T. Koschmann, D. Suthers, & T.W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The Next 10 Years!*
Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

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

Abstract. Efficient learning with cases requires discussion on the facts of the case as well as on their meaning. We investigated the focus (factual vs. abstract) of a case-based learning discussion when video was added to a chat-based learning system. Students whose first experience includes high-quality video, focus significantly more on abstract knowledge than students first exposed to chat-only or chat + low-quality video. We also found that these students expressed a preference for face-to-face discussion. We conclude that video may improve learning where 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. However, the specific effects of these additional cues on learning are not well studied. Here we investigate the availability of visual contact on learning 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 point of the case is understood and reflected upon (Guzdial, et al. 1996). 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.

BACKGROUND

The effect of non-verbal cues on dialogues is well studied. For example, focus shifts are introduced and accompanied by specific facial behaviors and gestures (Kendon, 1987; McNeill, 1992); establishing and maintaining common ground uses non-verbal cues (Chovil, 1991; Argyle & Cook, 1976). Less is known about how visual cues affect learning quality and efficacy. Some pertinent research has, however, 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 more on the use of knowledge than its meaning, while the opposite result is found in a system featuring chat and a facility for representing 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 comparing f2f with CMC, from which this study derived its hypotheses, Tscholl & Dowell (2005) found a similar pattern. The alternative explanation of differences on learning with and without visual contact focuses on the improved intimacy and immediacy of face-to-face dialogues. A certain amount of intimacy is valuable in case-based learning, where drawing parallels and seeing similarities and differences between (personal) episodic knowledge and the case under discussion is an important means of generalizing. Further, as collaborative learning is promoted by exchanging viewpoints, contrasting and critiquing, the closeness of visual contact may affect the dynamic of the dialogue (cf. Walther (1999))

Augmenting chat with a video channel may be a straightforward way to counter the downsides of chat alone. This research aims to contribute to the understanding of the impact of video on learning discussions.

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. We this by using 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. To measure how students used the video channel we employed eye tracking. We also measured user perceptions of the different media conditions with a short questionnaire.

Dependent Variables

Facts vs. General (F-G Coding)

We distinguished two types of foci: concrete foci, includes facts or detailed aspects of the case, and the other, abstract foci, 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 such objects into the discussion would require non-verbal communication, as such a shift would entail a change to, and maintenance of, a 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/factual (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

We measure coordination by assigning a score of 0, 1, 2, …to indicate the number of messages between the current and the topically related one. Messages starting as new thread were scored 0.

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 1. 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 with 8 questions was structured as a series of statements, gauging for usability, enjoyment and perceived video quality. Included was: "I would have preferred a face to face discussion". After discussing each case participants rated their strength of agreement with the statements on a 7-point scale.

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 judges' verdict (example: Figure 1). The task given was to explain the judge's verdict.

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 cause 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 1: An example case from the study

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 UCL 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. The version we used limited bandwidth usage for video to 400 kb/s.

Procedure

A short questionnaire given before the discussion probed basic demographic information and the participants' experience with chat rooms, instant messaging (IM) and video conferencing. At the end of each case discussion they completed a short questionnaire. 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

Design

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 1).

Order	Case 1	Case 2	Case 3
A (2x4)	Chat	Low (0.2fps)	High (25fps)
B (2x4)	High	Chat	Low
C (2x4)	Low	High	Chat

Table 1: Study design.

RESULTS

Dialogue Analysis

Figure 2 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 ($Z=3.068$, $p< 0.001$). However, we found no difference in chat coordination between the different media conditions.

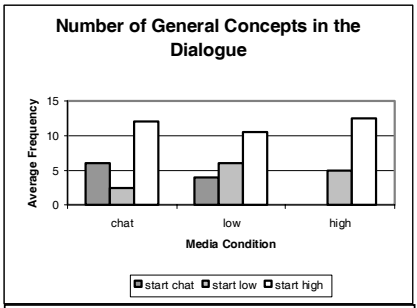

Figure 2: G's produced by condition

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. 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.

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 3 (the learning material display occupied the right and center of the screen; below left: chat window; mid left: large video; above left: thumbnails). Gaze density is clearly much higher in the chat window. Surprisingly, in the video conditions less that 10% of gaze is directed towards the large video window and even less towards the thumbnail window. Across media conditions there appeared to be large differences in the *revisits* made to different screen regions. As described above, a *revisit* is scored when the eye makes a single fixation outside a region before returning to that region. Revisits to both the focus and chat windows are 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.

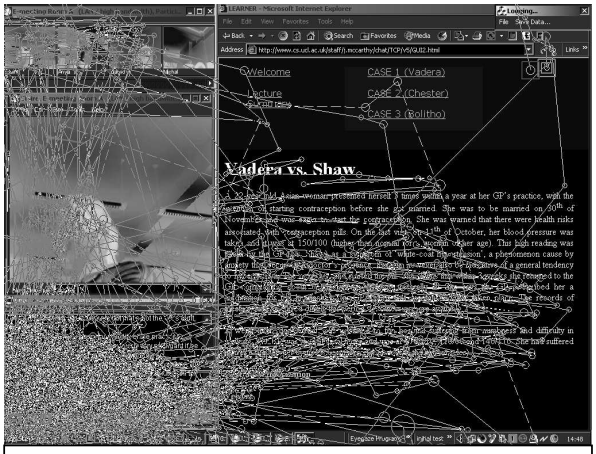

Figure 3: Gaze distribution.

One explanation of this result is that the eye is attracted to the motion 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. 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. Specifically, the video may be used for checking the reaction to a sent message. To investigate, we calculated two different measures. One records the proportion of messages of a particular type that are followed by glances to the video window. The other measures the mean number of glances made in these instances. As shown in Table 2, there are 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 contrast, though not significant, the mean number glances following factual statements is marginally higher than those following general concepts.

	High Frame Rate		Low Frame Rate	
	F	G	F	G
Prop. Of messages followed by video glances	71%	73%	52%	56%
Mean No. Of Glances	2.3	1.8	3.4	1.7

Table 2: 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 where they did glance at the

	High Frame Rate		Low Frame Rate	
	Sent	Read	Sent	Read
Prop. of messages followed by video glances	69%	70%	45%	63%
Mean No. Of Glances	2.3	2.3	2.8	2.6

Table 3: Glances to messages sent and read

video, 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 3). Expressed emotion during the chat sessions was almost completely absent.

CONCLUSIONS

The results show that adding high-quality video to a chat-based distributed learning environment has an effect on the discussion, attitudes and behavior of participants. However this was only observed when the students first used high quality video. Then, more general concepts in the discussion were produced, an effect that persisted through the remainder of the session. The eye tracking data show that compared to the chat window there were very few glances to the video focus window. However there was some evidence that people would frequently and briefly switch between the chat and video window, especially in the high frame rate condition. 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 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. 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. In other words, the first experience with high quality video may function as an *ice-breaker* - whose importance in education is well known (Lott & Lott, 1965; Meyers, 1997).

There are two innovations in this study 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 detail how people use the information available on the screen.

While a coding that distinguishes between factual and general concepts 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 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.

REFERENCES

Argyle, M. and M. Cook (1976). *Gaze and Mutual Gaze*. Cambridge, UK, Cambridge University Press.

Chovil, N. (1991). "Discourse-Oriented Facial Displays in Conversation." *Research on Language and Social Interaction* 25(1991/1992): 163-194.

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

Guzdial, M., Hmelo, C.E., Hubscher, R., Nagel, K., Newstetter, W., Puntambekar, S., Shabo, A., Turns, J., & Kolodner, J.L. (1997). Integrating and Guiding Collaboration. *Proceedings CSCL '97*, pp.91-99.

Hewett, B. L. (1998). The characteristics and effects of oral and computer-mediated peer group talk on the argumentative writing process. *Dissertation Abstracts International*, 59(09), 3435.

Kendon, A. (1987). On gesture: its complementary relationship with speech." *Nonverbal Behavior and Communication*. A. W. Siegman and S. Feldstein. Hillsdale, Lawrence Erlbaum Associates, Inc.**:** 65-97.

Lott, A. J. & Lott, B. E. (1965). Group cohesiveness as interpersonal attraction: a review of relationships with antecedent and consequent variables. *Psychological Bulletin, 64*, 259-302

Meyers, S. A. (1997). Increasing student participation and productivity in small-group activities for psychology classes. *Teaching of Psychology, 24*(2), 105-115.

McNeill, D. (1992). *Hand and Mind*. Chicago and London, The University of Chicago Press.

Tscholl, M., & Dowell, J. (2005). Differences in f2f and case-based CSCL. (*in preparation*)

Veerman, A. L., Andriessen, J. E. B. & Kanselaar, G. (1999). Collaborative Learning through Computer-Mediated Argumentation. In Hoadley, C. (Ed.) *Proceedings of CSCL99*, Stanford, Palo Alto. LEA.

Weiner, M., and Mehrabian, A., 1968, *Language within Language: Immediacy, a Channel in Verbal Communication,* Appleton, New York.