Low Bandwidth Video Conferencing

Flora Kundaeli

1. Abstract

This paper presents an overview of some of the factors to consider when developing a video conferencing (VC) application. Furthermore, it investigates previous work done on adapting VC systems for low bandwidth situations and gives a brief analysis of current online meeting tools and their limitations. It has been found that most VC systems are sensitive to bandwidth fluctuations and unstable connections. As a result, the video quality, audio quality, and audio-video quality have to be carefully selected to suit such circumstances. Moreover, not all meeting technologies work well, and can only support a few participants per meeting. The use of a server-to-client architecture, with many servers located all over the world, is currently the best architecture proposed for VC applications. Adding meeting facilitation tools, to web conferencing applications, is recommended, as it decreases the congestion at both the server and client ends.

2. Introduction

Business people, employers and employees alike spend a lot of their time at work participating in meetings. Post et al (2007) estimate that these meetings consume approximately 20-30% of their time. This has resulted in meetings becoming the focus of multidisciplinary research leading to the introduction of asynchronous meeting technologies (Reidsma et al., 2007). These technologies enable individuals from different locations to participate in and contribute to meetings. One such example of distributed meeting technology is video conferencing (VC), which offers the prospective of expanding participation in critical decisions, reducing transportation costs, and increasing the range of tasks that employees can accomplish without physically traveling to a central office (Emanuel et al., 1995). In a typical VC, a camera is mounted over a computer screen and an image of the remote person(s) with whom one is communicating is displayed on an on-screen window (Monk & Watts, 1995). Communication can then take place through various forms such as audio-video, video-chat or audio-video-chat.

Most existing VC tools, however, have shortcomings and limitations. Major well known issues involve dealing with unstable connections and low bandwidth. These conditions can greatly jeopardize the overall communication process, resulting in: distorted images; unsynchronized audio with image; or unclear audio delivery. Without good VC tools and effective meeting discipline, Web conferencing can become annoying, redundant, dysfunctional, unsatisfactory and waste people's time (Austin et al., 2006). This paper provides an overview of some of the factors to consider when developing a video conferencing (VC) application. It focuses on the video quality, audio quality and audio-video quality. Furthermore, it investigates previous work done on adapting VC systems for low bandwidth situations and gives a brief analysis of current online meeting tools and their limitations. It finally concludes by suggesting the best practices to employ when designing VC applications to support user's existing interaction skills for productive meetings, rather than require user's to adapt to the technology

2. Factors affecting video conferencing

The use of video in asynchronous meeting tools provides the users with a richer sense of presence and helps in the coordination of communication. It also facilitates emotional expression. Scholl et al (2005) explain that delivering high quality audio to larger groups remains a technical challenge since the available bandwidth has to be shared between users. Thus a larger group has less bandwidth for each person's video stream, imposing severe limitations on the quality and leading to a high required level of compression. This section takes a brief look at the following three factors that affect VC:

- video quality
- audio quality
- audio-video quality

2.1 Video Quality

There are many different factors that affect the video quality. However, only four are discussed in this paper. These include the codec; bitrates; frame rate and spatial resolution.

2.1.1 Codec

Carefully consideration should be taken when selecting which compression standard to use in order to avoid coding errors and problems in the video transmission. While using Microsoft's MPEG4 codec for compression in his experiment, Chen (2002) also mentioned three other low bandwidth video compression standards: the discrete cosine transform (DCT); the feature-outline; and the model-animation, which deliver usable video at less than 10kbps, 10kbps and 1kbps respectively. In addition, J. Kim and B.Kim (2011) stated that the MPEG4 Part-10 AVC/H.264 had become the new video coding standard approved by the joint video team (JVT) of ISO/ IEC MPEG and ITU-T VCEG. They argued that compared with other previous standards it had the ability to achieve a bit rate saving of more than 50% with the same quality.

2.1.2 Bitrates

Bitrates refer to the number of bits used to code a particular piece of data (bps). Jumisko-Pyykkö (2006) showed that higher bitrates of about 128kps, using XviD codec, resulted in better quality than bitrates as low as 80kpbs.

2.1.3 Frame Rate

The frame rate refers to the number of frames per second (fps). Bandwidth can be reduced by changing compression parameters or lowering the frame rate, which enables computational and bandwidth savings. However, reduction in the frame rate in some cases is problematic. Scholl et al (2005) state that when video is used only to provide a sense of presence, for example to identify basic emotions in a video chat, one frame every five seconds may be acceptable; however if complex emotions are to be portrayed, then 0.2fps will not suffice. An example is the Portholes project by Bly and Dourish (1992). They demonstrated that a frame rate of one update every five minutes could provide alertness in a work setting but may not be adequate for remote classrooms (Chen, 2002). His experiments showed that lowering the frame rate from 25 to 15 and 5 fps did not decrease a person's understanding of the content of the video and suggested that 5 fps may be the minimum required frame rate. However experiments have shown that video could still be useful at 1 fps. So clearly the choice of frame rate depends on the type of application to be designed.

2.1.4 Spatial Resolution (image size/picture ratio)

Spatial resolution refers to the number of pixels in each frame, or the dimensions of a frame in terms of height and width. Monk and Watts (1995) conducted an experiment to test the effect of image size on users. They had thirty two members of the general public work remotely from one another in pairs on some simple joint tasks. All the pairs had high quality audio links and were able to see one another's faces through an on-screen video image. For half the pairs this image was small (40 x 65 mm) and for the other half it was large (103 x 140mm). The conversations were analyzed and it was observed that the smaller video image resulted in formal and less fluent verbal interaction than the bigger image.

2.2 Audio Quality

Low bandwidth also affects the production of audio, resulting in a variety of audio distortions. Parts of the encoded audio may be deleted at low bitrates while coding high frequency content under low bandwidth situations (Jumisko-Pyykkö, 2006). He found that audio bitrates ranging from 16 to 24 kbps produced rewarding audio-video quality recommended the use of the AAC codec. Daly-Jones and Monk (1998) demonstrated through experimentation, the significant advantages of video conferencing over audio-only conferencing. They found that in the case of an audio-only channel, the absence of a visual channel resulted in less fluent conversations. These findings suggest the provision of a visual channel in addition to audio, for better group interactions.

2.3 Audio Video Quality (bitrate ratios)

Jumisko-Pyykkö (2006) observed that an audio-video bitrate of 24/76 kbps (6fps) was more pleasant than 16/84 (12.5fps). He noticed that when the overall quality reduced, viewers relied more on the audio information. His study also showed that the complexity of a scene also affected the relation of audio to video. The more complex a scene was, the more bits were needed for audio in low bitrates. Thus the scene complexity and the supposed clarity of the video affected the weight that was given to the audio channel. However finding the audio-video threshold ratio for different content might be vital for producing acceptable audio-video content with limited bandwidth.

3. Audio and video in meetings

In support of the above observations, Isaacs and Tang (1994) found that, compared with audio only, a video channel improved the ability to cooperate, express oneself and communicate more easily. They mentioned that the advantages of video depended critically on the nearly-instantaneous transmission of audio, even if it meant getting out of sync with the video image. On the other hand, when compared with face-to-face, it can be difficult in video interactions to control the floor, notice peripheral cues, easily point things out and manipulate real-world objects. Isaacs and Tang (1994) further suggested that, in order to fully enable rich interactions, video should be integrated with other distributed tools that enable natural collaborative behaviors within shared environments. Therefore one can opt to either use a video image with the audio (which is the preferred method) or just have clear audio depending on the technique one desires to employ.

3.1 Limitations of audio and video meetings

Isaacs and Tang (1994) indicated that, in face-to- face encounters, participants were able to more tightly coordinate their utterances, which improved their ability to reach mutual understanding faster than VCs. They stated that most video interactions do not allow participants to build on each other’s work, and manipulate real-world objects, nor do they allow users to look over each other’s shoulders to gain another perspective. However, according to Watts and Monk (1996), people were found to prefer viewing the face as opposed to having a view of the workspace. VCs also have a problem with resource sharing and switching to display the right person of interest in multiple video streams. A technique using video-follows-audio which automatically loads the current speaker into the audio window is used to deal with this problem (Scholl et al., 2005).

Another limitation has to do with floor control. A floor is a permission to access or manipulate a specific shared resource or a set of resources temporarily and floor-control refers to the mechanism that enables applications or users to gain safe and mutually exclusive or non-exclusive input access to the shared object or resource (Koskelainen et al., 2006). In Chan's (2004) turn-taking protocol, a user could be in three states: either in control of the shared resources or application; waiting for control; or observing their collaborator’s activities and actions. A user in control had privileges over the resource and could release it. Once released, a resource was passed on to the first user (users were queued) who had requested it, who then took control over the shared resource.

Another problem, revealed in a different study by Isaacs and Tang (1994) was that it was impossible to direct attention toward a specific person in a multi-way conference. Everyone sees you through the same camera, so if you are looking at one person’s video image, it appears to everyone as if you are looking at all of them. They observed that people tended to use each other’s names to address each other in such situations. To add onto this, the sharing of a single audio channel limited the ability of people to conduct side conversations, and pointing did not work either as it was difficult for the others to use spatial position to figure out who was being addressed. Pointing only worked when one wanted to focus attention on certain parts of their own environment.

4. Improvising for low bandwidth

This section gives a brief description of three experiments that were designed to deal with low bandwidth situations. Techniques to maximize the benefits that could be achieved for VC systems in such situations can be learned and adopted from these experiments. Chen (2002) did a gesture detection VC experiment using three different frame-rate: full-motion, gesture sensitive, and low-update. In gesture sensitive, the video image was updated on detection of a hand being raised. His data revealed that conveying postures alone (low update) was insufficient for small group discussions due to difficulties with floor control. However conveying gestures in addition to postures was a viable option if limited bandwidth would otherwise prevent using videoconferencing at all. Another study, by Takao (1999), examined the effects of: face to face meetings (FF); switching video (SV), which showed only the current speaker; and mixing video (MV), which showed each group member simultaneously. His results revealed that MV yielded better group decision quality than FF and that SV and MV showed no difference. Scholl et al (2005) used text chat to complement the video (video-chat) as opposed to the common audio-video version of communication. The video would display an image of the current chatter (video follows chat). Their findings revealed that most of the users found the application useful. However it would be more beneficial to incorporate video, audio and text chat in designing efficient adaptive VC systems to operate in low bandwidths conditions, so that when one feature is unavailable due to the low bandwidth, another feature can be used to balance and continue communication.

5. Comparison of some existing VC tools

Olson et al (2003) mentioned that the top four players in the Web conferencing market were WebEx, followed by PlaceWare, Raindance and Genesys. Chan (2004) also found that WebEx was considered the industry leader, as compared to other tools such as Microsoft LiveMeeting and Macromedia Breeze. Xu et al (2008) performed a comparison of ten existing asynchronous meeting tools. In their findings, PHProjekt supported the most features, followed by Microsoft Sharepoint Server, eGroupware, WebEx WebOffice and ZOHO Project among others. Other popular online chatting applications like Skype, MSN, Yahoo messenger, and Google talk can only support multi-party audio conference and a limit of 2-party video conference (Lu, et al ., 2010).

6. Conclusions and Design Implications

It has been shown that various tools and techniques can be employed to improve meeting effectiveness, and videoconferencing systems. In addition VCs are sensitive to bandwidth fluctuations and unstable connections. Thus specific factors such as codec, bitrates, frame rates, audio quality, video quality and audio-video quality have to be carefully chosen to suit the user's preferences and ensure pleasant use of the software tool. Furthermore most Web conferencing applications, in general, cannot provide very good quality to their end users and support a limited number of participants per meeting (Lu et al., 2010). Thus none of the applications work well. Lu et al (2010) suggest the use of a server-to-client architecture with many servers located all over the world as currently the best architecture for VC applications. In addition Austin et al (2006) talks about adding meeting facilitation tools known as Group Decision Support Systems (GDSS) as a supplement to VC over the Internet, as this introduces the least congestion at both the server and client ends.

References

Austin, T., Drakos, N., & Mann, J. (2006). Web conferencing amplifies dysfunctional meeting practices. *Gartner Research Report*, (p. 1-7). Gartner .Retrieved April 26, 2011, from http://data.vitusbering.dk/vbi/isi/Gartner-web_conferencing_amplifies_d_138101.pdf.

Bly, S., & Dourish, P. (1992). Portholes : Supporting Awareness in a Distributed Group Work. *Computing*, (p. 541-547). ACM. Retrieved April 26, 2011, from http://portal.acm.org/citation.cfm?id=142982

Chan, A. (2004). *Designing Haptic Icons to Support an Urgency-Based Turn-Taking Protocol*. *October*. (p. 1-159).

Chen, M. (2002). Achieving effective floor control with a low-bandwidth gesture-sensitive videoconferencing system. *Proceedings of the tenth ACM international conference on Multimedia - MULTIMEDIA '02*, (p. 476-483). New York, New York, USA: ACM Press. doi: 10.1145/641108.641109.

Daly-Jones O & Andew Monk, L. W. (1998). Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies*, *49*(1), (p. 21-58). doi: 10.1006/ijhc.1998.0195.

Emanuel, H., Niederman, F., & Shapiro, S. (1995). Online services as distributed meeting support software. *Proceedings of the 1995 ACM SIGCPR conference on Supporting teams, groups, and learning inside and outside the IS function reinventing IS* (p. 213–222). ACM. Retrieved April 26, 2011, from http://portal.acm.org/citation.cfm?id=212601.

Isaacs, E. A., & Tang, J. C. (1994). What video can and cannot do for collaboration: a case study. *Multimedia Systems*, *2*(2), (p. 63–73). Springer. Retrieved April 27, 2011, from http://www.springerlink.com/index/k6u5j47555154187.pdf.

Jumisko-Pyykkö, S. (2006). "I would like to see the subtitles and the face or at least hear the voice": Effects of picture ratio and audio–video bitrate ratio on perception of quality in mobile television. *Multimedia Tools and Applications*, *36*(1-2), (p. 167-184). doi: 10.1007/s11042-006-0080-9.

Kim, J.-H., & Kim, B.-G. (2011). Efficient intra-mode decision algorithm for inter-frames in H.264/AVC video coding. *IET Image Processing*, *5*(3), (p. 286-295). doi: 10.1049/iet-ipr.2009.0097.

Koskelainen, P., Ott, J., & Schulzrinne, H. (2006). *X. Wu,“ Requirements for Floor Control Protocols*. Retrieved April 27, 2011, from http://tools.ietf.org/rfc/rfc4376.txt

Lu, Y., Zhao, Y., Kuipers, F., & Van Mieghem, P. (2010). Measurement study of multi-party video conferencing. *NETWORKING 2010* (p. 96–108). Springer. Retrieved April 26, 2011, from http://www.springerlink.com/index/8X27W207513N2312.pdf.

Monk, A. F., & Watts, Leon. (1995). A poor quality video link affects speech but not gaze. *Conference companion on Human factors in computing systems* (p. 274–275). ACM. Retrieved April 26, 2011, from http://portal.acm.org/citation.cfm?id=223671.

Olson, G., Parham, N., Puetz, M., & Shami, N. (2003). FINAL REPORT EVALUATING THE WEB CONFERENCING TOOL CENTRA. *SI 689: Winter 2003 Computer Supported Cooperative Work*. (p. 1-40).Retrieved April 27, 2011, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.133.195&rep=rep1&type=pdf.

Post, W. M., Huis in ’t Veld, M. a a, & Boogaard, S. a a. (2007). Evaluating meeting support tools. *Personal and Ubiquitous Computing*, *12*(3), (p. 223-235). doi: 10.1007/s00779-007-0148-1.

Reidsma, D., Akker, R., Rienks, R., Poppe, R., Nijholt, A., Heylen, D., et al. (2007). Virtual meeting rooms: from observation to simulation. *Ai & Society*, *22*(2), (p. 133-144). doi: 10.1007/s00146-007-0129-y.

Scholl, J., Parnes, P., McCarthy, J. D., & Sasse, A. (2005). Designing a large-scale video chat application. *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA ’05* (p. 71-80). New York, New York, USA: ACM Press. doi: 10.1145/1101149.1101160.

Takao, S. (1999). The effects of narrow-band width multipoint videoconferencing on group decision making and turn distribution. *ACM SIGSOFT Software Engineering Notes*, *24*(2), (p. 109-116). doi: 10.1145/295666.295678.

Watts, L., & Monk, A. F. (1996). Remote assistance: a view of the work and a view of the face? *Conference companion on Human factors in computing systems: common ground* (p. 101–102). ACM. Retrieved April 26, 2011, from http://portal.acm.org/citation.cfm?id=257181.

Xu, J., Zhang, J., Harvey, T., & Young, J. (2008). A survey of asynchronous collaboration tools. *Information Technology Journal*, *7*(8), (p. 1182–1187). Retrieved April 27, 2011, from http://docsdrive.com/pdfs/ansinet/itj/2008/1182-1187.pdf.