Designing a large-scale video chat application

Jeremiah Scholl
Department of Computer Science and Engineering, Luleå University of Technology &
Norwegian Centre for Telemedicine
jeremiah.scholl@telemed.no

John D. McCarthy
Department of Computer Science
University College London
j.mccarthy@cs.ucl.ac.uk

Angela Sasse
Department of Computer Science
University College London
a.sasse@cs.ucl.ac.ukl

Peter Parnes
Department of Computer Science and Engineering
Luleå University of Technology
peter.parnes@ltu.se

ABSTRACT
Studies of video conferencing systems generally focus on scenarios where users communicate using an audio channel. However, text chat serves users in a wide variety of contexts, and is commonly included in multimedia conferencing systems as a complement to the audio channel. This paper introduces a prototype application which integrates video and text communication, and describes a formative evaluation of the prototype with 53 users in a social setting. We focus the evaluation on bandwidth and view navigation requirements in order to determine how to better serve users with *video chat*, and discuss how the findings from this evaluation can inform the design of future video chat applications. Bandwidth requirements are evaluated through user perceptions of video delivered using three different bandwidth schemes. For view navigation, we examine a system that automatically switches the video focus to the current "chatter", instead of requiring users to navigate manually to find the video steam they are interested in viewing.

Categories and Subject Descriptors
H.4.3. [**Communications Applications**]: Computer Conferencing, teleconferencing and videoconferencing. H5.m. [**Information interfaces and presentation**] (e.g., HCI): Miscellaneous.

General Terms
Performance, Design, Experimentation, Human Factors,.

Keywords
Chat, video conferencing, bandwidth sharing, collaboration

1. INTRODUCTION
Over the past few years, chat and instant messaging (IM) systems have become popular, not only with home users, but also as a communication tool within the workplace [10,12,21]. Despite being viewed as a 'media-poor' [25] form of collaboration, text-based systems continue to offer users several advantages. For example, they support a nice balance of both synchronous and asynchronous communication [11] and have shown to be effective for supporting unplanned, informal communication [14]. For these and other reasons, multimedia conferencing applications, such as Alkit Confero [1] and Marratech Pro [18], have been developed that give users the option of communicating via text chat, instead of only providing them with an audio channel. This allows each media to be used in the variety of situations where it is deemed most appropriate. For example, audio might be used for conducting a formal meeting and chat can be used for short impromptu discussions, similar to the way it is utilized in less media rich environments.

Video is generally used in these applications in order to provide a richer sense of presence [25,32], help coordination of communication [5,15] and facilitate emotional expression [8,23]. However, delivering high-quality video to larger groups remains technically challenging, since the available bandwidth has to be shared between users. Thus, the larger the group, the less bandwidth is available for each person's video stream, a fact that imposes severe limitations on quality. Another problem with larger groups is navigation of multiple video streams. With many video streams displayed on the screen, it is unclear, at any one time, which video stream is the important one. To tackle this problem, video conferencing systems need an effective technique for *view navigation* [10], to bring into focus the person of interest at all times.

In general, solutions to these and other problems related to video conferencing have been explored in a context where users communicate using audio. We take a complementary approach and have developed a prototype *video chat* application to investigate scenarios where users communicate via text chat. Our goal is to help designers improve multimedia conferencing applications by seeing how well certain requirements and strategies for video conferencing hold up when used in a *video chat* setting.

Bandwidth requirements were explored by implementing three different schemes for bandwidth allocation and video delivery. We addressed the view navigation problem by implementing a feature adapted from video conferencing, which we call "*video follows chat*". User responses to and perceptions of these features are examined in a subsequent formative evaluation.

We begin by describing the design decisions we faced during initial development of our prototype. The design of the evaluation, which was conducted with 4 groups consisting of 53 participants, and the qualitative and quantitative results follow. Finally, we discuss the implications of our findings for designers and researchers, and how they can be used to improve video chat applications.

2. PROTOTYPE DEVELOPMENT

2.1 Bandwidth Requirements

When a video conferencing application is faced with limited bandwidth supply there are less network resources available per person as group size grows. The large amount of progress that has been accomplished in the areas of efficient video coding and data transport mechanisms such as IP-Multicast and Application Layer Multicast can be leveraged in order to deliver higher quality video at a given bandwidth level, and thus reduce this problem. The default approach for dividing the supply, which is independent of the video codec and data transport mechanism, is to equally share the session bandwidth between users (referred to as *equal bandwidth sharing* for the rest of this paper). Bandwidth share for video can be reduced by changing compression parameters or by lowering the frame rate, which has the additional advantage of offering computational savings (less frames to decode) as well as bandwidth savings. However, in some contexts a reduction in the frame rate can be problematic.

An extreme example of this is when users are communicating via sign language, where at least 21 frames-per-second (fps) is recommended in order to support finger spelling [12]. In contrast, the frame rate requirements for audio/video conferencing are less strict, with various studies suggesting 5 fps as a minimum value [4]. Tang & Isaacs reported that people rate 5 fps as tolerable [26], and Watson & Sasse found that audio and video are not perceived as synchronized at less than 5 fps [30]. Studies of user behavior in video conferencing report no difference in task outcome and only slight differences in communication behavior when the frame rate is alternated between 5 fps and 25 fps while users design a tourist poster [15] or solve a jigsaw puzzle [20].

When video is used simply to provide a sense of presence, the bandwidth requirements may be much lower. The Portholes project for example, showed frame rates as low as 1 frame every 5 minutes to be adequate for providing distributed teams with a sense of group presence [7],

The requirements for *video chat* users have not been evaluated, and thus there is a lack of useful information for designers. The two questions we faced were:

(1) What are the minimum bandwidth requirements for video chat?

(2) How can we maximize video quality for video chat in large groups?

2.1.1 Minimum Bandwidth Requirements

Chat users are engaged in synchronous communication with each other, so *video chat* may have similar frame rate requirements to video conferencing. However, there are several important differences between video chat and video conferencing that may change user requirements. With video conferencing, users watch the video while listening to the audio at the same time. This is not the case with video chat. Both text chat and video are a visual medium, and will compete for the users' attention on the screen. This implies that users will focus less on the video when chatting than during audio/video conferencing, since they will be occupied reading chat messages and looking at the keyboard (unless they are a touch typist).

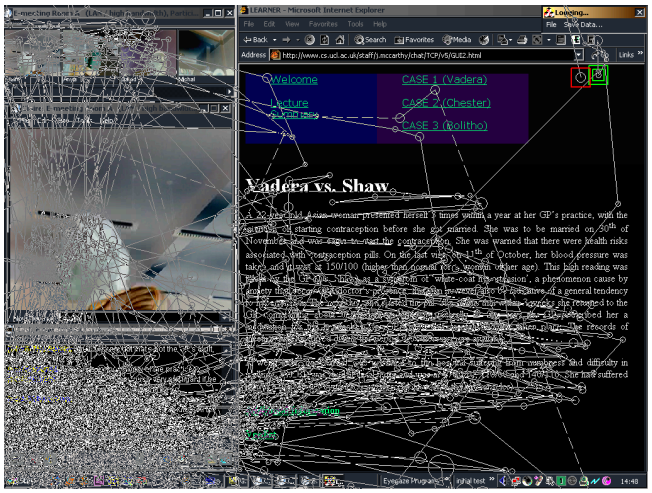

Figure 1: Gaze distribution while using video chat in an e-learning session.

Support for this argument is provided by a recent eye tracking study of small scale video chat in an educational setting [27] that shows that people spend around 70% of the time looking at the chat window but only 10% of the time looking at the video window. An illustration of the eye movements from this study is presented in Figure 1. As shown in the figure, gaze distribution is clearly much higher in the chat window, which is at the bottom left hand of the screen, than it is for the video window, which is placed directly above the chat window.

Also, because people are typing rather than speaking, the synchronization of facial movements with an audio stream may not be as much of an issue. Therefore, the 5 fps limit may not apply. Two types of information that are important for video chat are basic presence information and emotional expression. The bandwidth requirements for presence are known to be very modest but requirements for emotion recognition are less clear [23].

Drawing on emotion recognition research [9,23], we derived an estimate that one frame every five seconds (0.2fps) would be near the lower bound for emotion detection. El Kaliouby et al. [9] assembled video samples of naturally expressed emotions, and report that the average length of these emotions was approximately 5 seconds varying between 3 and 7 seconds. When people were presented with a sub-sample of 1 second of video from these clips, they could recognize simple emotions, (happy, sad), but were very poor at recognizing complex emotions such interest, boredom, and confusion. On the basis of these findings we predicted that with only a single frame from a 5 second period (0.2 fps), complex emotion recognition would be effectively blocked.

The effects of delivering video at low bandwidth are examined in the evaluation. If presence and basic emotional recognition (happy, sad) is all that is required for video chat, then one frame

every five seconds may be perfectly acceptable to users. However, if users need to identify more complex emotions then this bandwidth (0.2fps) will be unacceptable.

2.1.2 Maximizing Video Quality

When sharing bandwidth *equally* between participants, video quality will become very poor at some point. For example, the expected bandwidth/person is just 25 kbps for a 400kbps session with just 16 members. One strategy to mitigate the limitations on *equal bandwidth sharing* is to define certain users in the group as more important than others, and give them a larger share of the available resources. The rationale behind this strategy is that human communication patterns are generally uneven, and therefore result in users focusing more of their attention on some group members than others. If this focus can be detected then users will be better served by allocating more resources to video streams in the focus of attention.

The most basic version of *unequal bandwidth sharing* in video conferencing is Jacobson's *video silence suppression* [16], where a video stream is turned off if there are no receivers interested in viewing it. More sophisticated bandwidth schemes have also been explored that seek to fine tune the amount of bandwidth each sender uses, instead of just turning video-streams on or off [2,4,24]. This has been shown to be effective for providing floor control to e-learning participants over a limited bandwidth connection (modem) by allocating them an extra frame when they raise their hand to speak [4]. Others have investigated technical details of how to implement more general *unequal bandwidth sharing* schemes, which could improve video quality in a variety of situations and at a wide variety of bandwidth levels [2,24]. However, no user studies have been conducted to demonstrate their effectiveness.

We wanted to investigate the effects of a scheme of this type. If successful the technique could be used to deliver the experience of high quality video to large groups of users (15-50), since the bulk of the available bandwidth could be allocated to the subset of video streams that the users are attending to.

The client we developed had two types of video windows. Small *thumbnails* for each participant and a single large *focus* window (see Figure 2). Whenever a user clicked on a thumbnail, that video stream would be loaded into the focus window. To implement unequal bandwidth sharing, we used the contents of each client's focus window as the metric for user importance. Thus, if a particular user is in one or more clients' focus window, then that user is considered to be important and therefore needs more bandwidth.

The scheme operates by having each client send a message to the rest of the group when the contents of its focus window are switched from one group member to another. These messages are used by senders to adapt their bandwidth consumption by first allocating a minimal level of bandwidth to everyone, with the remaining bandwidth divided evenly among the "important" senders (defined as senders that appear in at least one client's focus window). Senders react *proactively* when they are "unimportant" (i.e. when they do not appear in anyone's focus window) by dropping their bandwidth usage to 1 frame every 3 seconds. This allows "important" senders to increase their bandwidth consumption *reactively* by measuring incoming bandwidth usage, and increasing their bandwidth consumption to "fill the gap" created by unimportant senders that do not use their "normal" share.

In the evaluation we examine whether such *unequal bandwidth sharing* increases perceived video quality over and above that for *equal bandwidth sharing*.

2.2 View Navigation and Video Follows Chat

Another challenge when delivering multiparty video is how to provide users with an adequate view of the available video streams within a limited screen space. A common technique is "click to focus" navigation, where clicking on a thumbnail loads that stream in to the focus window (see Figure 2). While this works well in many situations, it requires users to actively seek out and click on each person they want to view in more detail. This can be extremely tedious during an active discussion, where the current speaker constantly changes. To tackle this problem, view switching is usually automated during audio conferencing using a technique called "*video follows audio*" (also called "voice activated switching" [28]), which operates by automatically loading the current speaker into the focus window. This has shown to be successful with video conferencing but also has a reputation for being problematic because at times it can be difficult for systems to correctly predict the person of interest. For example, ambient noise may cause unwanted switches to occur that are unrelated to the flow of conversation.

We wanted to evaluate a similar system which we call *video follows chat* in order to see if it could offer some gain to video-chat users. With *video follows chat*, a person automatically appears in the focus window whenever they send a chat message. For this reason we will refer to the last person in the group to send a chat message as the *focus user*. *Video follows chat* offers two potential advantages in comparison to *video follows audio* in that problems related to ambient noise will not occur, and that freeing chat users from the need to click on video thumbnails will allow them to keep their hands on the keyboard in "chat position".

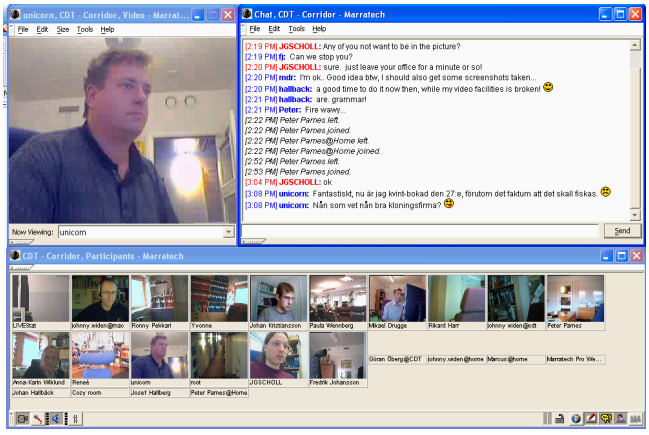

Figure 2: A screen shot of the prototype with the window arrangement used in the evaluation.

Video follows chat works in tandem with the *unequal sharing* scheme, so when users send a message they are shown in the *focus* window and also increase their bandwidth share. However, in implementing this feature we were again presented with a set of design decisions.

(3) When implementing video follows chat, should the video switch to a person when they start typing their message or after they have sent it?

(4) Should the people see themselves when they send a chat message?

Rather than explore this design space in testing all possible variations, we decided to make a particular set of decisions and based on a rationale or model of usage. The impact of these decisions would be examined in the subsequent evaluation.

2.2.1 (3) Switch Before of After?

The video channel can improve both coordination and emotional communication [15,23]. If the video switches to a person as they start to type, the video channel might improve coordination, serving the same function as the "X is typing" feedback implemented in popular IM applications [14,21]. However, we reasoned that watching people type could be tedious, and was unlikely to include much emotional communication, as most users would have their attention focused on the keyboard while typing. We predicted that emotional expression was more likely *after* they had sent a chat message and were anticipating a response. On this basis we decided to switch the video to a person immediately after they sent a message.

2.2.2 (4) See yourself?

V*ideo follows audio* systems typically do not switch the contents of a user's own focus window to themselves when they speak. The rationale for this choice is that users benefit more from continuing to view the previous speaker, as this allows them to judge reactions to their own speech, than they would from looking at their own video stream. One exception to this rule however has been noted by [6] who point out that new users may benefit by having a "confidence monitor" so they can see how others see them.

As we did not have any data either way, we decided to implement the standard *see others* setting to mimic the experience users get today from *video follows audio* and evaluate it's impact with new users. Thus, in our implementation, when a person sent a chat message, they would appear in everyone's focus window *except their own.* The message sender would see the person who sent the chat message before theirs.

3. FORMATIVE EVALUATION

The four issues addressed in the evaluation study were:

- Whether users felt they used the visual channel at all and if so what they communicated.
- User perceptions of video quality when delivered at the *low bandwidth*. [Decision (1)].
- User perceptions of video quality with *unequal bandwidth sharing* [Decision (2)].
- Their qualitative and quantitative rating of *video follows chat* with the design decisions we had implemented for this feature [Decision (3)-(4)].

As we were interested in the communication of emotions, we tested people in a social scenario where people had to introduce themselves. We expected that this would maximize interest in both the video stream and expressed emotions and would therefore be a strong test of bandwidth requirements. Also, we chose to model the real-life scenario of personal introductions because they typically take place in a wide variety of settings, both inside and outside the workplace and this would increase the ecological validity of our results over using a contrived task.

In each of the 4 chat sessions, 14-16 people used video chat to introduce themselves. The introductions where split into four rounds. In half the rounds, bandwidth was *equally shared* between participants, in the other half video was delivered in one of two conditions, either at the *low bandwidth* or *unequally shared* between participants.

After each round, participants completed a short questionnaire. To measure their perception of video quality, three different measures were taken.

I. An unlabelled 100 point scale [31]

II. Likert scales based on user descriptors of video quality [29]

III. A binary measure of acceptability [19]

To evaluate *video follows chat*, we probed whether they found the feature (a) useful or (b) annoying, and collected qualitative comments from users after each round. At the end of the evaluation we gave people a short questionnaire on whether they found the visual contact useful and whether they had used it to communicate.

3.1 Method

3.1.1 Participants

53 people participated in the study. Their average age was 26. 84% had previous experience with instant messaging, 70% had experience with chat rooms and 28% had experience with video conferencing. They were recruited from subject pools within XXXXX University and were paid $15 for participation. 26 were allocated to groups comparing *equal bandwidth sharing* with *low bandwidth*. 27 were allocated to groups comparing *equal* with *unequal bandwidth sharing*. In each session there were also two facilitators to guide people through the introductions and inform them when the rounds were over. Two groups were run in each condition. Table 1 summarizes the basic demographics of the groups.

Condition	Group (N)	M	F
low vs. *equal*	Group A (12)	7	5
	Group B (14)	5	9
unequal vs. *equal*	Group C (14)	7	7
	Group D (13)	10	3

Table 1. Sex distribution of each group in the study.

3.1.2 Equipment and Software

For our experiments, we modified a version of Marratech Pro, a commercially available multimedia conferencing tool. The commercial version of Marratech Pro uses an *equal bandwidth-sharing* scheme. We modified the client so that we could remotely

set the frame rate and bandwidth-sharing scheme actively used by each client during our experiments.

Marratech clients are used in conjunction with the Marratech E-meeting Portal (a license server and media gateway) to set up multimedia conferencing sessions. For the study we used an evaluation version of the portal, which limited bandwidth usage for video to 400 kb/s. Users were provided with flat screens, Logitech QuickCam Pro 4000 cameras, and Pentium 4 machines with 256MB memory.

3.1.3 Procedure

Before the start of the study, all participants completed a questionnaire that probed basic demographic information and the participants existing experience with chat rooms, instant messaging (IM) and video conferencing.

They were then told that they were here to evaluate a new communication tool, which they used to introduce themselves to each other in four different rounds. At the start of the first round the facilitators introduced themselves to reduce the uncertainty of what was expected and to help relax the participants. Different people introduced themselves in each round, as indicated by the facilitators, and at the end of each round they completed the questionnaire provided.

At the end of the session participants were given a final questionnaire to understand how they used the video channel and what they tried to communicate. Finally, they were debriefed about the precise nature of the study and given their participation pay.

3.1.4 Design

Table 2 summarizes the design for the evaluation. The different bandwidth schemes where alternated across the different rounds of chat. The order of conditions was counterbalanced across the groups.

Group (N)	Round 1	Round 2	Round 3	Round 4
A (12)	Equal	Low	Equal	Low
B (14)	Low	Equal	Low	Equal
C (14)	Equal	Unequal	Equal	Unequal
D (13)	Unequal	Equal	Unequal	Equal

Table 2. The study design for the evaluation.

4. RESULTS

To simplify the graphical presentation of the results the *Equal* ratings across conditions are combined into a single bar for all the figures subsequently presented. The t-tests however are conducted within each condition (A+B & C+D respectively).

4.1 Physical Video Quality

The bandwidth share of the *focus user* for each of the three schemes is shown in Figure 3. The *low bandwidth* scheme uses just 3.3kbps to deliver one frame every five seconds, *equal sharing* secured a bandwidth of 26.8kbps and for *unequal sharing* the mean bandwidth was 44.3kbps.

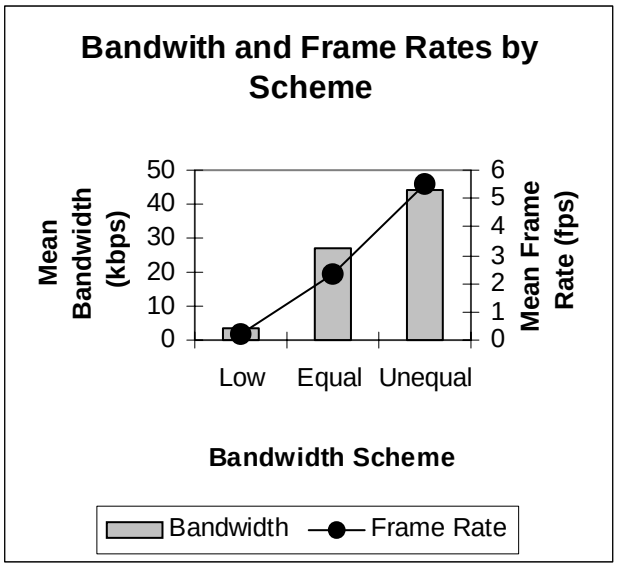

Figure 3. Bandwidth and frame rates for the *focus user* during each of the 3 conditions in the study

When translated to frame rates, *low bandwidth* was fixed at 0.2 fps, *equal bandwidth sharing* corresponded to a frame rate of 2.3 frames/second and the for *unequal bandwidth sharing*, the average was 5.5 frames/second.

4.2 Perceived Video Quality

On the 100-point rating scale there were significant differences between the bandwidth conditions. Participants rated *low bandwidth* significantly lower than *equal sharing* [$t(24)=3.86$, $p < 0.01$]. However, although the rating for *unequal sharing* was slightly higher than for *equal sharing*, this difference was not statistically significant [$t(23)=-1.32$, ns.]

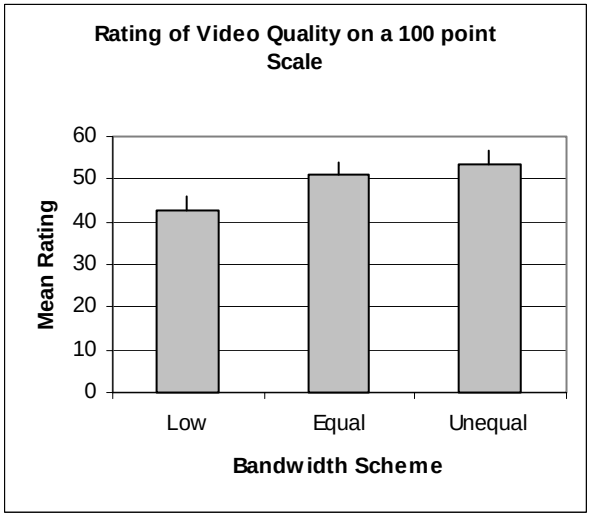

Figure 4 Quality rating on an unlabeled scale.

Differences were also found on the user descriptor scales of video quality. Relative to *equal bandwidth sharing*, users thought the video was less fast and smooth with *low bandwidth* [$t(24)=-3.80$, $p < 0.01$]. There were no differences in the perceived in picture clarity but, relative to equal sharing, video was seen as more slow

and jerky with the *low bandwidth* scheme, [t(24) = 2.32, p < 0.05], and *less* slow and jerky when bandwidth was allocated *unequally* [t(23) = -2.61, p < 0.05].

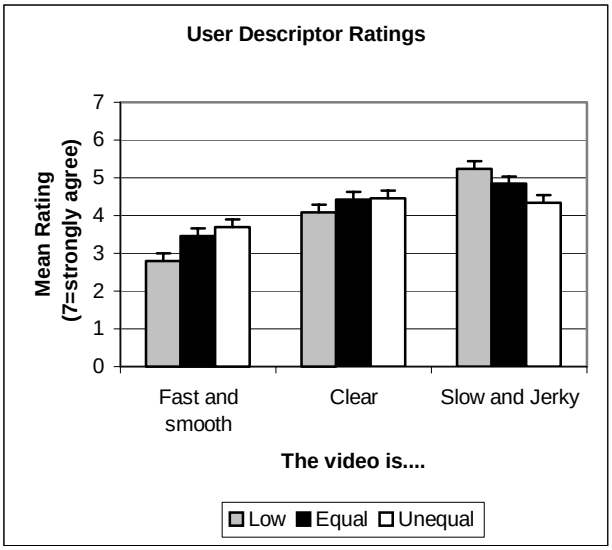

Figure 5. Unequal sharing is less slow and jerky.

Finally, on the measure of acceptability, a score of 2 indicates that the video quality was acceptable all of the time. A score of 0 means that quality was never reported as acceptable. On this measure, acceptability was lower with the *low bandwidth* scheme than with the *equal sharing* scheme [t(23) = 4.00, p < 0.01]. There were no differences in acceptability between the *equal* and *unequal sharing* conditions, [t(26) = 0.44, ns.].

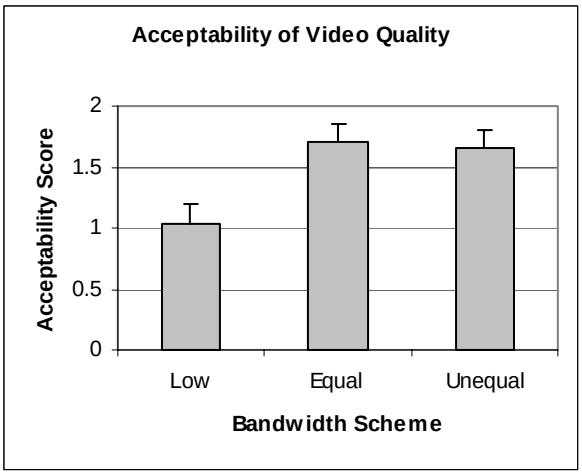

Figure 6. Low Bandwidth is unacceptable to users.

4.3 Visual Contact and Visual Communication

The qualitative data reinforces the interpretation that emotion recognition is important in video chat. 76% of subjects agreed that visual contact was useful and 45% of these *strongly agreed* that they "*used the video channel to communicate visually very often*".

When asked an opened-ended question about what *was* communicated, most people gave responses that mentioned emotions and facial expressions. Of the 34 comments recorded 25 (73%) referred to emotions and facial expressions. Interestingly, another 4 people made comments relating to self-consciousness and their attempts to mask emotions and body language. 3 people simply appreciated seeing who they were chatting with and putting names to faces. A sample of comments is shown in Table 3.

"Describe shortly what you tried to communicate through the camera"
"Smiling to communicate the tone of voice"
"Facial Expressions - explains better than words"
"Surprise and amusement at jokes. Emotions mainly"
"Aware of how I looked and more conscious of touching my face. Nice to see the others though"
"Was conscious of smiling - problem is that you can't pretend to be a 24stone ex-body builder from the Philippines"
"Good to put a name to a face. Don't have to be contrived – just see their natural proper character".

Table 3. User comments on what was communicated over the video channel.

4.4 Video Follows Chat

Participants gave different ratings on the usefulness of *video follows chat* in the different groups. Specifically, those who experienced both low and equal bandwidth schemes thought the feature was less useful [t (47) = 3.03, P < 0.01] and their responses were not significantly different from the neutral rating of 4, indicating "No Opinion" [t (24) = -1.19, ns.]. Thus, the feature was only found to be useful by groups who experienced the *equal* and *unequal sharing* schemes.

A similar pattern was observed when questioned whether they found the feature annoying. Those who experienced both *low* and *equal* sharing schemes were not significantly difference from the neutral rating, [t(24) = 1.571, ns.], whereas those who experienced both *unequal* and *equal bandwidth* schemes disagreed with the statement [t(23) = 2.71, p < 0.05].

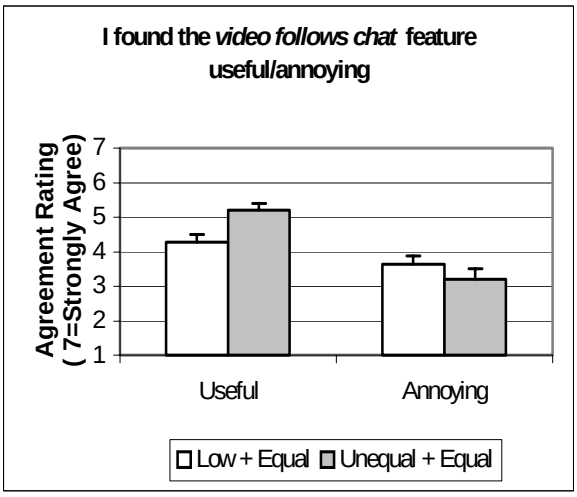

Figure 7. Video follows chat was useful and not annoying with higher bandwidth.

A broader range of opinion was gleaned from qualitative comments. As shown in Table 4, three themes emerged. First, as we expected, it was difficult to pay attention to both the video window and the chat window. Second, many people wanted to appear themselves when they sent a chat message. Finally, a number of people commented that the switching of the video was too fast.

Qualitative Comments
"Hard to watch video and chat - did one or the other"
"I can't follow both so tend to ignore video most of the time."
"I don't see myself in the large screen when I chat, is it supposed to be this way?"
"I'd like to see my face while sending. I don't like it that others can see me while I can't"
"Would be useful with fewer people, otherwise it's too fast to catch the video while someone is chatting."
"Sometimes video follows chat changes too quick"
"Too many people and thus chaotic"
"Bewildering. Could be good if fewer people and all familiar faces"
"Good idea, but hard to keep up, and when reading text there is movement in the peripheral vision. There's a lot going on the screen so quality isn't crucial as can't take it all in."

Table 4. Qualitative comments on the *video follows chat* feature.

5. DISCUSSION

5.1 (1) Minimum Bandwidth

Results on the three measures of video quality clearly indicate that participants noticed the difference in video quality when it was set to the *low bandwidth* scheme, and more importantly, that quality was seen to be unacceptable at this level. Only 30% of users found the *low bandwidth* scheme acceptable all of the time. These findings indicate that video chat with high quality video, offers users more than a simpler system with a single video frame accompanying each chat message. We argue that the important factor is the recognition of complex emotions. However, the video requirements for this appear to be lower than for audio/video conferencing. Here the minimum has been set at 5 fps, yet we find high levels acceptability for frame rates of just 2.5 fps with the *equal bandwidth-sharing* scheme. From this we conclude that a frame rate of 2.5fps, or 16 users in 400kbps session, is perfectly acceptable for effective video chat.

5.2 (2) Maximizing Video Quality

We have some evidence that *unequal bandwidth sharing* can improve perception of video quality such that users perceived it as less slow and jerky. However, on the other measures (rating scales and acceptability) the discrimination in quality is less clear. One reason for this may be that the scheme we implemented gave less bandwidth to the *focus user* than we expected.

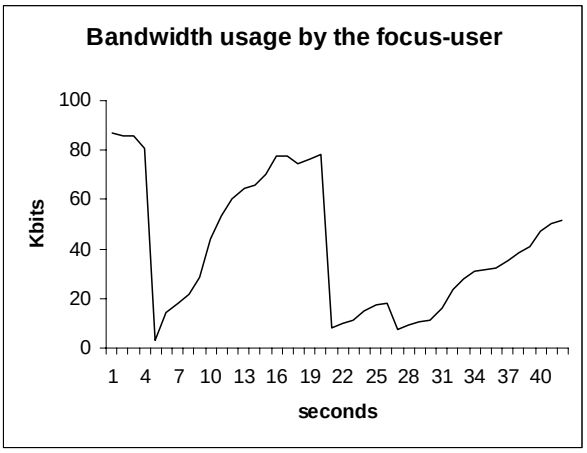

Figure 8: A graph of bandwidth consumption by the *focus user* when using *unequal bandwidth sharing*

This was because focus users increased their bandwidth consumption at a relatively slow rate in comparison to the pace of the chat. An illustration of this is shown Figure 8, which shows bandwidth consumption by the focus user over a 40 second period using *unequal bandwidth sharing*. Within this period the *focus user* changed 3 times, after approximately 5 seconds, 21 seconds and 27 seconds. Immediately after each switch, the bandwidth drops, because the new *focus user* is initially sending at a low send rate. However, as the scheme calculates the bandwidth available the focus user steadily increases its share.

The net result is that the *focus user* secured more bandwidth than with *equal sharing* but also experienced wide variations in bandwidth consumption. Figure 9 shows bandwidth consumption by the *focus user* during the first two rounds of Group D. During the first round, *unequal sharing* was active. During the second round *equal sharing* was in operation. With the *unequal sharing* scheme activated, the *focus user* secured a larger portion of the available bandwidth than with *equal sharing*, as is shown by the dashed line. However, the scheme also created wide variations in bandwidth consumption, including some transient periods where bandwidth consumption for the *focus user* fell bellow the average consumption for *equal sharing*.

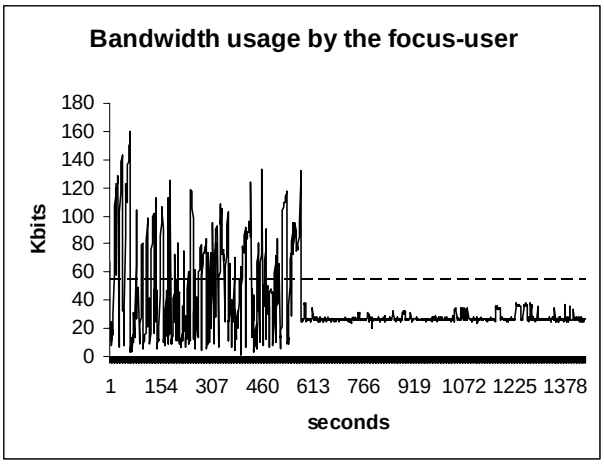

Figure 9: Bandwidth usage by the focus user with unequal and equal sharing schemes

Across groups, we found that our implementation of *unequal sharing* actually delivered *less bandwidth* than the *equal sharing* scheme 44.8% of the time. Yet, participants still rated the video as less slow and jerky. One explanation is that users focused more on the video when the frame rate increased, as there was more movement to attract attention. Thus the impression was of better quality video even though half the time the quality was actually worse.

With a few improvements to our scheme, it should be possible to secure a much larger portion of the session bandwidth for important users and also improve the reaction time of the algorithm. However, many real world factors may produce performance similar to that observed in our evaluation. The most obvious examples are large round-trip-times or packet losses on the network. These could interfere with the messaging process controlling bandwidth sharing and under these conditions we would expect the bandwidth adaptations to be delayed.

Thus, the evaluation suggests that *unequal bandwidth sharing* can still offer some gain, (or at the very least – no costs) even under such sub-optimal conditions.

5.3 Video Follows Chat

Generally, people found the *video follows chat* to be useful and on balance did not find the automatic switching annoying. However, qualitative comments from many users provided more detailed feedback on the design decisions we had made.

5.3.1 (3) Before or After?

Although no one commented directly on this decision there were numerous comments on the pace of the chat. As is common with large group chat [22] we observed multiple conversational threads overlapping and a very rapid turnover of messages. We would predict that had we implemented *video follows typing* instead there would be a slower pace of chat, greater coordination of process and fewer overlapping threads. However, the validity of these predictions and the consequences on user experience need to be evaluated in a further study.

5.3.2 (4) See yourself?

A few participants volunteered that they would prefer to see themselves when they sent a chat message, even though we did not directly ask them to comment on this feature. This self-referential feedback would increase awareness of what was being communicated through the video channel and might facilitate understanding of the channel. This parallels observations made on the early stages of video conferencing use [6]. In the future it would be interesting to evaluate this feature for both *video follows chat* and *video follows audio* in order to see if the "conventional wisdom" used in current designs holds true.

5.4 Other Issues

A number of people commented that it was hard to watch both the video and chat windows at the same time. Our interpretation of these comments was that following and understanding the chat demanded most of the users' attention. In future prototypes we intend to experiment with overlaying chat messages in the focus window, as shown in Figure 10.

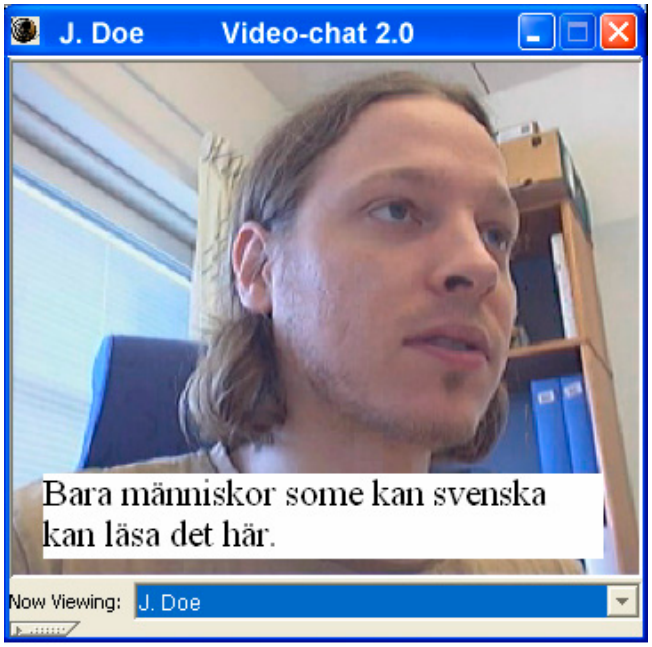

Figure 10. An example of a chat message overlayed in the focus window.

This type of design has been suggested recently by [3,17] and uses a display technique familiar from graphic novels where text and graphical information are presented in the same window. In implementing this design we would also need to ensure that people have enough time to read the message overlaid on the video display.

6. CONCLUSIONS

6.1 Limitations

This paper describes a formative evaluation that was conducted as part of an ongoing iterative design process, and the results should not be taken outside this context. We have only tested *video chat* in a single application setting that models a predicted use case in a wide variety of scenarios. The design requirements in other settings may not be the same as those reported here.

For example, the version of Marratech used in our experiments utilizes H.261, and thus delivers relatively low quality video compared to systems based on more modern codecs. It is possible

that this affected the experimental results and may be one reason why some users found "no difference" between variations in video quality. This may not be the case with systems that deliver higher quality video since in this case the video may attract a larger amount of the users' attention making them more sensitive to changes in video quality.

6.2 Substantive Conclusions

The vast majority of people we tested said they found both the video channel and *video follows chat* feature useful. But there are clear indicators in the qualitative comments that video chat could benefit from a number of design changes.

One issue was the lack of feedback to the person sending the message. The application might be improved if the sender also sees the video switching to them when they send a message. Another issue was the visual separation of video and chat, and the problems experienced when trying to follow both. A merging of these two streams, in a comic strip type layout might improve legibility. Whether the bandwidth requirements would be the same with this modified layout would require further study.

In its current form, in a social setting, the *low bandwidth* scheme of 0.2 fps delivers an unacceptable level of quality. Our explanation of this is that the communication of basic emotional expressions is blocked at this low rate. With *equal bandwidth sharing* and a frame rate of 2.5 fps the quality is acceptable all the time to the large majority (81%) of users. However, approximately half of users (51%) did find video quality to be acceptable even at 0.2 fps. This suggests an extreme insensitivity to bandwidth variations by a fair number of users during *video chat*. There is evidence that perceived quality can be further improved by using *unequal bandwidth sharing*. When supporting larger groups, (i.e. E-learning), such a scheme may be essential to maintain acceptable levels of video quality.

6.3 Methodological Conclusions

Any application requires a number of design decisions to be made. With single user applications it may be possible to evaluate alternatives with users prior to implementation. This is not feasible with applications designed for large groups. Our approach under these circumstances was to first make a set of design decisions, identifying our rationale for each one. In the formative evaluation we sought to map both the qualitative and quantitative measures onto the design decisions we had made. Although the mapping of questions to answers is loose we found sufficient information from the evaluation to identify problems with the current prototype that inform the iterative design process.

Of the four different measures we used to evaluate video quality, only the user-descriptor scales recommended by [29] brought out the differences in quality perception between *equal* and *unequal bandwidth sharing*. This would indicate that these scales are more sensitive to small variations in quality as they aligned with how people naturally describe video quality.

7. FUTURE WORK

Future work will primarily focus on further design iterations of the video chat application. The data collected during the evaluation point to a number of areas for improvement. These include implementing a version of the *unequal bandwidth sharing* scheme that reacts quicker to changes in user importance, overlaying chat messages into the video window (see Figure 10), and also evaluating user perceptions of *video follows chat* vs. *video follows typing*. In addition, future evaluations will incorporate eye tracking of users, to get a clearer picture of where users locate their visual attention while participating in video chat. Finally, we intend to go forward with ecological evaluations of future prototypes by deploying them in e-learning and video-corridor settings.

8. ACKNOWLEDGMENTS

This work was done within the VITAL Community project, which is supported by the Objective 1 Norra Norrland - EU structural fund programme for Norra Norrland. Support was also provided by the Centre for Distance-spanning Technology (CDT). Thank you is also given to Stefan Elf for his help developing the prototype system for uneven bandwidth sharing.

9. REFERENCES

1. Alkit Communications, http://www.alkit.se
2. Amir, E., McCanne, S., and Katz, R. Receiver-driven Bandwidth Adaptation for Light-weight Sessions. In *Proc. ACM Multimedia 1997*, ACM Press (1997), 415 – 426
3. Ayatsuka, Y., Matsushita, N., Rekimoto, J., ChatScape: a visual informal communication tool in communities. *Ext. Abstracts CHI 2001*, ACM Press (2001), 327 - 328
4. Chen, M. Achieving effective floor control with a low-bandwidth gesture-sensitive videoconferencing system. In *Proc. ACM Multimedia 2002*, ACM Press (2002), 476-483.
5. Daly-Jones, O., Monk, A. & Watts, L. (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 (1998), 21-58.
6. Dourish, P. A. Adler, V. Bellotti, and A. Henderson. Your Place or Mine? Learning from Long-Term Use of Audio-Video Communication. *Computer-Supported Cooperative Work*. (1996) 5(1): pp. 33-62.
7. Dourish, P., and Bly, S. Portholes: Supporting Awareness in a Distributed Work Group. In *Proc. CHI 1992*, ACM Press (1992), 541-547
8. Ehrlich, S.M., Schiano, D.J., Sheridan, K., Communicating Facial Affect: It's Not the Realism, It's the Motion . *Ext. Abstracts. CHI 2000*, ACM Press (2000), 251 - 252
9. El Kaliouby, R., Robinson, P., and Keates, S. Temporal Context and the Recognition of Emotion from Facial Expression. In *Proc. HCII 2003*, (2003).
10. Furnas, G.W. Effective View Navigation. In *Proc. CHI 1997*, ACM Press (1997), 367 - 374
11. Handel, M., and Herbsleb, J.D. What Is Chat Doing in the Workplace? In *Proc. CSCW 2002*, ACM Press (2002), 1 – 10
12. Hellström, G., Quality Measurement on Video Communication for Sign Language. In *Proc. HFT 1997*, 217 - 224
13. Herbsleb, J.D., Atkins, D.L., Boyer, D.G., Handel, M., and Finholt, T.A. Introducing Instant Messaging and Chat in the Workplace. In *Proc. CHI 2002*, ACM Press (2002), 171 – 178

14. Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D.J., Kamm, C., I M everywhere: The character, functions, and styles of instant messaging in the workplace. In *Proc. CSCW 2002*, ACM Press (2002)
15. Jackson, M., Anderson, A.H., McEwan, R., and Mullin, J. Impact of Video Frame Rate on Communicative Behaviour in Two and Four Party Groups. In *Proc. CSCW 2000*, ACM Press (2000), 11 – 20
16. Jacobson, V., DARTNET Planning and Review Meeting, 1991
17. Kurlander, D., Skelly, T., Salesin, D., Comic Chat In *Proc. SIGGRAPH 1996*, ACM Press (1996)
18. Marratech AB, http://www.marratech.com
19. McCarthy, J.D., Sasse, M.A., Miras, D., Sharp or smooth?: comparing the effects of quantization vs. frame rate for streamed video. In *Proc. CHI 2004*, ACM Press (2004)
20. Masoodian, M., Apperley, M.D., and Frederickson, L. Video Support for Shared Work-Space Interaction: An Empirical Study. Interacting with computers 7, 3 Elsevier Science Ltd. (1995), 237-253
21. Nardi, B.A., Whittaker, S., Bradner, E., Interaction and outeraction: instant messaging in action. In *Proc. CSCW 2000*, ACM Press (2000)
22. O'Neill, J., and Martin, D. Text Chat in Action. In *Proc. GROUP 2003*, ACM Press (2003), 40 –49
23. Schiano, D.J. , Ehrlich, S.M., and Sheridan, K., Categorical imperative NOT: facial affect is perceived continuously, In *Proc. CHI 2004*, ACM Press (2004), 49 - 56
24. Scholl, J., Elf, S., and Parnes, P. User-interest Driven Video Adaptation for Collaborative Workspace Applications. In *Proc. NGC 2003*, Springer-Verlag (2003), 3-12
25. Short, J., Williams, E., & Christie, B. The Social Psychology of Telecommunications. (1976) New York: Wiley
26. Tang, J.C., and Isaacs, E.A. Why do users like video? studies of multimedia-supported collaboration. In Computer-Supported Cooperative Work: An International Journal 1, 3 (1993), 163–196
27. Tscholl M., McCarthy J. D., Scholl J. The effect of video-augmented chat on collaborative learning with cases, In *Proc. CSCL 2005*
28. ViDe Video Conferencing Cookbook 4.0. http://www.videnet.gatech.edu/cookbook/
29. Watson, A., Assessing the Quality of Audio and Video Components in Desktop Multimedia Conferencing. (2001) PhD Thesis, University of London. Available at http://www-mice.cs.ucl.ac.uk/multimedia/projects/etna/watson.pdf
30. Watson, A., and Sasse, M.A. Evaluating audio and video quality in low-cost multimedia conferencing systems. Interacting with Computers 8, 3 (1996), 255–275
31. Watson, A., and Sasse, A., Measuring perceived quality of speech and video in multimedia conferencing applications. In *Proc. ACM Multimedia 1998*, ACM Press (1998), 55 - 60
32. Williams, E., Experimental comparisons of face-to-face and mediated communication: A Review. Psychological Bulletin, .84, 5 (1997), 963-976.