

Using Annotations in Online Group Chats

Na Li, Mary Beth Rosson

College of Information Sciences and Technology
Pennsylvania State University
{nzl116, mrosson}@ist.psu.edu

ABSTRACT

Annotating documents has long been a widely used strategy for distilling important contents and externalizing related thoughts and ideas in context. No one has studied the activity of annotating dynamic texts, such as online chat, although online conversation is an important communication media for global companies. In this paper, we investigate Instant Annotation (IA), a real-time annotation-enhanced chat tool. We contrast the use of the enhanced chat tool to a standard chat tool for multilingual groups doing a brainstorming and decision-making task. Results show that group satisfaction and perceived control of the conversation are enhanced for the participants who used IA. We also report new patterns of annotation use and discuss design implications for group chat tools.

Author Keywords

Instant annotation; conversation control; design; evaluation

ACM Classification Keywords

H.5.3 [Information interfaces and presentation]: Group and organization interfaces – Computer-supported cooperative work.

INTRODUCTION

Collaborative annotation tools provide users the ability to connect text-based comments or other reactions to an anchoring context in the source material; the annotations may also offer an additional space for interaction among users. Studies of collaborative annotation tools suggest that serious challenges arise for designers when the annotation features must be integrated with real-time communication and collaboration [4]. Because synchronous interaction already demands considerable attention, adding annotation as a subtask may be difficult; even more demand arises from the implied need to attend to and interact with others via the resulting annotations. For instance, Kelkar et al. [4] investigated a collaborative tagging system for real-time audio meetings that allowed users to annotate utterances as they are indexed on a timeline (the system did not support re-

plies or other elaborations of the annotations). They found that users were not very good at multitasking between tagging and active participation in the meetings, presumably because of the extra cognitive demands.

In a previous work [6], we described a different approach to collaborative annotation – a design concept called Instant Annotation (IA; see Figure 1). The IA design features a collaborative annotation sidebar that can be used to react to the text-based utterances of a group online chat. In the group discussion snippet shown in Figure 1, two collaborators are discussing a shared research project, and the annotations mark an important piece of content (due date), a meeting proposal, an idea and reaction, and a to-do item.

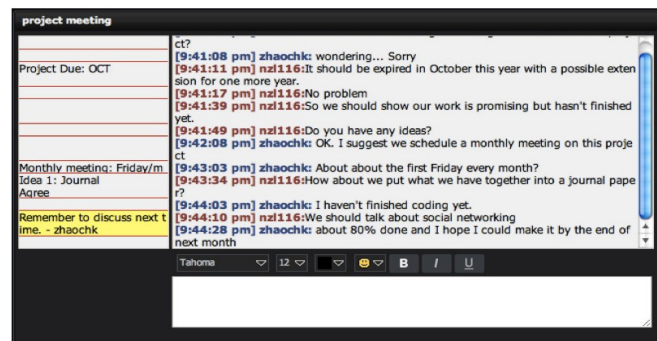


Figure 1. Screen capture of the IA prototype; annotations are in the side bar on the left.

Each annotation section in the IA sidebar is anchored to its context (i.e., a line of text from the chat) in the main chat window next to it. Users simply click on a section to make it active and type in an annotation. If a user wants to attach a second annotation to the same context, it is stacked immediately below the existing one. At any time, users can scan annotations currently in view or scroll up in the chat log to see previous annotations. For long or stacked (multiple line) annotation lists, the user simply mouses over the annotation section to expand it and show the full listing.

In this paper, we report whether and how a working version of the IA design might enhance brainstorming and decision-making tasks for small online groups. Specifically, we used the prototype to explore three research questions: How is an IA chat tool used? How is user experience affected by the IA functionality? What are the costs of IA in group chats?

USING INSTANT ANNOTATION IN GROUP CHATS

To answer the three research questions, we conducted a comparison study using mixed methods (behavior, surveys, interviews). In this paper we report data collected from two

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright 2014 ACM 978-1-4503-2473-1/14/04...\$15.00.
<http://dx.doi.org/10.1145/2556288.2557209>

conditions run at different points in time; they were identical except that in the first condition the IA tool was not available and in the second it was. We collected a mix of dependent measures, including the types of annotations made, ratings of discussion quality and satisfaction, and comments from a post-task interview.

Participant nationality (American or Chinese) was a second research variable. In an earlier study [5] we found that in multilingual group chats (American and Chinese speakers), the imbalances in English proficiency led to communication problems. Reasoning from participants' reactions to a design sketch of the IA concept [6], we hypothesized that the IA space might provide a second communication channel in online chats, and that it might compensate for some of the communication problems of non-native speakers.

Task & Procedure

We adapted a task from Freiermuth and Douglas [2]. Four participants formed a group to discuss how to spend \$5000 to support environmental sustainability; they were told to generate at least eight ideas in 15 minutes. At the end they were asked to choose the three best ideas. This combination of brainstorming and decision-making within the task make it similar to real world situations where formal discussions occur (e.g. in business or academic settings). In one condition (Non-IA), participants used AIM, which is a standard chatting tool. In the other (IA), participants used the IA tool depicted in Figure 1. After the task, participants completed a post-experiment survey. Finally, individuals were interviewed separately about their experiences.

Survey

The post-experiment survey included 10 Likert-style rating scales adapted from a prior study of online group discussion [1]; scale values ranged from 1= Strongly Disagree to 5=Strongly Agree. Six scales assessed *communication quality* (e.g., "The conversation seemed highly interactive" or "I found it difficult to keep track of the conversation"); four others assessed *satisfaction* (e.g., "I enjoyed communicating with my group members" or "I am satisfied with the group decisions that were reached"). In condition 2 (IA), an additional five scales probed reactions to the IA tool (e.g., "I could easily review messages using this chat tool" or "I am distracted by the annotations in the side bar").

Participants

Each condition included five groups of four participants; each group included two native and two non-native speakers. The groups in the two conditions were constructed to be similar in terms of education level, gender, and language proficiency. This enabled a valid comparison of individuals' group chat behavior and reactions when using two different chat tools, for native and non-native speakers.

DATA ANALYSIS AND RESULTS

We examined impacts of IA on both groups and individuals. We found that the presence of IA did not affect group performance, but that it did seem to influence *individuals'* chat experience, which we discuss below in three subsections,

working from a mix of qualitative (annotations and interviews) and quantitative data (frequencies and rating scales). We did open coding on annotations and interview transcripts to find common themes. A second pass over the interview transcripts' open codes was performed using axial coding to look for data that could help to explain interesting results in the quantitative analysis.

How Was the IA Tool Used?

The number of annotations generated by the four IA groups varied from 12 to 23, for a total of 90. Thus participants did indeed use the IA function, but there was also considerable variability in how much groups employed the feature. Table 1 summarizes the ways in which annotations were used, organized into five major use patterns.

Category	Definitions and Examples	N
Summarize	Summarize ideas ("idea 3: awareness competition to reduce energy"); Categorize ideas ("Idea 3.1"); Summarize decisions ("Top3 Ideas: recycling competition, PSU Green Week, Intersting Green activities")	64
Communicate	Ask questions ("Can anyone price this for us?"); Reply ("i'm not sure about the cost of monitor reporting, but i would think so - facilities usually has that info"); Express support ("Nice idea!")	13
Vote	Vote for preferred ideas ("vote from winston")	5
To do	Reminder of things to consider or check later ("ToDo: check how many dryers are in the building")	3
Add Information	Extend an idea ("Competition for greenest lab on campus"); Propose an idea ("Temperature should be able to adjusted in each room....I hope...")	5

Table 1. Five categories of annotations

Note that even though IA was designed as a side channel next to the main chat, people sometimes used the space to expand on an idea or propose a new idea. To better understand why users might add information in the IA space rather than in the main chatting window, we carefully examined the five cases of Add Information. In one case, a new idea was proposed in the side bar. The new idea pertained to the planning theme that was under discussion at that point in the main chatting window. In this group, group members had first identified several themes, then went back to develop ideas within each theme. The annotation was one such idea, suggesting that it was a convenient place to inject new ideas into what was already in the chat.

In other cases, Add Information annotations extended an idea that appeared in the chat; they were positioned right next to the original idea. These examples suggest that users' motivation for adding information in the IA space was to organize discussion content by topics or themes, perhaps with the result that these points of discussion also became more visually salient.

Looking more generally at the kinds of annotations produced it is clear that the most popular use was Summarize, which accounted for 71% of the annotations. In the inter-

views, many participants reflected that they relied on the summarizations in the sidebar to make their final decisions.

How Was User Experience Affected by the IA Tool?

To explore user experience impacts of the IA tool, we analyzed participants' ratings of their discussion experiences, focusing particularly on judgments of perceived quality of the communication process and their satisfaction with the discussion. To better understand differences in these ratings we also draw from the individual post-task interviews.

Quality of Communication (QC)

After confirming satisfactory internal reliability (Cronbach $\alpha = .73$), we aggregated the six rating scales assessing quality of communication into the QC construct. A two-way ANOVA with Tool and Speaker-type as between-subject factors revealed a main effect of Speaker ($F(1,36)=6.11$, $p<.02$), but no effect of Tool or interaction of Tool and Speaker. Native speakers provided generally higher QC ratings (overall mean 3.94) than non-native speakers (overall mean 3.41).

Although there was no main effect of Tool on overall quality of communication, inspection of the individual items comprising the QC scale suggested that the one item focused on *conversation control* ("I was able to take control of the conversation when I wanted to") was sensitive to the presence of the IA feature. For the non-IA condition, the mean ratings for native and non-native speakers were 4.11 and 2.60; when using IA, the means for both speaker types were higher (4.30 and 3.30). An ANOVA on this item revealed main effects of both Tool ($F(1,36)=4.34$, $p<.05$) and Speaker ($F(1,36)=34.66$, $p<0.001$), suggesting that both the IA tool and language fluency may have influenced perceptions of conversation control. Even though non-native speakers felt less control in general, speakers who used the IA tool felt that they could control the conversation more than those who did not use the IA tool. Given this pattern, we re-examined the interview transcripts, which provided three explanations about how and why the IA feature may have enhanced conversation control. We also derive design implications from this qualitative analysis, which are discussed in corresponding sections.

Mitigating Attenuation Blocking

The interviews suggest that IA may mitigate the "attenuation blocking" problem common in group meetings; this problem was discussed in the seminal work of Nunamaker [8], who emphasized the downsides of participants not being able to contribute comments at the time they felt to be most relevant. One of our participants shared her frustration in traditional online chatting and how she used IA to overcome it, *"I ended up, you know, I started typing something, and then I wiped it out, because somebody else already raised the issue. ... That's why I like the annotation, because you can annotate your comment that pertains to that person's chat comment by their side and it makes more sense because if you put it into the conversation then you had these two or three conversations going on at the same*

time, and you were technically commenting something that was chatted on three minutes ago." (Interviewee 12, IA)

This participant had difficulty taking turns in traditional chats because she is not fast enough to contribute to rapidly moving discussion points. Using the IA space, she could add to any point at any time, even after the group moved on to other topics. When other people saw her annotation, they might move back to the discussion point she commented on. This happened twice in one group, where a participant suggested that the intent of a comment like 'Nice idea!' would not be clear in a regular chat. However, in the IA space, such a comment could be posted next to an idea, causing the group to further discuss the idea that had been called out as a good one. A design implication from this observation is that tools for group brainstorming should provide secondary channels for users who are slower in expressing ideas, to help them contribute to rapid discussions.

Supporting Distributed Cognition

IA can be seen as a support for distributed cognition among team members via collective annotations of ideas and reactions during chat. For example, one participant stated that she put complementary information in the IA space with the goal of keeping side points on the side, *"I want to comment on that one point, and I felt that was a better place in the annotation than in the chat sequence. Because it's more of a supporting statement, I wanted to put in like 'To do:' what need to be done to get that idea going. And I just didn't feel that it wants itself to be in the discussion, it's more like a side note of ok heads up, we need to do this for this idea."* (Interviewee 10, study 2)

This participant described an example of how team members used IA to build external representations of their knowledge structures; Hutchins [3] argues that this is critical for complex collaborative behavior. In fact, Summarize, Communicate, Vote, To Do, and Add Information annotations are five different instruments for supporting group members' distributed cognition. With the IA space, it is easy to externalize one's understanding of or reactions to the conversation, thereby sharing it with other team members. The expectation is that the group as a result might enjoy a richer shared understanding; this in turn can explain why participants using IA felt more control over the conversations. A design implication from this observation is that online chat tools should have features to help group members coordinate shared cognition through external representations like the five types of annotations.

Balancing Conversation Participation

Wooley et al. [9] found that 'equitable talk' predicts better group problem solving. Although IA did not change performance, it seems to promote perceptions of a better balance in conversations, which helps to address issues stemming from differences in language fluency. In the Non-IA condition, when asked whether there was a leader in their group, all groups reported that native speakers assumed leader roles in their multilingual groups. However, in the IA condition, no groups reported having a leader. It may be

that the IA feature reduced the tendency for groups to have dominant individuals controlling the conversation. In fact, all of the IA groups emphasized that the member contributions of ideas and annotations were well balanced. Even talkative people who tend to dominate conversations found a way to balance their input. They waited a bit before talking, and spent time instead doing annotations for their group. One native speaker shared this strategy of balancing the conversation in his group, *"I was the one who was doing all the annotations. I had a couple of ideas, but I kind of slipped and let other people talk, while I was annotating. And once it slowed down, I added my own ideas."* (Interviewee 13, IA) A design implication from this observation is that tools supporting text-based chats should provide secondary activities to engage fluent participants who are waiting for input from the less fluent.

Group Satisfaction (SAT)

The post-task questionnaire also included four rating scales that assessed satisfaction with the group. After confirming satisfactory internal reliability (Cronbach $\alpha = .89$), we aggregated the scales to form the SAT construct. A two-way ANOVA on SAT revealed a main effect of Tool ($F(1,36) = 5.92, p < 0.05$), but no effect of Speaker and no interaction, suggesting that satisfaction with one's group differs in the two conditions. Specifically, participants who chatted using the IA tool were more satisfied with their group (overall mean = 4.43) than participants who chatted without the IA tool ($M = 3.95$). This might be due at least partially to their increased perceptions of conversation control.

What Are the Costs of the IA Tool?

Finally, the post-task questionnaire included two subconstructs designed to evaluate the cost of using the IA tool. Two 5-point rating scales assessed ease of using IA (EIA); three others probed the perceived cognitive cost of IA (CIA). The average values for both constructs were moderately positive, with a mean of 4.0 for EIA and 3.93 for CIA. We also conducted independent t-tests to contrast the perceptions of native and non-native speakers with respect to these measures but found no significant differences. This suggests that even for users with relatively high cognitive loads (e.g., non-native speakers), annotations combine well with chatting. Our interviews also probed perceptions of difficulty in using the IA tool. Users' comments were consistent with the rating scale data, suggesting that people can multitask between annotations and the main chatting task.

CONCLUSIONS AND FUTURE WORK

Our research has shown that adding instant annotations during real-time collaboration is both possible and has promising consequences for chat participants. In terms of cost, participants seemed able to multitask well between the main chatting window and the IA side bar. An analysis of the use patterns of the IA tool suggested diverse and subtle ways in which it can assist group communication. Partici-

pants not only used instant annotation as a memory aid in summarizing, they also used it to interact with others. Their satisfaction scores as well as perceived capability of controlling the conversation were higher. The qualitative interview data extended our understanding of IA from previous work [6]. Users' comments suggested that IA might enhance feelings of controllability by mitigating attenuation blocking, supporting distributed cognition and helping to balance conversations. Based on this analysis, we proposed some design implications for designing lightweight secondary conversation streams for group chat.

In the future it will be interesting to extend the use of IA to real-time text chats for larger groups, where richer and more flexible representations of the secondary channel may be needed. Other features that may help with larger groups are labels to aid awareness of the greater number of participants; notifications of new annotations as they appear; and nested replies for more complex concepts and annotations. Further discussion of related design ideas can be found in an earlier paper that first introduced the IA concept [6].

REFERENCES

1. Convertino, G., Asti, B., Zhang, Y., Rosson, M. B. and Mohammed, S. Board-based collaboration in cross-cultural pairs. In *Proc. CHI'06 extended abstracts*, ACM Press (2006), 658-663.
2. Freiermuth, M. and Jarrell, D. Willingness to communicate: can online chat help? *International Journal of Applied Linguistics*, 16, 2(2006), 189-212.
3. Hutchins, E. *Cognition in the Wild*. MIT Press, Cambridge, MA, USA, 1995.
4. Kelkar, S., John, A. and Seligmann, D. D. Some observations on the "live" collaborative tagging of audio conferences in the enterprise. In *Proc. CHI 2011*, ACM Press (2011), 995-998.
5. Li, N. and Rosson, M. B. At a different tempo: what goes wrong in online cross-cultural group chat? In *Proc. Group 2012*, ACM Press (2012), 145-154.
6. Li, N. and Rosson, M. B. Instant annotation: early design experiences in supporting cross-cultural group chat. In *Proc. SIGDOC 2012*, ACM Press (2012), 147-156.
7. Nguyen, D. T. and Fussell, S. R. How did you feel during our conversation? Retrospective analysis of intercultural and same-culture instant messaging conversations. In *Proc. CSCW 2012*, ACM Press (2012), 117-126.
8. Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D., and George, J. F. Electronic meeting systems. *Commun. ACM* 34, 7, 1991, 40-61.
9. Wooley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. Evidence for a collective intelligence factor in the performance of human groups. *Science* October 2010: 330 (6004), 686-688.