The effect of video-augmented chat on collaborative learning with cases

Michael Tscholl
Department of Computer Science
University College London
m.tscholl@cs.ucl.ac.uk

Jeremiah Scholl
Department of Computer Science
Lulea University of Technology
jeremiah@cdt.luth.se

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

Abstract. We investigated the effect of augmenting a chat-based learning system, with a low- and high-quality video link, on the focus (factual vs. abstract) of a case-based learning discussion. Efficient learning with cases requires discussion on the case as well as on the meaning of abstract knowledge. We find that students who first experience chat + high-quality video, focus significantly more on abstract knowledge than students first exposed to chat-only and chat + low-quality video. We also found that these students expressed a preference for face-to-face discussion. We conclude that video should be employed to improve case-based CSCL or in other methods where a balance between discussions on abstract and concrete knowledge are important.

Keywords: case-based CSCL, video, eye tracking, dialogue analysis, non-verbal communication

INTRODUCTION

Anthropological and social research has consistently revealed the importance of non-verbal cues for dialogue, discourse and information management. For example, focus shifts are introduced and accompanied by specific facial behaviors and gestures (Kendon, 1987; McNeill, 1992); establishing and maintaining a common ground uses non-verbal cues (Chovil, 1991), that become especially important if the object of discussion is complex (Argyle & Cook, 1976). While the importance of visual cues to communication has been studied extensively, there is little research on how visual cues affect learning quality and efficay.

Here we investigate this relationship by analyzing a case-based CSCL dialogue, distinguishing between foci on factual information and foci on abstract, general knowledge. Good learning with cases requires that generalizations are drawn from the factual information conveyed in the cases and that the principle or point of the case is understood and reflected upon (Guzdial, et al. 1996). Generating an abstract schematic representation of a case or relating factual information to abstract knowledge aids the insertion of this case into a “knowledge base” of cases that are reusable for solving further problems (Schank, 1987, Kolodner, 1993). While this research has been carried out in a case-based CSCL environment, its findings are relevant for all educational methods using discussion where a careful balance between factual and general knowledge must be fostered.

Earlier research has shown that the technology used for learning, affects what is talked about. For example, Veerman et al. (1999) found that in chat-only systems, students focus much more on the use of knowledge rather than its meaning, while the opposite result is found in a system featuring chat and a facility for the representation of conceptual knowledge. The student writing groups of Hewett (1998) using computers and oral communication focused on more abstract, global idea development, than their peers using only computers. And in observational studies of learning with cases, from which this study derived its hypotheses, Tscholl & Dowell (2005) found that students were much more likely to engage in a discussion on the general concepts and principles of the case when discussion was conducted face to face, than if conducted with a CSCL system featuring an unthreaded chat and a shared whiteboard.

The limitations of text-based communication has motivated the development of systems that augment chat, generally by supplying specific software tools and promoting, through additional tasks, their use. For example,
tools have been designed to facilitate and promote argumentation (Bell, 1997), to allow indexing of cases to aid organization, structuring and, as a result, memorization of problem solving episodes (Kolodner, 1997), or external diagrammatic representations have been used to foster the acquisition of rich interconnected knowledge (Suthers, 2003). While these applications have been successful in improving learning, they are generally tied to a specific system and also difficult to transfer to a different area of knowledge (e.g. science, arts, law, medicine). Augmenting chat with channels typical of F2F discussion may be a more straightforward way to counter the downsides of chat alone. An obvious choice is video. Video and text communication is now widely used for social communication with popular Instant Messaging (IM) applications. The possible impact of video on chat based CSCL is currently not well known. To understand the possible impact of video on learning discussions we first present a brief overview of how the addition of video might affect dialogue.

BACKGROUND

Communication over different media channels is generally understood using Social Presence Theory. Classical social presence theory (Short, Williams & Christie, 1976) was developed to describe synchronous communications involving face-to-face, audio, or close-circuit video telecommunication media. The concept of social presence was related to the social psychological constructs of Intimacy and Immediacy. Argyle and Dean (1965), saw Intimacy as “a function of eye-contact, proximity and conversation topic; changes in one will produce compensating changes in the others; eye-contact is generally sought after, but too much creates discomfort; for instance, eye-contact is reduced when people are placed very close together”. Immediacy was described as a measure of the psychological distance that a communicator puts between themselves and the object of their communication. In this context, Weiner and Mehrabian, 1968 showed that the choice of language (e.g. ‘We…’ as opposed to ‘I…’ or ‘You…) connotes different feelings of closeness and association.

As originally formulated, social presence theory did not deal with text-based communication yet within the construct is remains possible that Intimacy and Immediacy can still be maintained through language use. The importance of video, or visual cues, remains unclear. At one extreme, Walther (1999) concludes that visual cues have little place in CMC and that text-based interaction can foster exceedingly intimate and affectionate communication, to levels greater than those common in parallel offline ftf interaction. At the other extreme Rice and Love suggest that without visual cues communication “will be perceived as impersonal and lacking in normative reinforcement, so there will be less socioemotional (SE) content exchanged” (Rice & Love, 1987). While the emotional value of visual contact remains equivocal, even more elusive is an understanding of the importance of visual cues to signal topic shifts during a text-based discussion. If eye contact is important in face-to-face situations then its value may be muted over a video channel, where direct eye contact is impossible. On the other hand, it remains possible that the video channel could signal attitude, mood and agreement (e.g. nods of the head), and topic shifts much more effectively than text-based communication alone. To investigate this we examine the impact of adding a video channel to chat room discussion using the case-based-learning method.

RESEARCH QUESTION AND APPROACH

The primary research question was whether the addition of a video channel leads to more discussion of abstract concepts. More generally we were interested in whether the dialogue patterns differed across the media conditions.

To isolate different aspects of the video channel we studied two different video conditions, alongside chat only communication. In the low frame rate condition video was delivered at 1 frame every five seconds. This frame rate was adopted to give a sense of presence of the other students while effectively blocking non-verbal communication. As outlined recently by (Ehrlich et al, 2000), motion is an important pre-requisite for emotion communication, but at 1 frame very 5 seconds no motion is communicated and simple non-verbal gestures like nods and shakes of the head cannot be discriminated. In the high frame rate condition video was delivered at 25 frames/second. This is equivalent to motion quality delivered in television and film and is capable of delivering complex emotion and non-verbal communication.

To measure how students use of the video channel we employed eye tracking. The eye tracker records the location of gaze on the screen 50 times/second. Finally, we also measured user perceptions of the different media conditions with a short questionnaire.
Dependent Variables

Dialogue Analysis

Facts vs. General (F-G Coding)
The focus analysis, presented in detail in the next section, is the basis for the identification of abstract knowledge or notions as the current focus of discussion. We distinguished two types of foci: one type, the concrete one, includes facts or detailed aspects of the case, and the other, the abstract type, includes general concepts pertinent to the case, the principle or the ‘point’ of the case. What we wanted to capture with this distinction is whether an utterance is explicitly tied to the factual information conveyed in the case or whether it is only derived from it and positioned into the dialogue as a stand-alone object of discussion. We reasoned that bringing into the discussion such objects would require non-verbal communication, as such a shift would entail a change from a less to a more complex topic and subsequently the maintenance of the more complex topic.

All utterances or propositions within an utterance containing at least one reference to a specific detail of the case or the case as a whole (“but she refused to attend twice”) were counted as concrete/actual (F). Utterances focusing on the definition or ‘nature’ of concepts (“what is negligence actually”), utterances referring to the general consequence of applying a concept to the case (“…but then a doctor has to check every symptom every time. This is not practical”), and utterances mentioning the point of the case (“this is a he-said, she-said situation”) were designated as abstract/general (G). Utterances outside these criterions (such as “yes, I think so, too”) where categorized as the utterance they referred to (if the reference could be determined uniquely).

Coordination measure
Techniques from Pilkington’s discourse analysis scheme for educational dialogues (Pilkington, 1999) have been employed to derive a measure for coordination. The analysis scheme allows coding of single utterances into topics and foci from which, through the designation of parent- and child-topics, the topic structure of a sequence of chat messages can be revealed. In a discussion, many utterances are observably topically related to previous ones; for example, they may add information to a previous assertion, or question, confirm or critique it. However, multiple threading, a persistent phenomenon in chats, disrupt the sequence of messages. To measure this, messages topically related to the immediately previous are designated with a score of 0. Similarly, messages that started a thread, i.e. that introduced a new topic were also scored as 0. Those related to the one before the previous message are assigned a score of 1, etc. For example, if message 16 were topically related to message 12, it would be assigned a score of 3.

When it wasn’t possible to determine the reference of a message, i.e. some relation to previous messages was not observably topical, the message was classified as ambiguous and did not contribute to the coordination measure. When summed over the entire discussion this gives a measure of the coordination of the dialogue. If the total is zero then all either start a new thread or refer to the previous one. Figure 1 illustrates coding of the G-F measure and the coordination measure: R is the number of messages between the current and the referred one and L the F-G coding, T the time of sending message, and S the sender.

Naming
Chat users are clearly aware of difficulties of coordination, and compensate it partially by direct referencing a message through naming its sender. Such instances of peer naming were also counted.

Eye-Tracking Measures
The eye tracker records gaze position on the screen 50 times/second as a series of X-Y data points. To understand where people look during the discussion we...
categorized the screen into 4 Regions of Interest. These were, (i) the participants’ thumbnails, (ii) video focus window, (iii) chat window and (iv) the browser containing the learning material. An illustration of these regions and the layout used in the evaluation is shown in Figure 2.

The measure of Gaze % is the proportion of (raw X-Y) gaze samples that are recorded in different areas of the screen. As an indicator of transitions we also recorded the Revisits to the different regions. A revisit is counted when gaze moves briefly out of one region for a single fixation before returning back to that region. It has been associated with the need to seek additional information.

**Questionnaire Measures**

The short questionnaire was structured as a series of statements. After discussing each Case participants rated their strength of agreement with the statements on a 7-point scale, where 7=strongly agree, 4=no opinion and 1=strongly disagree. Table 1 summarizes these statements.

<table>
<thead>
<tr>
<th>Questionnaire Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found it easy to understand the arguments made by other students</td>
</tr>
<tr>
<td>There were few misunderstandings between people</td>
</tr>
<tr>
<td>I found it easy to speak my mind and get my point of view across</td>
</tr>
<tr>
<td>I enjoyed discussing the case with other students</td>
</tr>
<tr>
<td>I would have preferred a face to face discussion</td>
</tr>
<tr>
<td>I thought video quality in the large window was very good</td>
</tr>
<tr>
<td>I thought video quality in the window showing all the participants was very good.</td>
</tr>
<tr>
<td>I feel exhausted and drained after the discussion</td>
</tr>
</tbody>
</table>

Table 1: Questionnaire statements.

**E-LEARNING STUDY**

Groups of 4 students discussed three cases under three different media conditions: chat only, low frame rate (0.2fps) and high frame rate (25fps). The three cases were actual cases of medical negligence, including a short description of the case details and a report on whether the doctor was found to be negligent or not. An example of one of the cases is given below (figure 3).

The **Vadera** case: A 22-year old woman presented herself 3 times within a year at her GP’s practice, with the intention of starting contraception before she got married. She was warned that there were health risks associated with contraception pills. On the last visit her blood pressure was taken and it was at 150/100 (higher than normal for a woman of her age). This high reading was taken by the GP (Dr. Shaw) as a symptom of ‘white-coat hypertension’, a phenomenon caused by anxiety that occurs in a doctor’s presence, that can however also be indicative of a general tendency to hypertension. The next day, she started the pill. A week later, the plaintiff was admitted to the hospital suffering from numbness and difficulty in walking. Her BP was read several times and was at 170/110, 110/60 and 140/110. She had suffered a stroke that left her completely paralysed.

The statistical evidence does not link taking contraceptives with stroke, over the population as a whole.

**Verdict:** the judge did find the GP negligent but not liable.

Figure 3: a case study material

The students were asked to discuss the case, express their agreement/disagreement with the reported verdict and explain their reasons for them. After discussing each case the students completed the short eight-item questionnaire. At the end of the evaluation we gave people another short questionnaire on whether they found the visual contact useful and whether they had used it to communicate.

**Method**

**Participants**

24 people participated in the study. 16 were female and 8 male. The mean age was 26. They were recruited from subject pools within XXXXX University and were paid $15 for participation. They were tested in groups of 4
people. All groups conducted discussion of cases in all three conditions, *Chat Only*, *Low Frame Rate* and *High Frame Rate*. One member of each group was eye-tracked to understand their attention patterns.

**Equipment and Software**

For our experiments we used a modified version of Marratech Pro, a commercially available multimedia conferencing tool that includes media such as audio, video, chat and a shared whiteboard. The Marratech Pro client is 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.

The commercial version of the Marratech uses a classical bandwidth-sharing scheme where the bandwidth allocated to each media is set to a static limit, with all clients given an equal portion of this bandwidth. We modified the client so that we could remotely set the frame-rate used by each client during our experiments.

**Procedure**

Before the start of the study all participants completed a short 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 that they would use to discuss medical negligence cases with other students. At the end of each case discussion they were to complete a short questionnaire provided. The same questions were asked after each case and the questionnaire layout encouraged active comparison with previous responses.

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.

**Design**

Table 2 summarizes the design for study. Two groups of 4 students were assigned to one of three different variations that counterbalanced the order of media conditions with a Latin Squares design.

<table>
<thead>
<tr>
<th>Order</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (2x4)</td>
<td>Chat</td>
<td>Low (0.2fps)</td>
<td>High (25fps)</td>
</tr>
<tr>
<td>B (2x4)</td>
<td>High</td>
<td>Chat</td>
<td>Low</td>
</tr>
<tr>
<td>C (2x4)</td>
<td>Low</td>
<td>High</td>
<td>Chat</td>
</tr>
</tbody>
</table>

Table 2: Study design.

**RESULTS**

**Dialogue Analysis**

Figure 4 shows the number of general concepts recorded in the dialogue. As illustrated in the figure there were significantly more general concepts produced by the two groups whose first experience was high frame rate video ($Z=3.068$, $p<0.001$).

![Number of General Concepts in the Dialogue](image)

However, we found no difference in chat coordination between the different media conditions and no difference in the frequency of naming.
Questionnaire Data

Analysis of the questionnaire data also revealed effects contingent on the order in which media conditions were experienced. For example, students said they found it harder to speak their mind if their first experience had been with chat only \( [F(4,42) = 2.7, p < 0.05] \). This difference disappeared when they used high frame rate video.

Preference for a face-to-face discussion also interacted with the order in which media conditions were experienced. As shown in Figure 5, those who were first exposed to high frame rate (25fps), expressed a clear preference for face-to-face discussion \([ F(2,21) = 5.35, P < 0.05]\). By contrast, those exposed to the low frame rate (0.2fps) on the whole said they would not prefer a face-to-face discussion. Those exposed to chat first did not have an opinion either way.

![Figure 5: Preferred a face-to-face discussion.](Image)

Despite these differences, the perception of video quality itself was very similar across the low and high frame rate conditions. Although people thought the quality was slightly better in the high frame rate condition, ratings were only marginally lower at 0.2fps (Mean = 4.5) than at 25fps (Mean=5.1). \([F(1,21)=5.7, p < 0.05]\).

Eye-Tracking Data

The eye tracking data illustrate that the video actually received very little of users’ attention. A visualization of the gaze distribution is presented in Figure 6. Gaze density is clearly much higher in the chat window in the bottom left of the screen.

![Figure 6: Gaze distribution to the application.](Image)

A summary of gaze to the different screen regions is presented in Figure 7. This graph shows the proportion of time the users gazes at different screen objects in the three different media conditions.
As illustrated, the bulk of the users attention is focused on the chat window. While not statistically significant there appears to be a decreasing tendency to attend to the chat window with higher frame rate video and a corresponding increasing tendency to attend off screen. Surprisingly, in the video conditions less that 10% of gaze is directed towards the video focus window and even less towards the participants’ thumbnail window. While this analysis suggests that gaze distribution is almost identical in different media conditions a very different story is given from an analysis of revisits. As described previously, a revisit is measured when the eye makes a single fixation outside a region before returning to that region. As shown in Figure 8 the revisits to both the focus and chat windows are a function of media conditions and much higher in the high frame rate condition \([F(2,4) = 14.5, p < 0.05; F(2,4) = 7.56, p < 0.05]\). This indicates in the high frame rate condition there are many instances where a glance is made from the chat to the focus window and vice versa.

One obvious explanation of this result is that the eye is attracted automatically to the motion present in the high video condition. If this were the case, we would expect the eye to be drawn to the video immediately after it switches to the latest message sender. However, only 3% of all video switches were immediately accompanied by a glance to the video focus window. Much more common was that the student would first read the message and then glance at the focus window.
To investigate further, we conducted a post-hoc investigation of eye movements in the “Start high” condition. Specifically we were interested in whether the eye movements were related in any way to shifts in the dialogue from discussing facts to general concepts of the case. When we replayed the recordings of the discussion with the eye movement overlaid we observed many instances where the eye would read a message in the chat window and then flick briefly to the video focus window as though seeking more information from the person who sent it. Similarly when the person being eye tracked sent a message, they would often flick to the focus window as if to check for a reaction to the message they sent. We were interested in whether the frequency of these actions differing during periods when the students discussed general concepts. If these eye movements were important to maintain or signal the topic shift then we would expect more glances to the focus window when general concepts were being discussed. To investigate this we calculated two different measures. One measure records the proportion of messages of a particular type that are followed by glances to the video window. The other is a measure of the mean number of glances made in these instances. As shown in Table 3, there is a higher proportion of glances to the video window in the high frame rate condition. However, there is clearly no difference in the proportion of glances following factual (F) statements vs. general (G) concepts. In fact, although not significant, the mean number glances following factual statements is marginally higher than those following general concepts. This is in the opposite direction to what we would have predicted.

<table>
<thead>
<tr>
<th></th>
<th>High Frame Rate</th>
<th>Low Frame Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td>Prop. of messages followed by video glances</td>
<td>71%</td>
<td>73%</td>
</tr>
<tr>
<td>Mean No. Of Glances</td>
<td>2.3</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 3: Glances during different phases of discussion.

As an additional analysis we calculated equivalent measures to investigate how they used the video window when they either sent a message themselves or read a message from someone else. One measure was the proportion of times they glanced at the video after they sent a message (to check response) vs. the proportion of times they glanced at it when someone else sent a message (to check intention). For the messages they did glance at the video after we calculated the mean number of glances/message. Again however, we found fewer glances to the video window in the low frame rate condition but no differences in glances following messages sent vs. messages read (See Table 4). Informally, we also observed that expressed emotion during the chat sessions was almost completely absent.

<table>
<thead>
<tr>
<th></th>
<th>High Frame Rate</th>
<th>Low Frame Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sent</td>
<td>Read</td>
</tr>
<tr>
<td>Prop. of messages followed by video glances</td>
<td>69%</td>
<td>70%</td>
</tr>
<tr>
<td>Mean No. Of Glances</td>
<td>2.3</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Table 4: Glances to messages sent and read.

**DISCUSSION**

The results show that adding high-quality video to a chat-based distributed learning environment has an effect on both the discussion and attitudes of the participants. However this was only observed when the students first session used high quality video. For these groups the students consistently produced more general concepts in the discussion, even in later sessions where the video channel was no longer present. To examine use of the video channel we employed eye tracking. Compared to the chat window, there were very few glances to the video focus window, however there was some evidence that people would switch between the chat and video window, especially in the high frame rate condition. To examine whether these glances might explain the differences in dialogue we conducted a post-hoc analysis relating eye movements to the statements made. But we found no evidence that the video channel was used in any way to regulate the dialogue. Even if such evidence were uncovered it would not explain the production of general concepts for the “start high” groups when they communicated by chat alone. Thus, although the video has clearly had an effect, we find no evidence that this effect is due to a regulation of dialogue through the visual channel.
A much simpler explanation is that the first experience with high frame rate video changes the way students feel about each other. In terms of social presence theory, the video may increase the intimacy between students in the group. To some extent, this intimacy may be unfulfilled. Although students seek eye contact, because of the position of the camera, it is never actually made. This leaves them feeling as though they would do better to meet face to face. Despite this, these students’ feel more connected to others in the group and are much more willing to shift discussion from the factual to the general. In other words, the first experience with high quality video may function as an ice-breaker that increases group cohesion. The educational importance of such ice-breakers is well known in standard educational settings (Lott & Lott, 1965; Meyers, 1997) and here we have evidence that simply adding high quality video to a chat discussion significantly changes its educational effectiveness.

CONCLUSIONS

Methodological
There are two innovations in user measurement that are fruitful for future research. Firstly, the technique to code concrete and abstract references in the dialogue revealed interesting differences that identify subtle changes in communication patterns under different media conditions. Secondly the use of eye tracking helps to isolate in great detail how people use the information available on the screen.

To evaluate the educational effectiveness of CSCL dialogues we would recommend a coding that distinguishes between factual and general concepts. While this is particularly suited to the Case Based Learning method we use here it can also be used in other approaches such as problem-based-learning (PBL, cf. Barrows, 2002)

On the eye-tracking front we would caution against using a simple measure of gaze % to different screen regions. As shown by the re-visit and post-hoc analysis the gaze % measure can hide subtle but important differences in how people use information available on the screen.

Overall we would recommend adopting a multi-dimensional approach to identify how dialogue, attitude and eye data can be tied together to give a more detailed picture of user performance.

Substantive
The findings suggest that the pattern of learning using chat systems may be quite different with and without high quality video. Students whose first experience is with chat or low quality video tend to focus on the facts of the case and do not share their thoughts on general concepts relating to the cases being discussed. In comparison, those whose first experience is with high quality video, discuss general and abstract concepts much more readily. A stronger conclusion is that what is learnt is different with and without high quality video. In analyzing how student use the video channel we found that it is not used directly to support shifts in the dialogue from factual to general. Instead we conclude that high quality video helps promote group cohesion and it is this connection gives students the confidence to switch the discussion to more abstract concepts.

Whether such cohesion might be obtained through long-term use of text-based channels is unclear. But our findings certainly suggest that in the short term, first experience with high quality video changes the educational value of the dialogue. Thus, for first meetings, or one-off CSCL sessions, the addition of a high quality video channel is strongly recommended.

ACKNOWLEDGMENTS
We thank CHI, PDC and CSCW volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document. Some of the references cited in this paper are included for illustrative purposes only.

REFERENCES
Barrows, H. (2002). Is it truly possible to have such a thing as dPBL? Distance Education, Vol. 23, No. 1, 2002.


