Distance Education via IP Videoconferencing: Results from a National Pilot Project

Anna Watson
Department of Computer Science
University College London
Gower Street
London, WC1E 6BT
a.watson@cs.ucl.ac.uk

M. Angela Sasse
Department of Computer Science
University College London
Gower Street
London, WC1E 6BT
a.sasse@cs.ucl.ac.uk

ABSTRACT

Internet Protocol (IP)-based videoconferencing technology can offer a low-cost means of collaboration and resource sharing on a national or global scale. This is potentially of interest to many users, especially in non-profit sectors such as education and healthcare. However, it has been questioned whether a best-effort network service can provide the reliability and quality required to support teaching and learning activities. To evaluate the technology, a 9-month pilot project of distributed teaching activities between 13 UK universities was set up. We present and discuss the issues involved in gathering and analysing data in a large-scale project with real users engaged in learning activities. The results suggest that incorrect equipment set-up and user behaviours cause most of the perceived problems, rather than network irregularities.

Keywords

IP videoconferencing, evaluation, field trials, audio/video quality.

INTRODUCTION

IP videoconferencing [1] offers the opportunity of low-cost collaboration and resource sharing on a national or even global scale. Since it allows point-to-point and multi-way interaction via audio, video and shared workspace tools, it makes possible the sharing of academic and educational resources between different institutions, and increases the ease with which differently located research groups can collaborate. Enthusiasts predict that its benefits for research and teaching will be on the same scale as email and the Web. Sceptics, on the other hand, question whether an IP-based - i.e. best-effort - service can provide the reliability and quality required to support these activities, particularly teaching at a distance.

In order to establish the feasibility and value of communicating in this way, it is necessary to conduct large-scale field trials using the technology between different sites and across different disciplines. Both subjective and objective data must be collected, to increase understanding of the end user requirements and the optimal network configurations. As the findings reported in this short paper indicate, laboratory-based simulations would not be sufficient to achieve this aim.

BACKGROUND TO RESEARCH

A research project, PIPVIC-2 (Piloting IP-based VIdeoConferencing) [2], set out to gain a greater understanding of the issues that may be encountered in running a large-scale IP videoconferencing service. Involving 13 academic institutions and 150 participants in a range of educational activities, PIPVIC-2 formed the largest collaborative project of its kind ever undertaken. The project was unique in its sheer scale and diversity, and collecting meaningful subjective and objective performance data presented a challenge.

Issues in evaluating IP-based videoconferences

Previous research has identified key issues and problems to beware of in collecting evaluation data in IP videoconferencing [3]. For example, network conditions can change rapidly, leading to packet loss and fluctuations in delivered audio and video quality during the course of one conferencing session. It is known that perceived quality is affected negatively by packet loss. However, there are other sources of 'objective' degradation, such as background noise, or problems caused by the hardware used, such as 'leaky' headsets or inadequate lighting. The impact of these objective factors is difficult to assess, especially since network conditions, hardware and set-up will all be different for different end-sites.

Conventional audio-video subjective quality measurements are taken at the end of a session, but when participants report that quality was "bad at times" over an hour-long session, to which times are they referring? How can we establish *when* quality is not good enough, and how can we know *what* is responsible for the perception of poor quality?

In PIPVIC-2 we took steps to address this issue by time-matching network statistics and subjective opinions. When registered opinion is poor, but network conditions are good, a better understanding of the effects of other objective factors can begin to be established.

OBJECTIVE AND SUBJECTIVE DATA COLLECTION

The audio and video conferencing tools used in the project were modified so that they logged reception reports gathered from other participants in a conference. These log files record factors such as packet loss and frames per second. This affords a means of recording (objectively) how a participant's media stream is received at other sites.

Subjective evaluation data was collected from some 150 participants in a range of activities encompassing small group working, tutorials, seminars and lectures, in various topics including Russian, Sociology, Art History and Business. The conferencing activities were mainly, but not exclusively, desktop rather room-based. Over 500 hours of videoconferencing experiences were evaluated in total.

Subjective data collection took place through

- Paper-based questionnaires that addressed perceived audio and video quality, and the adequacy of the quality for the task at hand.
- Specially designed web-based rating forms, which were necessary to achieve a more time-dependent measure of perceived quality: by ensuring that quality ratings were given at certain points during a conference, a closer match between subjective ratings and objective conditions could be derived.

KEY FINDINGS AND INTERPRETATION

Although further analysis is yet to be conducted on the wealth of data collected, the overall finding in the project was that the quality of the audio and video was mostly good, and adequate for the purposes of the session. Although relatively few participants reported problems, the problems that *were* reported are very revealing. Starting with the most commonly cited problem, these were:

- Missing words or incomplete sentences;
- The variation in volume between participants;
- The variation in quality between participants;
- Delay in the delivery of the audio signal.

Missing words would normally be attributed to the occurrence of packet loss, but through consultation of the objective reception reports, and cross-matching with the time that the web rating forms were completed, it is clear that in many cases missing words are *not* due to packet loss. The likely cause is therefore either hardware (a faulty microphone used by the speaker) or software (over-enthusiastic silence suppression in the audio tool).

The variation in volume and quality between participants are again not network factors, but due to hardware (inadequate headset or soundcard) or user behaviour (moving microphone away to cough, then not replacing it near the mouth; insufficient raising of the volume bar on the audio tool).

The delay in the delivery of the audio signal, long a major concern of network providers, was raised as an issue of concern comparatively infrequently. It is clear that many of the key usability issues in a future IP videoconferencing service will be focused at the end-workstations rather than in the network.

DISCUSSION AND RECOMMENDATIONS

Traditionally, concern over the feasibility of IP videoconferencing has focused on network issues such as packet loss and round-trip delay. With an increase in available bandwidth and greater reliability of networks, however, debilitating bursts of packet loss are becoming rare. Additionally, results from this pilot show that delay is not of such concern to end users compared to more mundane aspects, such as discrepancy in volume between participants, or the quality of the headset in use. The impact of these factors would not have been apparent in controlled, lab-based simulations of videoconferencing, but the combined approach of objective and subjective measurement undertaken in PIPVIC-2 has revealed their importance. The findings can now be used to inform the design of meaningful controlled studies to assess the relative weights of these factors.

The results from the project are highly encouraging for the future use of IP videoconferencing in educational arenas, but it is clear that user training and configuration/set-up support will be required. In addition, the technology would benefit from some additional functionality, such as automatic volume control and synchronisation between the audio and video streams. We advocate a combination of large-scale field trials and lab-based studies to further identify and improve the critical elements of successful IP videoconferencing.

ACKNOWLEDGEMENTS

PIPVIC-2 is funded by UKERNA, which manages the operation and development of the UK's Joint Academic Network. The willing participation of staff and students at the PIPVIC-2 project sites is gratefully acknowledged, particularly that of Rachel McEwan at Glasgow University.

REFERENCES

1. Deering, S.E. Multicast routing in internetworks and extended LANs. Proceedings of SIGCOMM '88 (Stanford CA, August 1988), ACM Press, 55-64.
2. http://www-mice.cs.ucl.ac.uk/multimedia/projects/pipvic2/
3. Watson, A. and Sasse, M.A. Measuring perceived quality of speech and video in multimedia conferencing applications. *Proceedings of ACM Multimedia'98* (Bristol, UK, September 1998), ACM Press, 55-60.