view window, a remote view window and a talk window to send text to the remote partner.

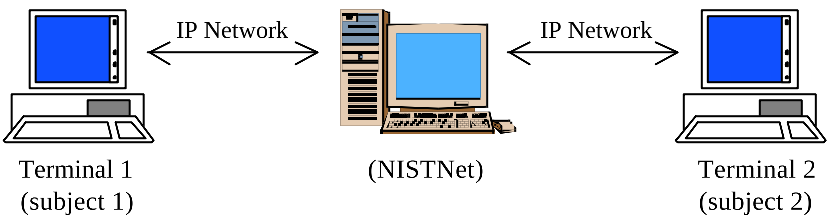

Figure 1: Testbed configuration

3. Result

Figure 3.1 shows loss effects on video. Video was found to be error free for packet loss below 3%, (MOS is between 2.9 and 3.5). At 4%, MOS drops to 2.6 or 2.7, and video degradation can be perceived. On approaching 8% to 30% packet loss, error in the video then became apparent and became unusable. Overall MOS for video, obtained by system using G723.1, shows lower results although video settings are not changed. This raises suspicions as to whether the change in audio quality would cause the change in perceived video quality and vice-versa.

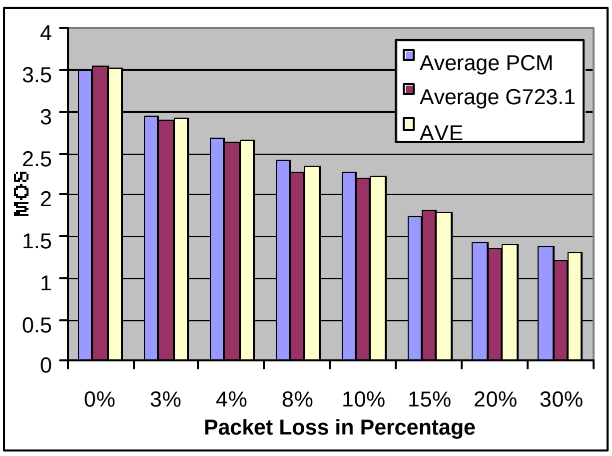

Figure 3.1: Packet Loss Effects on Video

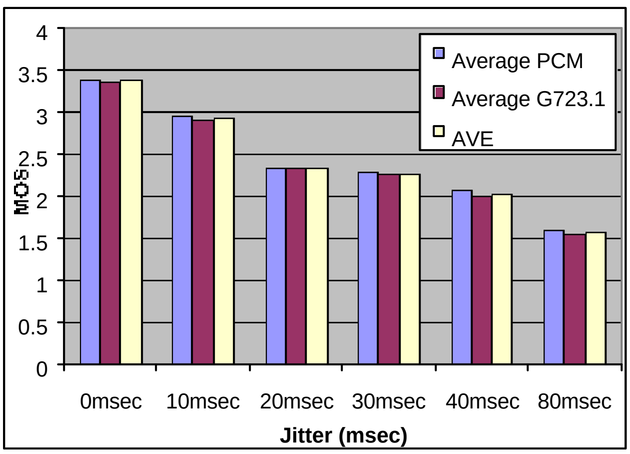

Figure 3.2: Jitter Effects on Audiovisual Overall

The MOS rating for loss effects on audio, for system that using PCM (µ-Law), is Good (MOS 3-4) at 0% to 8%. Whereby, the system employing G723.1 produced lower MOS (2.5) at 8% loss. This indicates that systems employing PCM (µ-Law) are more tolerant to packet loss. Packet loss on audiovisual overall is generally the same as that on video.

Figure 3.2 shows jitter effects on perceived audiovisual quality overall. Between 0msec-10msec, audiovisual quality is Good, i.e. MOS 3.4-2.9 for PCM and 3.3-2.7 for G723.1. At jitter between 20msec – 30msec, both systems score barely above Poor quality (i.e. 2.2-2.3 MOS). At 40msec jitter and above, the quality is Poor. For jitter effects on video, at 10msec, the scores are just below the Good (3) threshold, i.e. 2.81 (PCM) and 2.79 (G723.1), although the quality is still acceptable. At 20msec, MOS are 2.5 (PCM) and 2.4 (G723.1). At 30msec jitter, both systems produce Poor video. As expected, PCM gives better results, for jitter effects on audio (i.e. at 0msec–30msec, MOS are 2.7-3.9). At 40msec, audio started to degrade and became poor at 80msec.

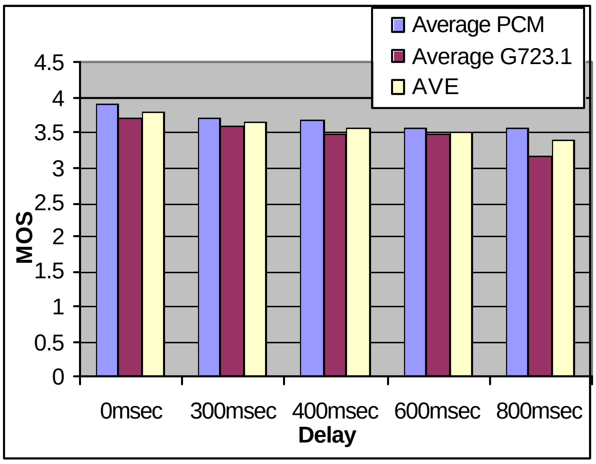

Figure 3.3: Delay Effects on Audio

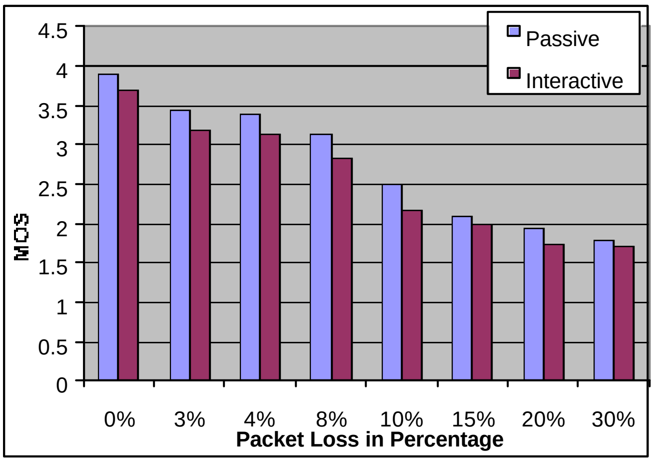

Figure 3.4: Packet Loss Effects on Audio Passive Vs Interactive Communication

Figure 3.3 shows delay effect on audio. As can be seen from the result, audio is less susceptible to delay as compared to packet loss and jitter. The MOS are around 3.2 to 3.7 for the range of 0msec to 800msec delays. Delay also has little impact on video. For example, for the range of 0msec to 800msec delay, MOS are around 2.9 and 3.6. The same pattern of results is repeated for audiovisual overall.

Figure 3.4 shows loss effects on audio, comparing the results obtained from passive communication and interactive communication, as suggested in Stage 2. It is evident that passive communication produces higher MOS than interactive communication. Results given by video quality and audiovisual quality overall follow the same pattern of results given by loss effects on audio, but with lower MOS.

Figure 3.5 shows the result of Stage 3. As expected, system Config. 1, as in Stage 3 produces much higher MOS (3.3-3.8). However, the overall score for System Config. 2, under congested network, is Poor (MOS 1.8-2.4). The overall rating for general performance of NetMeeting under ideal network is either Good (4) or Fair (3).

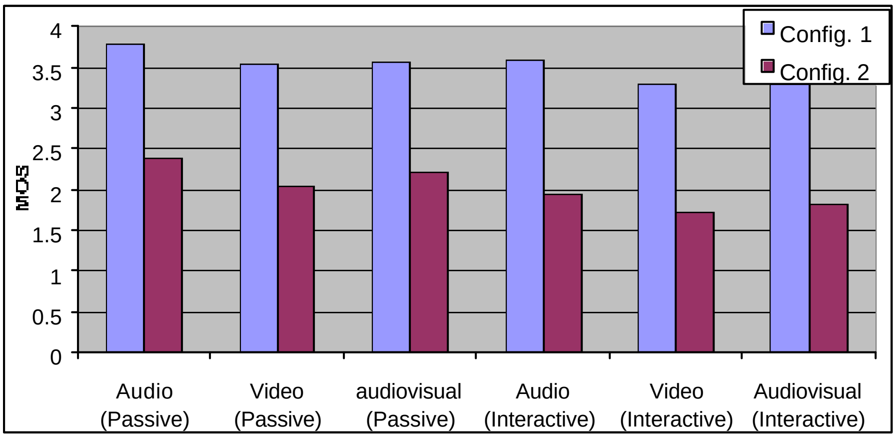

Figure 3.5: System Config. 1 vs System Config. 2

4. Discussion

Our observation indicates that, to assess video quality in videoconferencing is a very complicated issue since the frame rates are constantly changing. Subjects found it hard to give the score for video in a limited period of time. Some subjects felt dissatisfied in giving only one score, as their mind fluctuated between two or more scores from one moment to another during the test. This is due to the fact that the frame rate is high if the subject is relatively still, but as the movement of the subject increases, then either frame rates will vary inconsistently or the amount of video artifacts are likely to increase. It was observed that frame rate varies with the motion content, the level of detail, and the percentage of image that changes from one frame to another. However, assessing audio quality is less complex since the audio degradations are relatively significant as the network constraint increases.

When evaluating audio quality only, the MOS rating is high but when evaluating the combined quality of audio and video, the rating drops to almost similar to video rating. This implies that video quality contributes an important element in benchmarking the overall performance of desktop conferencing system.

By considering the result obtained in Stage 2 (refer to Figure 3.4), task performance has small effects on both audio and video, with the difference is only below 0.5 MOS. Our conclusion here, the task performance designed to carry out the assigned task was insufficient to obtain a more outstanding result. Another possibility is that the subjects were not fully trained and were therefore unable to perform the task exactly as required. In the future, each specific task performance must be carefully designed and the subject must be well trained so that the task will be conducted coherently in order to obtain a more reliable result. Generally, the result shows that the task performance effects on both audio and video became apparent for packet loss around 8% - 10%, for audio and around 4% - 10%, for video. Whereby, for jitter it is around 20msec and 40msec.

5. Conclusions and Future Work

Observations so far indicate that audio quality, video quality, and overall audiovisual quality are susceptible to packet loss and jitter, but are less susceptible to delay. Throughout the test, audio quality is higher when compared to video quality and overall audiovisual quality. Assessment of audio is very straightforward. Assessment of video, however, is very complex as its quality varied during the study, from very acceptable to almost useless.

Throughout the test, the best MOS rating for both audio and video is between Fair (3) and Good (4), although Good (4) MOS is seldom given. None of the subjects gave an Excellent (5) MOS rating. Thus, it is evident that NetMeeting or IP videoconferencing in general, is in its infancy with substantial improvements needed to achieve higher performance. We also learnt that traffic related network factors (such as packet loss, jitter, and delay), CPU power, CODEC, and task performance are all vital in maintaining the quality of video and audio service in NetMeeting.

The work presented in this paper will eventually lead to the characterisation of the factors that are necessary for audio and video optimisation in multimedia conferencing system. Future work will include, investigate more on issues such as lip synchronization. Currently, it is understood that lip synchronization is difficult to achieve in most low cost videoconferencing system due to the fact that the audio and video signals are transmitted via separate channels, and then reassembled at the receiving end. With this in mind, we will study a different approach in sending audio and video streams i.e. to combine them in the same packets. The work will then proceed to investigate the problems inherent in multi-modal transmission. A prototype system will be implemented in order to enable a baseline comparison against existing conferencing systems. The associated results will be the focus of future publications.

References

[1] Flanagan, J.L. (1965), Speech Analysis, Synthesis and Perception. Springer-Verlag, Berlin

[2] Gillian M. Wilson and M. Angela Sasse (2000), Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality. In S. McDonald, Y. Waern & G. Cockton (eds.) Proceedings of HCI200: People and Computers XIV - Usability or Else! Springer, pp. 327-339, September 5th - 8th, Sunderland, UK. ISBN 1-85233-318-9.

[3] A. Watson & M. A. Sasse (1996), Evaluating Audio and Video Quality in Low-Cost Multimedia Conferencing Systems. Interacting with Computers, Vol. 8 (3), pp. 255-275.

[4] Wilson, G. & Sasse, M.A. (2000), Investigating the Impact of Audio Degradations on User, Subjective vs Objective Assessment Methods. In C. Paris, N. Ozkan, S. Howard & S. Lu (eds.) Proceedings of OZCHI 2000:Interfacing Reality in the New Millennium, pp135-142, December 4th - 8th, Sydney, Australia. ISBN 0-643-06633-0.

[5] Hollier, M.P., Rimell, A.N., Hands, D.S., Voelcker, R.M. (1999), Multi-modal perception, British Telecom Res. Labs., Ipswich, UK BT Technology Journal, vol.17, no.1, Jan. 1999

[6] Bouch, A., Watson, A. and Sasse, M.A. (1999) QUASS – A Tool for Measuring the Subjective Quality of Real-Time Multimedia Audio and Video. Proceeding of HCI '98, 1-4 September 1998, Sheffield, UK

[7] Microsoft NetMeeting Home Page<URL: http://www.microsoft.com/windows/netmeeting/>

[8] Isaacs, E. and Tang, J. (1994), 'What Video Can and Cannot Do for Collaboration: A Case Study', Multimedia Systems 2, 6~73

[9] NIST Net Home Page <URL: http://snad.ncsl.nist.gov/itg/nistnet/ >