Evaluation of Interference During Collaborative Document Development

Jeffrey D. Campbell
Information Systems Department, UMBC
1000 Hilltop Circle, Baltimore, MD USA 21250
jcampbel@umbc.edu

Abstract

A problem with collaborative development of documents is that people can interfere with each other's work. If unresolved, this interference can result in an incorrect document. A evaluation scenario using the development of a database design is proposed as a good context for evaluation of collaborative systems for the development of both text and diagram documents. Several metrics and measures are described for quantifying the occurrence of interference. In particular, five survey questions were added to a standard usability instrument to address interference and protection.

1. Introduction

People working together at the same time on the same document can interfere with each other's work. They could make conflicting or inconsistent changes in a text editor. The two dimensional aspect of diagrams adds complications not found in linear text. The term "interference" is used here to mean an action, or actions, taken by a person that are not consistent with the intentions of another individual. By this definition, two people can cooperate in acting on the same item without it being considered to be interference. Such interaction can reflect the synergy possible with collaborative work. The focus of the research is on the development of documents that are a final output of the task, not just an ad hoc aid to communications like text chat or an electronic whiteboard. For the final deliverables, correctness is essential. Undetected interference can cause inconsistencies or errors that can make the document practically worthless. At a minimum, interference results in time lost resolving the conflict.

Some argue that the system does not need to explicitly prevent interference. Stefik et al. [1] observed that normal social protocols tend to lead people to avoid impacting other's work without permission. However, unintentional interference is outside of the scope of social protocols which can only work when people are aware of the interference.. Various strategies have been discussed for avoiding, preventing, detecting or correcting such interference. The need for such 'concurrency control' techniques in Computer Supported Cooperative Work (CSCW) and the problems with it have been discussed for at least the last decade, e.g. [2], [3] and [4]. Concurrency control corresponds to the 'Coordination of Action' and 'Protection' components of the mechanics of collaboration [5]. It occurs at the Technology level in the Collaborative Framework of [6]. The nature of interference and the impact on the people using the system is important because this information is needed to select the best control technique to use.

The next section presents a scenario that has been used for evaluation of both diagram and text document development. This database design scenario is promoted as one that has broader applicability. Sections 3 and 4 discusses the metrics and measures [6] used to quantify interference. The paper concludes with a discussion of key points.

2. Database design scenario

This section discusses the selection of the database design scenario and the details of its usage in two studies. The participants were each provided with a catalog from the same retail/mail order business and instructed to develop a database design for that company. They were instructed to start with the portion of the design that represents the items the company sells. The catalogs provided examples of the data requirements for these items. The catalogs could also be used to choose between design alternatives, for example, using "product ID" as the primary key instead of "product ID, size and color".

2.1. Scenario selection

The database design scenario has a number of desirable attributes:

Multiple Output Formats: The design can be represented in a variety of diagram and textual styles. An Entity-Relationship (ER) [7] diagram, (Figure 1) has rectangles representing entities, ovals representing attributes and diamonds showing relationships between entities. A frequently used text format to express the design is based on

Proceedings of the 10th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE '01)
1080-1383/01 $10.00 © 2001 IEEE

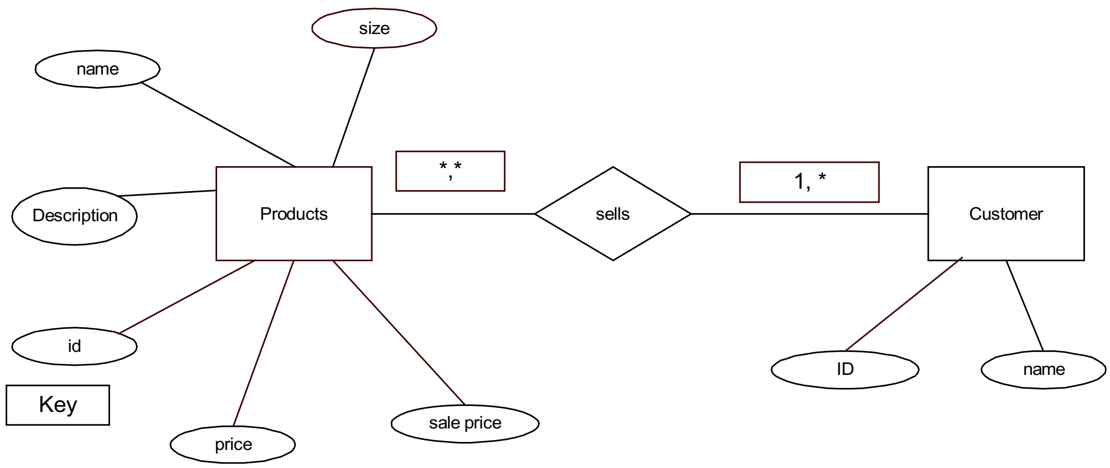

Figure 1 Sample ER diagram as drawn by participants.

mathematical set notation. Table 2 shows the same design with entity names followed by lists of attributes in parentheses. Table 3 is an implementation of the design using SQL, a common language for database definition. While textual development of a database design may not encompass all the tasks found in writing prose documents [8], it was sufficient for the interference studies. If prose is necessary, the scenario could be extended to add a memo or design report. Table 4 is a brief narrative describing the design illustrated above. Since the scenario can result in a different diagram and text formats, any variation in collaboration measures associated with format can be investigated.

Task Type: Using McGrath's classification[9], this is an intellective task. There is anecdotal evidence of usability testing of diagram and word processing software with scenarios in which the participants must recreate an existing document. This approach was not expected to adequately represent the group interaction and time necessary to create the design as well as to represent it in the document.

Realistic: A preliminary survey of information technology graduate students [10] and ethnographic observations of software engineering groups [10] confirmed that database design tasks are often the responsibility of a group of analysts. While such tasks are not currently supported by synchronous groupware, testing such applications with this scenario is not unreasonable.

Participants: Database design is taught in courses in numerous academic departments including computer science, information systems, information science and management information systems. With this variety of academic backgrounds, participants are likely to be more diverse and hopefully more representative of information technology professionals than a discipline specific scenario. For studies run in an academic environment, the multiplicity of courses should increase the participant

Table 1 Sample database design using mathematical set notation.

Products(<u>ID</u>, Name, Description, Size, Price, SalePrice)
Customer(<u>ID</u>, Name)
…

Table 2 Sample database design using SQL statements to define the database tables.

```
CREATE TABLE Products (
    ID Number,
    Name varchar(40),
    Description varchar(120),
    Size varchar(20),
    Price number(5,2),
    SalePrice number(5,2),
    Constraint ProductPK Primary  Key (ID));
CREATE TABLE Customer(
    ID number,
    Name varchar(40),
    Constraint CustomerPK Primary Key (ID));
…
```

Table 3 Sample database design as text narrative

There are two main entities and corresponding tables in the database. The first one is Products. The attributes of products are a unique identifier called ID which is the primary key for the table. Additional attributes include the name of the product, description, size, price and sale price. The second entity and table is called Customer and has a unique ID as the primary key and another attribute for the customer name…

pool. The task is similar enough to assignments in an introductory database course, that students should have the knowledge necessary to participate.

Correctness Assessment: It is also relatively easy to evaluate how well a design represents the data for a given set of business requirements. For example, the rules for database normalization can be applied to the design. This provides a clearer basis for the evaluation of the effectiveness of the software than graphically similar, but less well defined diagrams such as mind mapping diagrams or general text development tasks.

Open Ended: The task is large enough that it is unlikely that any participants will complete the entire task in an experimental session of reasonable length. This helps to eliminate upper bound limits on measurements of efficiency and productivity.

Unscripted: The scenario does not prescribe how the work is to be done, only the end document type is specified. This allows a group of participants to organize the tasks as they choose and select the appropriate transition tasks and times.

Group Size: The scenario can be used with individuals or groups. Especially with an expansion in the scope of the design, (for example to include order processing, warehousing, and supply chain issues,) the group size could easily be increased. A single user version of the ER software is currently under development to provide comparison for group and individual productivity, efficiency, correctness, etc.

Synchronicity: In addition to real-time document sharing, the scenario could be used for asynchronous work with either check-in/check-out or merge strategies.

Communications Support: Based on observation of groups in the initial studies, some level of communication services are needed to support the design task. However, these can range from text chat through audio to audio/video at varying quality levels.

Beyond the Lab: While the initial work was done as a lab study, database design tasks seem to be frequent enough in software engineering environments that additional ethnographic or field studies of the collaborative system could be completed.

Diagram Extensibility: Another factor in selecting ER diagrams for the initial study is that they are similar to many other types of diagrams based on nodes and edges, so it might be reasonable to extend the results to a broader range of diagram types.

2.2. Scenario implementation

The initial studies using this scenario involved groups of three people creating ER diagrams or text documents using set notation. Their work was synchronous. They each worked on separate computers. The shared document was updated after the completion of each action by any participant. They could talk to each other, but were unable to see each other's computer monitors. Figure 2 illustrates a typical session.

Figure 2 Initial study showing two participants with a divider between them

3. Metrics

Data collected during the initial studies included system data logs, observer notes and audio video recording. In addition, participants completed a demographic survey, a usability questionnaire, and participated in a brief interview.

User activity, such as adding, moving, connecting, disconnecting and deleting shapes, was recorded by the collaborative diagram software. Keystrokes and mouse clicks were recorded for the text version. The experimenter made contemporaneous observations of the group. Those notes were recorded using a custom application similar to those used in video analysis which wrote to the same system data log file.

A video recording was made from the computer monitor of the observer. This was scrolled and zoomed by the observer to display a reasonable overview of the current activity. Individual microphones recorded conversation from the three participants onto the stereo audio channels of the video tape. One participant's audio was recorded on the left channel, one on the right channel and the third on both channels. Signal processing could be used to separate the signals for analysis. In the analysis performed to date, headphones have been sufficient to identify who was speaking based on which ear or ears the sound was coming from.

The demographic data included questions about academic and professional background, familiarity with various software engineering techniques, experience with software packages and knowledge of business functions relevant to the scenario. There were questions that measured how well the participants knew each other.

The survey given to the participants after the session is the Computer System Usability Questionnaire (CSUQ)

[11] with additional questions specific to collaborative work. These questions are included in Table 4. Question 19 focused on awareness. Question 20 was the primary question about interference. Question 21 was designed to quantify the impact of any interference. Questions 22 and 23 were written to identify the frequency of over and under protection of their work from interference.

A factor analysis in a manner similar to [11] was performed on the responses from all 63 participants (57 diagram, 6 text) in the initial studies. This sample is smaller than Nunnally's [12] guidelines for identifying stable factors which suggests 100 participants for this size survey. Five factors were found to underlie the twenty questions analyzed from the enhanced survey. The first three factors roughly correspond to Lewis' factors. The last two factors are new and are uncorrelated with the questions in the base CSUQ survey.

The new questions added to this survey contribute only to factors 4 and 5, not to any of the factors identified by Lewis. Questions 20 (unexpected changes) and 21 (interference impact) compose factor 4 which can be described as the interference factor. Factor 5 consists of questions 22 and 23 about shape access and can be thought of as a protection factor.

Table 4 Additional usabiltity survey questions with Likert scale labels.

19	It was easy to know who was working on a shape. (strongly disagree - strongly agree)
20	Other people changed shapes I was working with in ways I did not expect (Never - Always)
21	If there were unexpected changes, how much did that impact your work? (No impact - Severe)
22	The system prevented me from working on shapes that I wanted to use. (Never - Always)
23	The system kept me from working with shapes that other people really were NOT using. (Never-Always)

4. Measures

The metrics described in the previous section were used to compute measures for interference and sharing. A third group of measures are more general and were computed as a basis for comparison for future studies with different output formats, concurrency control approaches, etc. A fourth section describes a useful visualization of the data. The measures are described here in the context of diagrams, but they can be applied to text documents by substituting "word" for "shape."

4.1. Measures of interference

The primary results are the rates of interference. This can be reported as the number of incidents per unit time or the number of incidents per shape (or word). Rates for specific types of interference based on particular actions, e.g. simultaneous text change for a diagram shape, can be reported as a rate per action occurrence.

The participants' impression of frequency and severity of interference was measured in the usability survey and observed through their contemporaneous comments and during the follow up interview.

Since two people working on the same shape is only interference if their intentions are different, another measure of interference is the percentage of same shape work that interferes versus cooperates.

The final measure of interference is the amount of time spent in recovering from any interference detected by the participants. This can be measured in person-minutes and in elapsed time.

4.2. Measures of Sharing

If one ignores the case when two people add new items in the same previously empty space, interference occurs only when more than one person acts on the same shape. The extent to which people work on more than one shape is important for interference research.

Sharing can be measured as the percent of shapes that are acted upon by more than one person. It could be important to distinguish situations where one person essentially gives the shape to another person and does not work with it again from situations where the most recent user of the shape changes back and forth. Determining the number of changes in person acting on a shape compared to the number of different people working on the shape measures this distinction. Each of these measurements can be computed for all shapes or for shapes of a particular type (e.g. entity shapes vs. attribute shapes in an ER diagram).

Closely related to sharing is a measure of the extent of overlapping actions taken by different people on different shapes. This essentially is a measure of how much parallel work was done. This can be quantified as the percent of actions taken by people within some short time (e.g., .5 seconds) of an action by another person.

4.3. General Measures

Standard concepts of productivity and efficiency can be easily measured in terms of shapes per unit time or correct shapes per unit time (to consider quality, not just quantity). Similarly, time to complete predefined portions of the scenario is another measurement.

The correctness of the design could be evaluated by a panel of experts. Self-evaluation by the participants could show differences between different systems or modes.

It may be possible to get some sense of individual contribution to the final document be examining the number of items (shapes or words) created by each participant that remain in the final document. Another measure of contribution might be the number of items that each individual was the last person to modify. (Such a measurement might have to adjust for minor actions, such as cosmetic movement of shapes. Otherwise the person who did a final centering of the diagram on the page would receive credit for all of the shapes.)

There may be some value to analyzing the types of actions and the types of shapes. If significant patterns are found, this increased understanding of diagramming behavior could guide development of better interference detection and control services. For example, one could measure the number of actions per shape and look for differences in the number of actions for different types of shapes. For example, different control mechanisms might be appropriate if it was confirmed that entities are rarely edited but have frequent connections while attributes are frequently edited but only connected once.

4.4 Visualization

Diagram interaction graphs, for example, Figure 3, were developed as a concise visualization of the sessions. The graph represents the sequentially assigned shape number on the vertical axis and time on the horizontal axis. The actions taken by each participant are represented by a different symbol on the graph. The graphs are especially helpful when color is used to represent the different participants. One can easily look at graphs from multiple sessions and see different patterns of interaction.

Actions on the same horizontal line represent actions taken involving the same shape. One can see how many and which shapes were used by more than one person. For example, the horizontal line in Figure 3 marks activity for shape 19. It was created (11:47) by participant C and modified a minute later. At 11:50, participant D worked with it and later (11:53) C worked with it again. Shapes in the same vertical line were acted on at the same time. These usually represent two shapes being connected together or disconnected from each other. Other causes of simultaneous actions are group moves and copy/paste of several shapes at once. For example, in Figure 3, the vertical line show participant C connecting shapes 2 and 19 at 11:53. Potential interference and cooperation is also observable as closely intermixed actions by multiple people on the same shape.

Future enhancements for the diagram interaction graphs will include an indication of the type of shape (entity, relationship, attribute, line). The type of action could also be represented with other symbols for adding, moving, deleting, connecting, or disconnecting shapes.

Sufficient data was recorded in the system log files to allow the session to be replayed and the diagram recreated. This would allow for repeated analysis and could support changes in time scale (fast forward, slow motion) and a different view (zoom or pan) from that recorded on the contemporaneous video.

5. Discussion

The interference results for collaborative diagramming have been reported in [10]. Work with collaborative text editing continues.

Database design development seems to be a reasonable task for collaborative system evaluation. The design document can be a diagram or a text. A reasonably wide range of people in an academic or software engineering environment are familiar with database design and ER diagrams. There are accepted standards for database design, making it relatively easy to evaluate the quality of the database designs to measure system effectiveness and efficiency.

Four of the new usability survey questions were found to correspond to factors identified here as “interference” and “protection”. The fifth question related to awareness and was not identified with any factor. The relative lack of awareness support in the systems being tested may have resulted in highly variable responses with lower correlation to the other questions. Neither of the two new factors overlap with the factors identified by Lewis. None of Lewis’ questions were significant components of the new factors. The conclusion is that the new questions appear to be a useful addition to the CSUQ questions for use with CSCW evaluations. The factor analysis should be repeated when more data is available to confirm these preliminary results.

The diagram interaction graphs provide a concise visualization for the large amount of log data recorded. The trick of recording three audio channels on a stereo videotape could have general applicability to usability evaluations with three participants.

6. References

[1] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar, "WYSIWIS revisited: Early experiences with multiuser interfaces," *ACM Transactions on Office Information Systems*, vol. 5, pp. 147-167, 1987.

[2] C. A. Ellis and S. J. Gibbs, "Concurrency control in groupware systems," presented at *Proceedings of the ACM Conference on Management of Data*, New York, 1989.

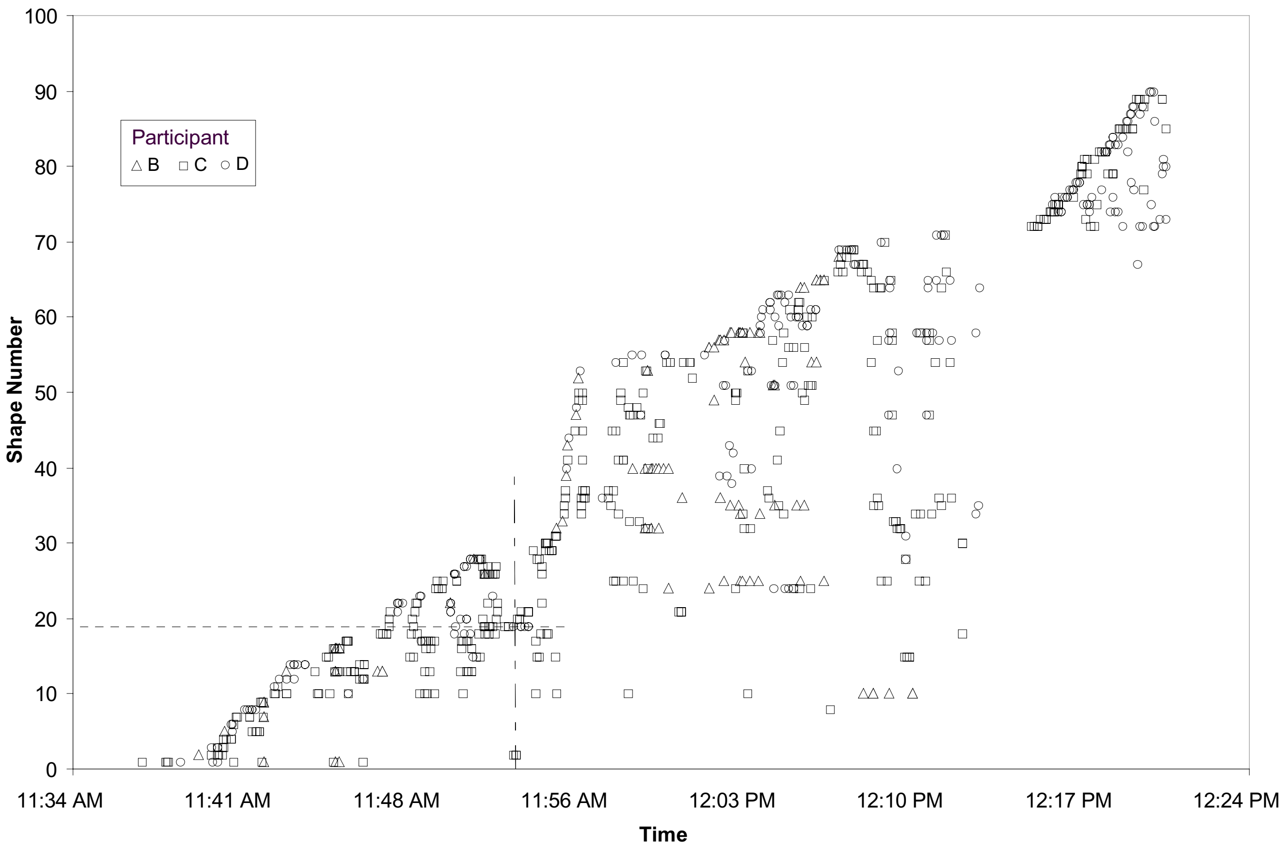

Figure 3 Diagram Interaction graph showing actions taken on shapes by participants B, C and D.

[3] S. Greenberg and D. Marwood, "Real time groupware as a distributed system: Concurrency control and its effect on the interface.," presented at CSCW '94 Proceedings of the Conference on Computer-Supported Cooperative Work, Chapel Hill, North Carolina, 1994.

[4] C. Allison, "Concurrency Control for Real Time Groupware," presented at Proceedings of Concurrent Engineering: Research and Applications 1994, Pittsburgh, PA, 1994.

[5] C. Gutwin and S. Greenberg, "The Mechanics of Collaboration: Developing Low Cost Usability Evaluation Methods for Shared Workspaces," presented at IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE '00), Gaithersburg, MD, 2000.

[6] J. Drury, L. Damianos, T. Fanderclai, J. Kurtz, L. Hirschman, and F. Linton, "Methodology for Evaluation of Collaboration Systems," DARPA Intelligent Collaboration and Visualization Program July 15 1999.

.

[7] J. E. McGrath, *Groups: Interaction and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[8] J. D. Campbell, "Consistency Maintenance for Real-Time Collaborative Diagram Development," in *Information Science*. Pittsburgh, PA: University of Pittsburgh, 2000.

[9] P. Chen, *The Entity-Relationship Approach to Logical Data Base Design*. Wellesley, MA: Q.E.D. Information Sciences, Inc., 1977.

[10] L. Flower, *Problem-Solving Strategies for Writing*. New York: Harcourt Brace Jovanovich, Inc., 1981.

[11] J. R. Lewis, "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use," *International Journal of Human-Computer Interaction*, vol. 7, pp. 57 - 78, 1995.

[12] J. C. Nunnally, *Psychometric Theory*. New York: McGraw-Hill, 1967