Quantifying Collaboration on Computationally-Enhanced Tables

Meredith Ringel Morris, Terry Winograd
Stanford University Department of Computer Science
353 Serra Mall, Stanford, CA 94305, USA
{merrie, winograd}@cs.stanford.edu

Vision

Recent research in our group has focused on tabletop interface design, with an emphasis on identifying design guidelines that facilitate group work. We have developed several collaborative tabletop applications, and have evaluated them through formal usability studies and informal walkthroughs. However, while we have had success in evaluating the usability of specific aspects of the tabletop interfaces (aspects of what Pinelle et al. [5] refer to as "taskwork"), the impact of the interface on the "teamwork" dynamics is far more difficult to judge. We hope that this workshop will provide a stepping-stone for developing a standard for quantitatively and qualititatively describing the impact of groupware technology on collaboration.

Experiences and Challenges

We have designed and evaluated several four-user tabletop applications for DiamondTouch tables [1]. For example, we have developed collaborative software for selecting music for movie soundtracks that combines the shared tabletop with individual audio information for each group member [4]. More recently, we have been developing an application in cooperation with biologists that allows them to assign metadata to photographs they take in the field, and then to collaboratively perform queries over this set of photos.

Based on our experiences designing and evaluating these collaborative tabletop applications, we have identified several potential methods of measuring the degree of and quality of collaboration within the group. While Pinelle et al. [5] identify some "mechanics of collaboration," we have found these are not suited to the type of evaluation we are doing, since their mechanics are intended as a "discount usability" checklist for use during early prototyping stages, whereas we need criteria that can be observed/measured when studying a finished design with groups of real users. We therefore propose several metrics that can be collected when presenting an interface to a group of users. We assess their reliability in quantifying collaboration, and we discuss the practical challenges associated with collecting each metric.

(1) Amount of talking: The percent of time a group spends talking can be a good indicator of the degree to which they actively collaborated. While it isn't perfect (some key communication might occur through other channels, such as gesturing), when combined with other measures it can be useful. The distribution of how much individual group members talk can also be revealing – this can indicate whether the speech is truly collaborative (e.g., all group members speak with comparable frequency) or is simply a monologue by one individual who has taken over the task. However, measuring the speech/silence ratio in a group environment is difficult and subjective in practice. One possibility is to have one or more observers of the live event (or a videotape of the event) use a stopwatch to directly collect this data. Another possibility is automated analysis – extracting the soundtrack of a video recording of the session and using software to determine what percentage of that sound is silence. Both methods are challenging – using human timers is tedious and time-consuming, while computer analysis can be difficult if there is other noise in the environment (e.g., noise generated by the application itself or background noise such as fans or computer humming).

(2) Types of talking: In addition to measuring how much talking occurs, it may be useful to break this talk down into various speech acts. An understanding of the relative frequency of various types of speech (planning strategy, coordinating access to resources, asking advice of other group members, etc) in the various study conditions can contribute toward an understanding of how the technology has altered the

group's collaborative style. A standardized list of the types of speech acts that are relevant specifically to the study of shared-display groupware would make this process more meaningful.

(3) Distribution of actions among group members: For devices that can attribute identity to inputs (such as the DiamondTouch table, or a system with multiple mice), it is possible to automatically record which user performed each interaction. This data can be analyzed to give an indicator of collaboration by looking at the distribution of touches/interactions among users. A roughly equal distribution indicates a very different collaborative style than a distribution where one user has performed most of the actions. This analysis can be further broken down according to the type of interaction performed by each user, which can reveal strategies such as specialization.

(4) Location of interactions: For devices that can associate input events with specific users, it can be informative to analyze the location on the shared device where these inputs occur. Do users only interact with objects that are located near them on the table, or do they often perform actions in central regions of the table, or even in the areas closest to other group members? Such data can be indicative of collaboration – for instance, by revealing the presence of a shared region of the table where all users touch.

(5) Number of people that handle each object: Another measure of collaboration can be the number of people that interact with key objects in the application, such as digital photos, puzzle pieces, etc. A high score on this measure indicates that group members passed items among each other (multiple people handle each object), rather than working completely independently (only one user handles each object). Although this is often a reasonable indicator of collaboration, it does not necessarily reflect other possible methods of collaborating – for instance, instead of passing an object to other group members, one group member might ask out loud for others to look at it, but might do all of the handling of the object herself. Hopefully, this latter strategy would then be captured by our "amount of talking" measure.

(6) Reorientation of objects: People often orient materials such as text or images toward others to indicate willingness to share them [2]. For tabletop systems built using software such as DiamondSpin [6], which allows arbitrary re-orientation of objects, recording how often users reorient items can be indicative of the degree to which they are actively collaborating on those items with other users. It is not a flawless indicator, however – in our work, we have observed users employing several workarounds for the orientation problem, including users rotating their heads to read sideways text alongside another user rather than reorienting the item in question, or passing an item around the table (which could be captured by the "number of people that handle each object" measure), or moving the item to a central area for simultaneous viewing by all group members (which could be captured by the "location of interactions" measure).

(7) Task outcome: The outcome of the task the users are completing with a groupware application can be an indicator of successful collaboration. In theory, groups that collaborate more effectively should produce better results, since they have the input of more people's knowledge and skills. However, in our experiences testing tabletop groupware we have found that this is not always the case – sometimes "too many cooks spoil the broth," and users "over-collaborate" by second-guessing each others' answers, resulting in lower scores for groups that collaborated the most. This demonstrates the need for assessing not only *how much* users collaborate, but also *how effectively* they collaborate.

(8) Number of corrections: The degree to which group members correct or modify their work can be interpreted as an indicator of collaboration. In particular, we have observed that many of the groups that collaborate effectively adopt a strategy of double-checking pieces of a group task that have been completed by other group members. This tends to result in a higher number of items that have values assigned and then re-assigned, as compared to groups that do not collaborate closely in this manner. This indicator is not as useful for groups that adopt a serial strategy, such as a group where all members simultaneously focus on one item at a time. In that case, most of the "changing one's mind" about the correct answer occurs verbally, with the final agreed-upon choice being assigned only once, and no subsequent checking step. Hopefully the "amount of talking" measure would capture this difference in strategy.

(9) Time: The time taken to complete the task is a standard measure for single-user programs, but is not straightforward as a measure of how well groupware supports collaboration. For instance, a longer

completion time might indicate more collaboration, as it could reflect an increased amount of time spent in discussion with other group members, or an increase in time taken as a result of double-checking each others' work. However, a decrease in total time taken could also indicate high collaboration, for example if it reflects the fact that the group developed an effective strategy of parallelizing aspects of the task. A more effective use of time as a metric of collaborative activity might be to examine the time spent on specific activities (e.g., talking, interacting with various components of the software, etc.) rather than treating task time as a single entity.

(10) Learning: Something which can be difficult to assess, but is a good indicator of the collaborative benefit provided by an application, is the degree to which group members learn from each other during the task. Lave and Wegner [3] note that the phenomenon of legitimate peripheral participation, through which novices absorb knowledge and skills from observing the actions of more experienced co-workers, can be an important learning mechanism. If relevant for the chosen task, the degree to which users learn from working with others could be assessed through pre- and post-task questionnaires or interviews.

(11) Self-reports: One cheap method of ascertaining the amount and quality of collaboration is to ask group members themselves, either through post-task questionnaires or through interviews. This subjectively-quantified data can be useful, especially as a means for evaluating how some of the more speculative quantitative measures apply to particular tasks and groups.

(12) Strategy type: Understanding the impact of interface design on collaborative styles is challenging; coming up with a standardized taxonomy of SDG work strategies would be useful in this regard. In our usability studies, group work strategies could be classified roughly into one of three categories: parallel (all group members perform similar actions in parallel), serial (all group members focus together on one item at a time), and assembly-line (all group members work in parallel on different aspects of the task). It is not necessarily clear that any of these basic strategies is more or less collaborative than others in general, although some might be more effective for a particular task.

Workshop Goals

By participating in the CSCW 2004 Workshop on Methodologies for Evaluating Collaboration in Co-located Environments, we hope to learn from other participants about successful strategies they have developed for measuring and describing the amount and quality of collaboration among users of groupware technologies. We also hope to obtain feedback about our proposed metrics for collaboration, and perhaps, as a group, to begin to develop standard methods of measuring and classifying collaborative activity that can ultimately make knowledge transfer and results-sharing more effective for the CHI and CSCW communities.

Bio

Meredith Ringel Morris is a doctoral candidate in computer science at Stanford University. Merrie's research interests include human-computer interaction, computer-supported cooperative work, and tabletop interfaces. She received an MS in computer science from Stanford University and a BS in computer science from Brown University. She can be contacted at merrie@cs.stanford.edu.

Dr. Terry Winograd is a professor of computer science at Stanford University. His area of interest is human-computer interaction design, with a focus on the theoretical background and conceptual models. He directs the teaching programs in human-computer interaction and HCI research in the Stanford Interactivity Lab. He can be contacted at winograd@cs.stanford.edu.

References

1. Dietz, P. and Leigh, D. DiamondTouch: A Multi-User Touch Technology. *Proceedings of UIST 2001*, 219-226.

2. Kruger, R., Carpendale, M.S.T., Scott, S.D., and Greenberg, S. How People Use Orientation on Tables: Comprehension, Coordination, and Communication. *Proceedings of Group 2003*, 369-378.

3. Lave, J. and Wegner, E. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, UK, 1991.

4. Morris, M.R., Morris, D., and Winograd, T. Individual Audio Channels with Single Display Groupware: Effects on Communication and Task Strategy. *Proceedings of CSCW 2004*, in press.

5. Pinelle, D., Gutwin, C., and Greenberg, S. Task Analysis for Groupware Usability Evaluation: Modeling Shared-Workspace Tasks with the Mechanics of Collaboration. *ACM Transactions on Human Computer Interaction*, 10(4), December 2003, 281-311.

6. Shen, C., Vernier, F., Forlines, C., and Ringel, M. DiamondSpin: An Extensible Toolkit for Around-the-Table Interaction. *Proceedings of CHI 2004*, 167-174.