# Lessons Learned: Using Contextual Inquiry Analysis to Improve PDA Control of Presentations

**Karen Cross, Adrienne Warmack, and Brad Myers**
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{kcross@andrew.cmu.edu, asw@cs.cmu.edu, bam@cs.cmu.edu}

## ABSTRACT

Contextual Inquiry is a method developed by Beyer and Holtzblatt for grounding design in the context of the work being performed. In this paper, we describe how we adapted the method successfully to perform analyses of pre-existing videotaped presentations. Our goal was to find improvements for a slide presentation program currently in development, called ‘SlideShow Commander,’ which runs on hand-held "Personal Digital Assistants" (PDAs). Contextual Inquiry provided meaningful data on the structures and typical problems found in presentations, on which we based our design ideas. We then further analyzed the Contextual Inquiry data, beyond what Beyer and Holtzblatt suggest. This new step provided a means to prioritize the design suggestions, as well as a way to motivate the potential commercial usefulness of the software. Deciding upon the value and direction of further effort is essential for software development; by using our adapted form of Contextual Inquiry, we were able to make and defend these decisions.

**Keywords**

Contextual Inquiry, Personal Digital Assistant, Palm Pilot, Usability Evaluation Method.

## INTRODUCTION

The Pebbles SlideShow Commander [10] uses personal digital assistants (PDAs), such as 3Com Palm Pilots and Windows CE devices, to control a presentation in Microsoft PowerPoint running on a PC. It provides a handheld-means for a speaker to navigate through PowerPoint slides, view notes associated with the current slide, view the list of slide titles, and point or draw on the screen using the PDA. The current system uses a serial cable, infrared (IR), or radio frequency (RF) transmission to enable two-way communication between the hand-held device and the PC.

The Pebbles group designed and implemented the current version of SlideShow Commander without using any formal evaluation of the domain for which it was intended — public presentations — or the usability of its proposed design. Further development and assessment of the software, however, was hampered by the difficulty of studying the impact of the software on the particular domain of formal presentations.

We decided not to design a formal experiment to assess the usability of SlideShow Commander because of the inability to control conditions. Presenters must be familiar with both the content of the talk they are giving and the software to be tested. Otherwise, any difficulties with information flow affected by the software would often be overshadowed by more severe presentation difficulties. If we relied on participants to use their own talks, the results could be confounded with order effects. Speakers improve with practice, navigating across familiar slides more easily, referring to notes progressively less, and incorporating questions occurring in front of practice audiences into the content of the talk. Even if we could balance for order across subjects, each talk is so individualized that comparing across subjects seemed like a poor measure of improvement. Thus, the nature of the domain made the experimental approach less likely to produce authentic and meaningful results.

Further, the nature of the software itself made quantitative assessment using any approach difficult. The impact of the software could potentially be an absence of sporadic interruptions in communication, rather than the presence of any positive affect. Initially, we had little idea of the common styles of presenters – their reliance upon the Notes page in Power Point, the amount and types of AV equipment used, the means by which they changed slides, or their use of the cursor to point or draw on the screen. Background literature searches revealed a large quantity of information about how one *should* give a presentation (e.g. [5][6]), but little on what people *actually* did. One exception to the latter is Nelson et. al. [8], which included observations of fourteen presenters while giving talks. Although these authors found some common trends, such as the reliance upon paper copies of slides and the difficulty in switching between multiple presenters and files, the authors did not provide a breakdown of who the presenters were or what types of talks they were giving. Without these details, it was hard to rely on the validity and generalizability of those results. Thus it was difficult to determine a clear measure of the potential impact of the software based solely on the Nelson study.

In order to gain an understanding of the ways people actually use PowerPoint, and to measure potentially subtle effects, we decided to study the domain both with and without use of the SlideShow Commander. The method that seemed the best for this goal was Contextual Inquiry (CI), a technique originally developed by Beyer and Holtzblatt [1][2]. This method recommends observing work as it occurs in its own context, and using a graphical modeling language to describe the work process and to discover places where technology could overcome an observed difficulty. We chose this method because it would provide actual data about the detailed structure of the presentations themselves. This would, in turn, help us to: 1) predict how the proposed software would impact presentations; and 2) guide future development of the software.

CI centers upon live interviews with current or potential users at the workplace, based on previously determined foci. We had been taught the technique from graduate lectures in combination with a textbook by Beyer and Holtzblatt[1], and had used the technique exactly as Beyer and Holtzblatt recommended in several class projects. However, for the current study we had access to a large library of previously videotaped presentations. Rather than discard this valuable (and easily available) resource, we decided to adapt the CI technique to use video instead of live observations.

A literature search revealed an abundance of articles about how to perform CI (e.g. [1][2]), but unfortunately not as much on the actual experience of normal, non-expert practitioners using it in the field. The best reference we found for our case studies [11] provided some insights into the variety of domains to which the technique can be successfully applied, but not many practical suggestions that we could specifically apply to our domain. We decided to use the technique nonetheless, agreeing with Beyer and Holtzblatt's argument on the importance of context, and the usefulness of their graphical representation of the information from the work – what they call 'work models.'

The advantage of a live CI is that you can ask the participants questions and prompt them for explanations. However, this would not be appropriate for observing presentations, where an uninterrupted observation is actually preferable. In cases where interruptions are a distraction, Beyer and Holtzblatt recommend taking copious notes and interviewing the participants immediately after the observation period. Unfortunately, we did not have access to presenters immediately after their talks, so we decided to piece together motivations and goals from the evidence of the videotapes alone. These inferences were fairly easy to make based on the verbal comments and behaviors of the presenters when the problems occurred, although we may have lost some information about personal frustrations and desires.

We further differed from Beyer and Holtzblatt's method by performing additional analyses based on the problems identified in our work models. Because we had the advantage of videotape, we were able to measure the duration of each problem (where applicable) and rate the severity. These ratings enabled us to build evidence that the software authors of SlideShow Commander could use to prioritize their development efforts.

We used our adaptation of CI to study a varied selection of presentations that did not use SlideShow Commander ("pre-software talks"). From this analysis, we produced sets of models that gave us an idea of the flow and characteristics of presentations we were studying. Further, these models revealed interruptions in the talks, henceforth called 'breakdowns.' Many of these breakdowns – including the most frequent ones – could potentially be remedied by SlideShow Commander. By analyzing the breakdowns, we found evidence that at least potentially, the software would produce smoother talks with fewer difficulties. To confirm this analysis, we introduced the software into two practice talks given by novice users, ("post-software talks") and performed CIs on these talks to see if the communication was indeed smoother. We complemented this analysis with a questionnaire designed to assess the user satisfaction of the software.

We found that CI using pre-existing videotapes produced an abundance of data that we used to develop our design ideas. Because we had videotapes, we saw problems that a single live viewing while taking copious notes would not have detected. Further, the videotapes allowed us to return to earlier talks to concentrate on detecting subtle problems that were only revealed in later tapes. Finally, the further analyses we performed on the severity and duration of the breakdowns, which could not have been

| | Presentations | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | CI-1 | CI-2 | CI-3 | CI-4 | CI-5 | CI-6 | CI-7 | CI-8 | CI-9 |
| **Group Characteristics:** | | | | | | | | | |
| Remote audience | No | No | No | No | No | Yes | Yes | Yes | Yes |
| Composition | Hierarchy | Hierarchy | Peer | Peer | Peer | Peer | Hierarchy | Hierarchy | Hierarchy |
| Audience size | <25 | >50 | 25-50 | 25-50 | 25-50 | <25 | <25 | <25 | <25 |
| **Task Characteristics:** | | | | | | | | | |
| Main goal of talk | Latest wrk | Visionary | Exp. res. | Demo | Exp. res. | Latest wrk | Prog. rpt | Demo | Prog. rpt |
| **Context Characteristics:** | | | | | | | | | |
| Time Pressure | No | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Evaluative Tone | Critical | Critical | Support. | Support. | Critical | Support. | Support. | Support. | Support. |
| # Presenters | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 |
| Gender | M | M | F | Both M | F | M | M | M | M |
| Power Point Used | Yes | Yes | Yes | No | Yes | Yes | Yes | No | Yes |

Table 1. Characteristics of the Presentations Selected for Analysis. ‘Latest wrk’ stands for ‘Latest work;’ ‘Exp. res.’ stands for ‘Experimental Results;’ ‘Prog. rpt’ stands for ‘Progress Report;’ and ‘Support.’ stands for ‘’Supportive.'

determined without using videotapes, gave us the evidence with which we could convince the software developers of the importance of our design suggestions. The advantages of using video as an analysis tool is well documented in the literature (e.g. [4]). Our results showed that its usefulness carries over to other analysis techniques, such as CI.

In this paper, we describe the technique we used, the results from analyzing pre-software talks, the confirmation of the results using the post-software talks, and our evaluation of the method we used.

## METHODS FOR “PRE-SOFTWARE” TALKS

### Types of presentations

Nine academic talks presented at Carnegie Mellon University between May 1998 and June 1999 were analyzed by videotape. In this paper, the terms ‘presentation’ and ‘talk’ will be used interchangeably to mean talks given with or without interruption to an audience, followed by a question and answer period. All of the presentations we examined through CI covered various topics in computer science. Presentations can be classified on several dimensions: remote or face-to-face talks; formal, informal, or practice talks; demonstrations, experimental results, or theoretical perspectives; and multiple or single, male or female, presenters. We selected talks that provided as wide a cross-section of these types as possible.

The description of each talk in terms of its group, task, and context characteristics is summarized in Table 1. These characteristics were selected from the analysis of group support systems presented in Nunamaker et. al. [9]. Values for some of these characteristics were obvious from the presentation. Others — the task, group composition, time pressure, and evaluative tone for each talk — were determined by analyzing the number and types of questions from the audience, as well as the time spent on demonstrations compared to straight lecture. PowerPoint was used in 7 of the 9 talks — these presentations contained an average of 35 slides each. One of the talks (CI-2) was a formal presentation by a nationally renowned person in the computer science field to members of Carnegie Mellon University in a large lecture hall. Two talks, CI-4 and CI-8, were predominantly aimed at presenting demos of new applications and hardware to a small audience eager to see new technology. Other talks presented progress reports to other members of the same research group. Still others presented experimental and other usability studies to argue for the success of their ideas and designs. Across all of these talks, some presenters were welcomed with a supportive tone, few critical questions, and many shared experiences; others were interrupted frequently by criticisms or contradictory evidence.

Four of the talks analyzed (CI-6 through CI-9) were presented simultaneously to a local audience and to a remote audience communicating over NetMeeting software and live video-conferencing. The remote audience was located in Europe; the local audience was in the United States. The other six talks were presented locally only. Each presentation was divided between a 30 to 45 minute lecture period, and a question and answer period that ranged from 15 to 30 minutes in length. Both the lecture and the discussion periods were modeled using CI.

## Contextual Inquiry

We created work models using the method of CI as described by Beyer and Holtzblatt [1][2]. In our study of formal presentations, 'work' consisted of giving a public presentation. As part of this method, we drew four types of work models for each presentation: sequence, cultural, flow, and physical. As Beyer and Holtzblatt explain, **sequence models** show "the detailed work steps necessary to achieve an intent;" **cultural models** illustrate "constraints on the work caused by policy, culture, or values;" **flow models** represent "the communication and coordination necessary to do the work;" and **physical models** illustrate "the physical structure of the work environment as it affects the work" [1, p. 86]. The final work model described by Beyer and Holtzblatt, the **artifact model** — a detailed view of the tools used by people for their work — was excluded because of the limited view we had of the user's specific behaviors on their PC from the videotape.

While creating each model, we identified breakdowns in the flow of the information between presenter and audience. 'Breakdowns' were considered to be any interruption, confusion, or awkwardness in the flow of information between the presenter and audience. We used the breakdowns to inspire our design ideas. So far, except for the use of videotapes, we conformed to the proscribed method of CI.

At this point in our analysis, however, we added a final step to Beyer and Holtzblatt's procedure. We took our master list of breakdowns, and classified each breakdown according to the severity ratings shown in Table 2. Breakdowns were judged to be the same across different presentations if they had the same potential fix. We also recorded the duration of each breakdown, to the extent possible. We then analyzed this master list to understand the severity and frequency of these problems, and the impact of the software we were analyzing.

## RESULTS – PRE-SOFTWARE TALKS

### Common characteristics

We recorded the characteristics of the presentations we analyzed to ensure a variety of talks were studied. More than any other characteristic, the main goal of the talk was the biggest factor in the flow of the presentation. Talks that were demo-intensive used a greater number of artifacts, and had a resulting higher number of breakdowns in flow. Switching from one medium to another, or one application to another, almost invariably caused mistakes or delays. Talks that presented experimental results or future visions typically had a lot of carefully detailed slides, and thus a lot of slide navigation issues.

The talks presented simultaneously to both a remote and a local audience had a universally supportive tone, but this could have been due either to the novelty of the medium or the fact that the talks were within a single research group. Although we would not have been surprised to observe some gender effects on presentation style, the models we developed did not differ along this characteristic, apart from the observed tendency for more females in the audience to ask questions when the presenter was female. (Because of the source of the videotapes, the audience was roughly the same in many of the talks, and so such differences in behavior may not have solely been due to a higher percentage of female audience members with female presenters).

As predicted by [8], the one talk we analyzed with multiple presenters did have some difficulty during transitioning from one presenter to the other. However, unlike [8], we noticed that a bigger difficulty with multiple presenters was in keeping to the required time limit. This talk was the only one to go knowingly significantly over the time allowed, even though this time limit was the roughly the same as in other talks.
The tone and audience composition affected the cultural models for the different talks. More critical tones and more hierarchical (rather than peer) audience structures resulted in a more ’demanding’ cultural relationship (question-answer), as compared to the more collaborative and sharing cultures in the other groups. Audiences in the critical, hierarchical talks requested specific information, and demanded responses to various objections to the points raised in the talks. Audiences in the supportive, peer talks offered suggestions to improve the research, and generated more of a two-way discussion between the presenter and the audience.

Our work models identified a variety of presentation activities and breakdowns, many of which were overlooked upon first viewing the presentation videotapes. For example, it was not until an audience member specifically asked for a reference to be written on the board (in the fifth talk we examined) that we realized the difficulty most audience members must have when references are given only verbally, or skimmed over in slides. None of the talks we studied used handouts, so audience members had no way to find the references after the talk. Once we noticed this breakdown in a single talk, having

| Rating | Explanation |
|---|---|
| 5 | Breakdown *cannot* be resolved. The presentation cannot be continued and must be postponed until later. |
| 4 | Parts of a presentation *cannot* continue. The presenter omits part of the presentation. |
| 3 | Breakdown causes the presenter to alter the presentation or style of speaking to continue. Causes serious delay in presentation (at least 8% of the speaking time, or five minutes in an hour presentation). |
| 2 | Breakdown causes disruption to the flow of the presentation. |
| 1 | Breakdown causes minor inconvenience to the speaker or audience. |

Table 2. Severity Ratings of Breakdowns

videotapes allowed us to return and look at the other tapes and notice additional places where this problem had originally been overlooked.

When the same types of models were compared across different talks, as Beyer and Holtzblatt recommend, common structures and problems were revealed. The roles of host, technician, advisor, and fellow research member were repeated across many of the talks. Audience members wanted information as clearly and as completely as possible. Presenters wanted to impress the audience, but they also wanted to stimulate interesting discussion and help their audience understand the material.

**Physical models** for each presentation described the arrangement of the room, the positioning of the presenter and audience, and difficulties that arose because of these physical properties. Due to the nature of communication over a distance, the talks where audience members were distributed across two locations required a different physical arrangement from the purely local talks. In particular, remote talks used more artifacts, including cameras, screens, and sound control boards. All of the presentations, however, used a seated audience and a standing presenter; at least one screen that showed the changed slides, started demos, operated the VCR, and spoke. Physical breakdowns included presenters tripping over cords and having difficulty reaching controls for VCRs or slide advancement. Figure 1a shows a physical model for CI-6.

**Sequence models** illustrate the basic sequence of activities observed in a presentation. The steps include showing slides, demos, or video clips at appropriate times; and responding to questions from the audience. Examples of breakdowns we observed in the sequence models included applications not launching successfully, or presenters making errors while navigating either forward or backward through their slides (Figure 1b).

**Flow models** reveal the spread of information across various artifacts and people. The flow models for the talks showed that the flow of information was often interrupted, either briefly (for a few moments), or significantly (for almost 10 minutes during one presentation). Further, some talks, in particular the remote ones, were overwhelmed with artifacts: multiple electronic devices that the speaker and technicians had to manipulate. Breakdowns we observed in the flow models were revealed most frequently when the information flow from one artifact to the presenter or the audience was interrupted, such as when extraneous application windows blocked part of the screen, or when sound controls were not adjusted properly (Figure 1c).

**Cultural models** illustrate the wants and desires of participants, and how different roles interact. The presentations under study contained similar cultural models: in each, there were two main roles, the presenter and the audience. The presenter desired feedback, support, and questions; the audience desired clear and organized information. Each presentation, however, also included one or more other subsidiary roles in the cultural model: the technical support; the host; advisory committees; and members of the presenter's immediate research team. All of the people fulfilling these roles helped to make the communication between the audience and the presenter as smooth and as informative as possible. Most breakdowns in cultural models occurred when the desires of the presenter to give information, and the audience to receive information, were obstructed. For example, the behavior of several presenters indicated a desire to be near the audience, but this conflicted with the need of the audience to receive visual information from the slides and for the presenter to stay out of the line of sight of the screen (Figure 1d).

**Work Models.** Contextual Inquiry work models of CI-6, a presentation given to a research group distributed across two locations. The physical model (Figure1a) illustrates the physical work environment. The sequence model (Figure1b) illustrates how work takes place over time. The flow model (Figure1c) illustrates the flow of information and work between people, places and artifacts. The cultural model (Figure 1d) illustrates influences on people and their work.

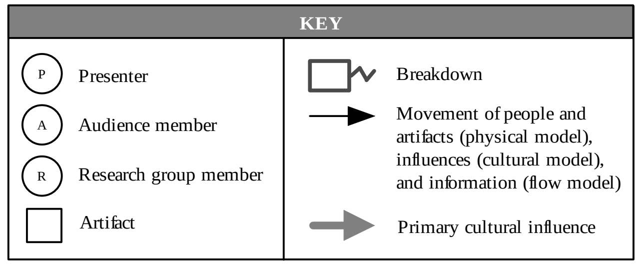
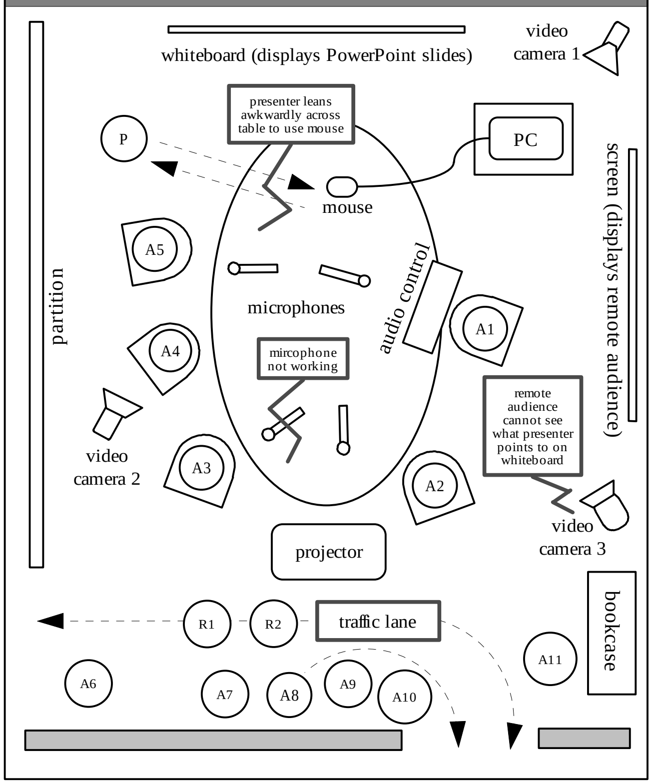


**Figure 1a. Physical Model**



| Figure 1b. Sequence Model (partial model shown) | | |
|---|---|---|
| **TRIGGER** | **INTENT** | **BREAKDOWNS** |
| 18. Question from remote audience member | Answer question | Audio unintelligible. Local members instruct remote members to adjust audio compression setting. |
| 19. Comment from remote member | Respond to comment | Audio unintelligible. Local members instruct remote members to reconnect. |
| 20. Comments from local members | Respond to comments by refering to slide from earlier in presentation. | Presenter tries to return to slide. Presenter rapidly searches through slides but cannot find it. |
| 21. Question from local member | Answer question | Presenter tries again and eventually finds slide. |
| 22. Local member asks presenter to bring up previous slide | Go backwards one slide | Presenter tries to go back one slide but goes forward one slide instead. |
| 23. Remote audience reconnected | Continue discussion | |
| 24. Question from remote member | Answer question | |
| 25. Comment from local member | Respond to question | Presenter flips through slides searching for "System Architecture" slide. |

**Figure 1c. Flow Model**

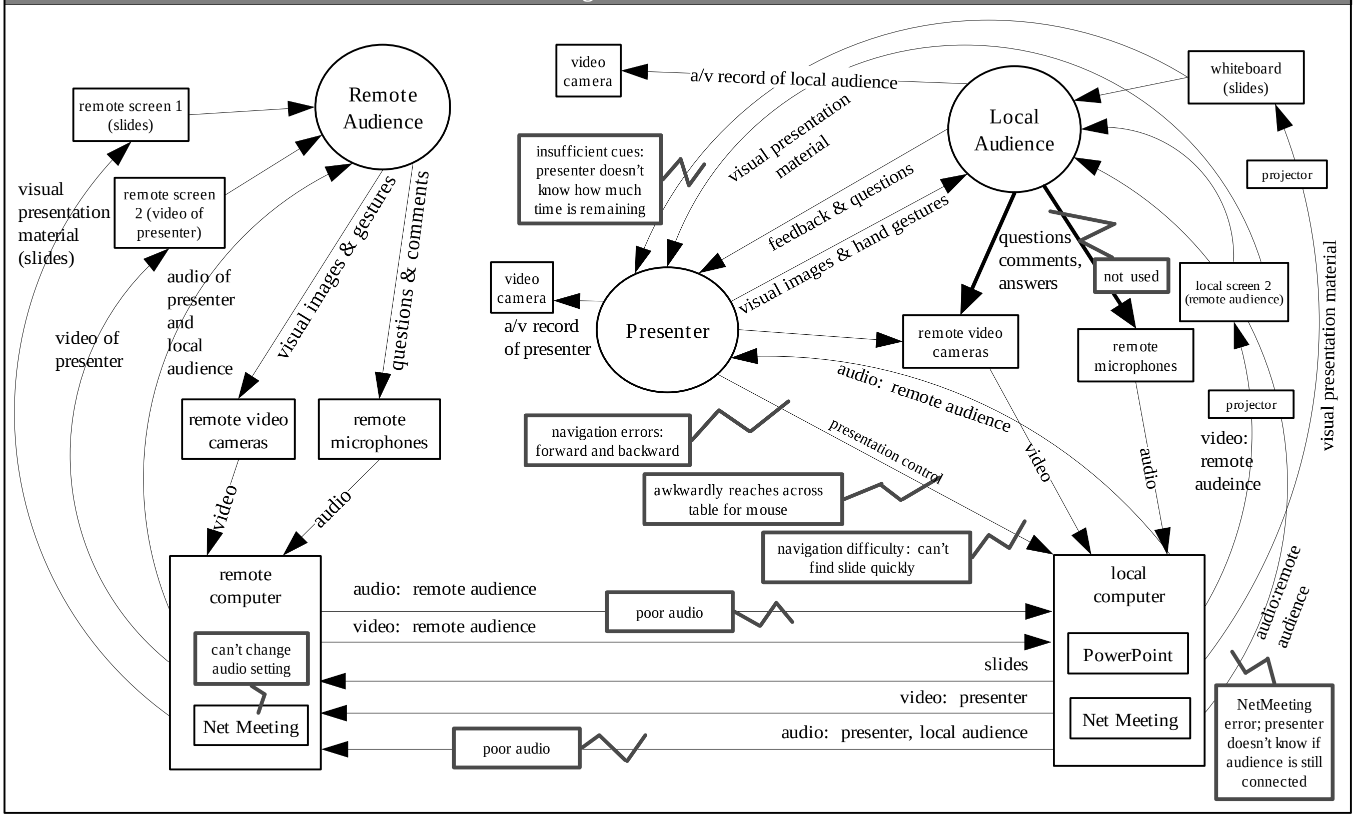
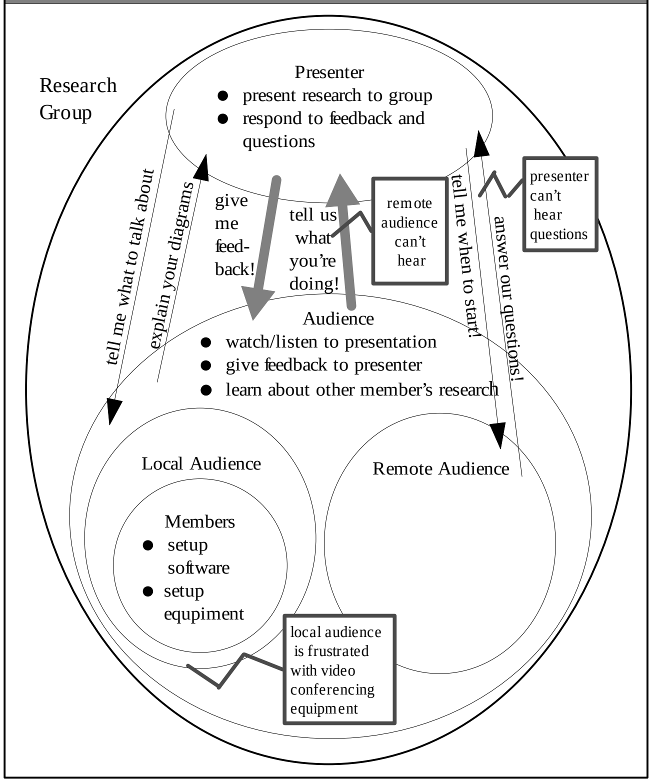


**Figure 1d. Cultural Model**



| Description | Model | % of Talks | Count (over all talks) | Average Severity | Average Duration (each time) |
|---|---|---|---|---|---|
| 1. Changing slides is difficult and awkward because of the placement of the mouse or laptop. | Physical | 67 | 166 | 1.2 | 2 sec |
| 2. Presenter loses track of time, must ask for verbal update. | Sequence | 44 | 6 | 1.5 | 55 sec |
| 3. Reference provided is incomplete or skimmed over; audience members would be unable to find it after the talk. | Cultural | 44 | 6 | 1 | 19 sec |
| 4. Camera view is unclear or pointed at wrong information. | Flow | 33 | 3 | 1.7 | 60 sec |
| 5. Audio level for demos is not set correctly. | Flow | 33 | 3 | 2 | 46 sec |

Table 3. The most frequent breakdowns observed in talks. '% of Talks' is the percentage of talks in which the breakdown occurred at least once. Count is the total number of instances of the breakdown observed across all talks. Severity is the average severity rating across all instances of the breakdowns. Average Duration is the average length of a single instance of the breakdown.

### Breakdowns

Thirty-eight unique types of breakdowns were found in the presentations by using Contextual Inquiries, with a total of 229 instances of breakdowns observed. Breakdowns were present during an average of 8.7 minutes of each talk. The most frequent breakdowns are summarized in Table 3. Each talk had an average of 34 instances of breakdowns, ranging from minor (Severity = 1) to moderately severe (Severity = 4).

The most frequent breakdown was the physical awkwardness of changing slides. Six out of nine presenters walked to one spot to talk, then turned and walked to a distance typically 3 feet away, positioned themselves, advanced slides using their PC, and then returned to the original spot from where they chose to talk. Often, the PC was on a low table, or difficult to reach, compounding the problem. This procedure wasted a total of between 30 and 78 seconds during each talk in which it occurred, with an average of 48 seconds. Another presenter found a less time-wasting solution: staying next to the slide control throughout the lecture part of their presentation, and then moving to a spot away from behind the podium and closer to the audience for the duration of the discussion period.

The second-most frequent breakdown was an inability of presenters to keep track of time or be aware of how much time they had remaining. Six of the presenters asked an audience member for a time check at some point during their lectures.

None of the observed breakdowns reached the highest severity — a permanent and premature end to the talk. However, three different talks had major breakdowns, requiring significant portions of the talk to be skipped. One had a demo that could not be shown because the PC lacked Shockwave. Two of the remote talks contained significant periods of time when the remote audience could not read the presentation slides, because of an unfocused camera and problems with the settings of the NetMeeting program.

## SLIDESHOW COMMANDER

The software we evaluated, SlideShow Commander [10], is a PDA-based application designed to aid in the delivery of PowerPoint presentations. It provides tools for presenters to navigate among slides, access presentation notes, and draw and gesture on the displayed slide.

The initial version of the application we studied, version 2.3, contains three main views: ‘Notes,’ ‘Titles,’ and ‘Scribble’ (Figure 3). The ‘Notes’ view displays the textual content of the PowerPoint Notes for the current slide. The ‘Titles’ view displays a list of the slide titles in a presentation. The presenter can navigate to a slide by tapping the stylus or a finger on the title. The ‘Scribble’ view allows the presenter to use the stylus to point or draw on the screen. Within any of these views, a presenter equipped with a PDA running SlideShow Commander can move linearly forward or backward through a series of slides by issuing “next” and “previous” commands on the PDA using simple gestures with a stylus. Finally, each of the four physical buttons is assigned to different modifier keys: <control>, <shift>, <alt>, <func>. These physical buttons could be used to send specific commands to the PC, such as F1, which displays a help screen.

## RESULTS

After we analyzed all of the “pre-software” talks, we took the completed master list of breakdowns and determined which could have been eliminated using version 2.3 of SlideShow Commander. We also proposed both short-term and long-term improvements to the software, which resulted in versions 2.4, 3.0, and x.0 of the software, shown in Table 4.

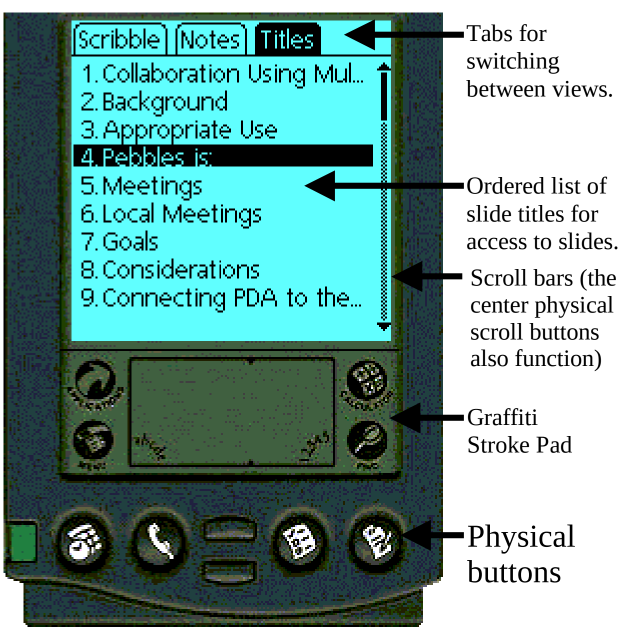


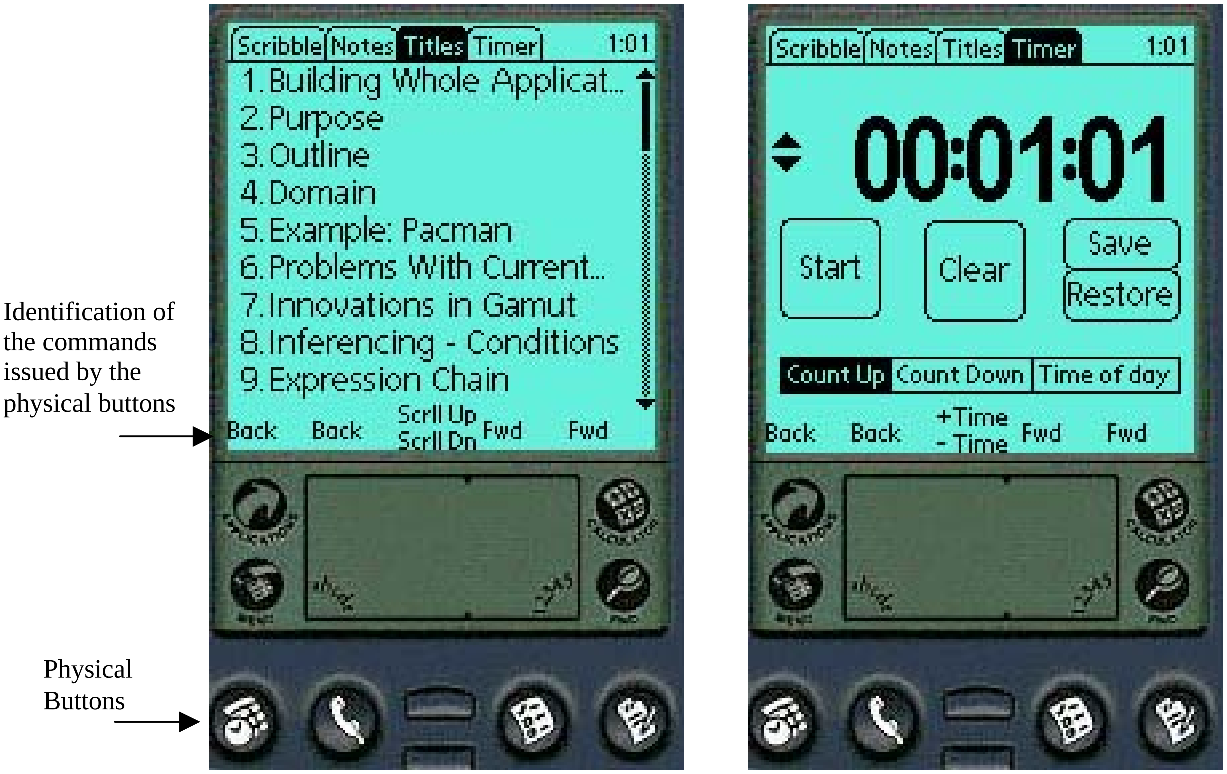

Figure 2. Titles View of SlideShow Commander v. 2.3

The initial version of the software we examined, 2.3, could have prevented certain breakdowns in the "pre-software talks." For example, we observed that presenters wasted time walking back and forth from behind the podium to the screen or audience, and returning to change slides (breakdown #1 on Table 3). Each time they returned, they had to reposition their hands on the keyboard, finding the correct 'forward' or 'backward' key. The original version of SlideShow Commander we studied, 2.3, would have prevented this breakdown by allowing the user to change slides away from the PC using the PDA. Further, the frequency of this breakdown made us realize that changing slides was a key goal for presenters, and should be supported by using physical buttons on the PDA, rather than requiring stylus gestures to advance. Thus, as an immediate fix to the software, we recommended that two of the hard buttons should be set as a default to 'next slide' and 'previous slide'. The developers took our suggestion and implemented this functionality in version 2.4 of the software.

Other breakdowns suggested important changes to the software that could be made in the short term. For example, many presenters lost track of the length of their talk, and had to ask verbally for the time remaining (breakdown #2 on Table 3). Not only was this an interruption in the flow of the presentation, but we knew from our own experience that it makes the audience aware of the time, and they start looking at their own watches, further decreasing the attention they are paying to the presenter. If a presenter knows that she only has 30 minutes in which to talk, she should be able to set a timer to count down from 30, so she knows when she is near the end of her allotted time. The authors of the software implemented this suggestion in version 3.0 of the software (see Figure 3).

Other breakdowns that we found suggested long term directions of development. For example, in every presenter's first attempt to play sound either locally through the PC or transfer it via NetMeeting software to a remote audience, the sound was not adjusted properly. One long-term development would be to allow the PDA to change the volume, and perhaps other controls, on the PC. The long term developments, combined under version 'x.0' in Table 4, have not yet been implemented by the software authors.

We used our measurements of the frequency and severity of breakdowns to argue for the importance of our design ideas (Table 5). Although the most frequent breakdowns are fixed by version 2.3 of the software, the most severe breakdowns (severity = 4) would not have been prevented by using either current or any foreseeable future version of SlideShow Commander. These breakdowns involved missing software on the laptop, wrong hardware for demonstrations, and remote audiences not being able to read the contents of slides using the cameras and NetMeeting. However, these breakdowns were rare – each being unique to only one presentation. We discovered that it is the small but common breakdowns in communication flow that would best be served by SlideShow Commander.




a. Titles View b. Timer View

Figure 3. Design Changes in SlideShow Commander 3.0. (a) Addition of user-defined hard button commands for basic slide navigation controls; (b) Addition of a timer to the application.

**Introduction of SlideShow Commander into presentations**

We performed a preliminary test to confirm our CI-based predictions. To do this, we analyzed two presentations in which presenters used SlideShow Commander to control a PowerPoint presentation. Both presentations were practice talks in which the presenters were practicing before peers and advisors for an upcoming formal presentation.

The first talk, T-1, was given by three presenters, two of whom used version 2.3 SlideShow Commander to point, access their notes, and advance slides (the third presenter was only responsible for the demo portion of the talk). The second talk, T2, was given by a single presenter who used version 2.4 of SlideShow Commander.

All of the presenters had prior experience with delivering PowerPoint presentations, and their self-rated PowerPoint presentation experience levels ranged from intermediate (5/10) to expert (10/10). None had used SlideShow Commander before. While all of the presenters had some prior experience using PDAs, their self-rated experience level on PDA use ranged from beginner (2/10) to expert (10/10). All of the users had a brief (20-30 minute) introduction to the software and the opportunity to practice with it several days before their presentation. We analyzed their practice talks using CI as well as questionnaires designed to assess the usability and satisfaction of the various features in the software. The questionnaires asked them to rate the ease of use of different features of

| Version | Description |
|---|---|
| 2.3 | Initial version of SlideShow Commander. |
| 2.4 | Addition of flexibility to assign hard buttons to various PowerPoint controls, such as ‘forward’ and ‘back’ slide. |
| 3.0 | Addition of a timer, which can display the time counting up or down, measure the duration of a presentation, and save times for later reference. |
| x.0 | Future long-term developments, such as providing a way to control other devices such as VCRs and Projectors in conjunction with the PC; and including a ‘task switcher’ to enable fluid transfer between one running application and another, and launching of additional applications. |

Table 4. Different versions of SlideShow Commander.

| Version | Average % of Talks | Count (Breakdowns Prevented) | Average Severity | Changes would save this amount of time in a single talk on average | Total time saved per talk on average |
|---|---|---|---|---|---|
| v. 2.3 | 20 | 178 | 1.5 | 54.7 sec | 54.7 sec |
| v. 2.4 | 0 | 0 | 0 | 0 sec | 54.7 sec |
| v. 3.0 | 21 | 19 | 1.25 | 13.5 sec | 1 min, 8.2 sec |
| v. x.0 | 24 | 15 | 1.7 | 4 min, 48 sec | 5 min, 56.2 sec |
| None | 13 | 17 | 2.1 | 2 min, 43 sec | 8 min, 39.2 sec |

Table 5. Improvement from SlideShow Commander. All data refers to the number of breakdowns observed in ‘pre-software’ talks that would not have occurred had the presenter used the features added in that particular version of the software. ‘None’ refers to breakdowns that occurred in the presentations but were outside the bounds of current or future versions of the software.

the software on a scale of 1 to 10 (where ‘10’ was ‘easy to use’). They were also asked to rate the importance of each feature to them on a scale of 1 to 10 (where ‘10’ was ‘very important’).

The only additional functionality that version 2.4, used in T-2, contained over version 2.3, used in T1, was the assignment of the “forward slide” and “backward slide” commands to the PDA physical buttons (see Figure 2). In version 2.3 of the software, these four buttons were assigned to special modifier keys such as <control>. None of the “pre-software” presenters that we observed, however, did anything more in PowerPoint than start the presentation, move forward and backward in slides, and switch to the thumbnail view to find a slide. While the ability to navigate through slides was included in version 2.3, we observed that the speakers in T-1 had some difficulty with performing the stylus gestures. This observation provided additional support for the most immediate fix requested — assigning the physical buttons to promote a quick, easy way to navigate among slides.

The talks in which the presenters used SlideShow Commander revealed absences of: 1) awkward leaning to reach the slide control; 2) physically being constrained closer to the PC away from the audience, and 3) navigational difficulties that had been observed with the non-SlideShow Commander talks. Presenters stood straight throughout the presentation, rather than repeatedly bending down to the laptop, as had been observed in the ‘pre-software’ talks. The presenters also chose to speak physically away from the PC. This provides some confirmation of our earlier observation that a PC tethers the presenters away from their desired speaking location. None of the presenters advanced to the wrong slide by pressing the wrong buttons, which occurred in some of the non-SlideShow Commander talks.

Presenters rated the ease of use and importance of the software reasonably well (mean ease of use = 6 and importance = 5, where 10 = ‘very easy/important’), and were most satisfied with the slide navigation feature of the software. The presenters’ rating of slide navigation as ‘easy to use’ varied from 7.5 (version 2.3) to 10 (version 2.4).

The different speaking styles of the presenters affected their ratings of the importance of the various features. For example, two presenters admitted that they were unfamiliar with their own slides, and relied heavily on the Notes feature of the software. Both of these presenters rated the importance of this feature as high (mean = 9). These results provide additional information that could be used in marketing the product and comparing it to its nearest competitors, which lack the ‘Notes’ feature (e.g. [3][7]).

## EVALUATION OF THE METHOD

CI analysis of the talks that did not use SlideShow Commander provided a successful means for predicting the usefulness of the software, as well as predicting what types of improvements should be made to the software. By using videotapes, we were able to make measurements that simple notetaking of observations (as recommended by Beyer and Holtzblatt) would not have made possible. We capitalized on the ability to return repeatedly to the videotape for confirmation of the identity, duration, and severity rating of the breakdowns. We used this data, as shown above, to more systematically classify each breakdown and its impact.

We originally learned the CI technique in a graduate-level HCI class. One of us relied primarily upon lectures and feedback from the instructor; the other depended more upon the Beyer and Holtzblatt text [1]. We judge that the text would have been sufficient by itself either in the initial apprehension of the technique, or in a later 'refresher,' although having an experienced HCI professional teach the course we took certainly made us both more confident in the accuracy of our models.

The most time-consuming part of the technique was drawing the models and timestamping the major events in the models. Initially, we transcribed the tapes so that we could have a written record of the events. Once we had modeled the first three talks, however, we found that generating a full transcript was extremely time-consuming and did not add significant value to the analysis. We thus switched to creating a partial transcript that only included potentially relevant material, such as discussions surrounding major breakdowns and useful quotes that supported our inferences or confirmed the severity of some of the breakdowns. Once we no longer used a full transcript, the models took approximately 2 - 3 times the length of the actual presentation to draw. We both drew them while initially viewing the tape, and then typically watched the tape a second time to garner quotes and details for the models. While drawing the models, we worked in parallel, drawing all four types of models for a given talk simultaneously, as recommended by [1].

With our data, we could also assess the usefulness of drawing each of the types of work models (Table 6). Certain models, however, were more difficult to draw than others, though they all became easier to draw with practice.

**Physical models** were the simplest models to create, requiring a view of the layout of the room and attention to overt indications of the smoothness, direction, and frequency of movement of the presenter (Figure 1a). Rough drafts of these rarely took more than ten minutes to draw. The physical models, however, were the least useful in terms of identifying surprising breakdowns that we had not seen before and would not have predicted. For example, the breakdown where a presenter tripped over a cord was obvious from simply watching the tapes (and hearing the chuckles from the audience), and thus did not need a physical model to make it evident. Because of its low cost in terms of time and attention, we recommend drawing these models, but not spending much time on the fine details, unless this has a specific use for the type of software or hardware being developed.

**Sequence models** were also fairly easy to create and draw, given the heavily sequential nature of presentations (Figure 1b). Although the steps were easy to identify, the modeling process still took a long period of time (up to an hour) in order to record interesting quotes and details of the sequential process. Our notes for the sequence models also became useful when returning to the tapes later in the modeling process while trying to find (or re-confirm) the presence of particular breakdowns. This model type might have been more difficult to draw in other domains in which more work occurs in parallel, such as studying secretarial work with its boundless interruptions. The sequence model was, with our domain, the type that provided us with the most detailed, defensible breakdowns.

**Flow models** were the most complex of the models we drew (Figure 1c). They required careful attention to the implicit ways in which information was transmitted between the audience, the presenter, screen, laptop, and other sources. Flow models were time-consuming to produce and required substantial effort to diagram the complex relationships and interactions of people, artifacts, and information. However, they revealed the immense complexity involved in giving a smooth presentation that we would not have predicted given our own informal beliefs about this domain. A presenter must keep track of the many directions information can come from at all times, make sure applications are launched on time and without difficulty, ensure that the correct slide is on the screen, and continue to speak at the right pace. Difficulties in any of these activities caused breakdowns that were often highly relevant to the software we analyzed.

For example, cameras and projectors being out of focus and sound being incorrectly adjusted, led directly to one of our long-term suggestions for the software – the ability to control the audio/visual aspects of presentations.

**Cultural models** were the hardest to draw conceptually (Figure 1d). Determining the desires and expectations of the presenters and audience required careful attention to overt behaviors, and the inference that these behaviors indicated certain mental states. We may have had difficulty because of the inability to interview presenters after their talks. Without this resource, we occasionally felt we were doing more inferring than observing, and were not absolutely confident of our cultural claims. Fortunately, often the presenters and audience members expressed their frustrations and desires during the course of the talk, probably because most of the talks were within a relatively small academic community that knew each other. Information from the cultural models would probably have been more detailed, with a greater number of desires and frustrations revealed, had we been able to interview the presenters after their talks.

## FUTURE WORK FOR SLIDESHOW COMMANDER

From our ‘pre-software’ contextual inquiries, we created a list of design suggestions inspired by our observations. By using our severity, duration, and frequency ratings, the value of further development was ascertained. We also were able to present to the developers the strong long-term potential of a more general PDA presentation tool that would allow for control of PC sound, projection focus, and VCR operations from the PDA.

Further, observing the interaction between audience and presenter has given us ideas for the functionality of an ‘audience mode’ for SlideShow Commander. This might be useful in helping the audience take notes, record references, and refer to slides during the question and answer periods of a presentation. Currently, multiple PDAs running SlideShow Commander can be connected to the same presentation, but there is no way for one PDA to block input from other PDAs, so audience members could disrupt presentations. In the future, we hope to help design an interface where the presenter's PDA can control the presentation while allowing the audience to occasionally share control in order to move to slides or point during discussions.

| Model | Average % of Talks | Count | Average Severity | Total Time Spent in Breakdowns Over All Talks (9.5 hours of talks were studied) |
|---|---|---|---|---|
| Physical | 33 | 169 | 1.2 | 5 min, 3 sec |
| Sequence | 17 | 24 | 2.0 | 7 min, 21 sec |
| Flow | 17 | 20 | 1.9 | 63 min, 37 sec |
| Cultural | 16 | 16 | 1.2 | 2 min, 38 sec |
| TOTAL | ------ | 229 | ------ | 78 min, 39 sec |

Table 6. Characteristics of the breakdowns initially detected by different types of models.

## FUTURE WORK FOR PERFORMING CI

The two most arduous aspects of performing CI are referencing the videotape thoroughly within the models (by timestamps) and drawing the models themselves neatly. Both are necessary to facilitate the process of confirming breakdowns and substantiating design ideas during later in-depth discussions of the data. Referencing the videotape by matching timestamps to key breakdowns and quotes was not difficult, but doing so carefully was painstaking, and slowed the speed at which we could properly document our models. We estimate that the initial sketches of the models would have been up to 30% faster if this step had been automatized and these timestamps could have been automatically inserted into our models. Second, drawing the models electronically was too time-consuming to do real-time (up to 2 hours per model, with four models needed for each presentation). As a result, we drew rough paper sketches while watching the videotapes, and then painstakingly remade each model neatly using a combination of Adobe Illustrator, Microsoft Word, and SmartDraw when formally presenting the results. To improve both referencing videotapes and drawing models, a tool digitizing and analyzing video for use in CI analysis would be an immense help. Ideally, this would enable the user to mark the ’key quotes’ or ’key moments’ that they need to reference in their models, and match these key points to the respective model. With enough planning, this tool could eventually create the models automatically and include accurate references to the data.

## CONCLUSIONS

Using CI, we analyzed various types of presentations and observed their common features and breakdowns. We predicted that some of these breakdowns would be remedied with our software, and introduced the software into two talks to test our predictions. These talks supported our predictions, in that SlideShow Commander talks revealed an absence of the most frequent breakdowns. Presenters could advance slides, point, and look at Notes with a minimum of physical movement. Our adaptation of the CI technique worked well using videotaped analysis, particularly for models where the inner thoughts and beliefs of he user could be inferred from their outward behavior, such as the flow models. Capitalizing on video to confirm and measure breakdowns is a useful addition that others should follow when trying to provide more quantitative support for their design decisions.

## ACKNOWLEDGEMENTS

We would like to thank the Pebbles group, in particular Benjamin Bostwick and Rob Miller, for their assistance in enabling us to study their software, and their willingness to implement our design suggestions. We would like to thank all of the presenters we studied, especially Ralph Gross and the members of the Interactive System Labs. We would also like to thank Catherine Copetas at Carnegie Mellon University for providing access to her library of videotapes. Finally, we would like to thank Daniel Yocum for his support throughout this endeavor.

## REFERENCES


1. Beyer, H. and Holtzblatt, K. *Contextual Design: Defining Customer-Centered Systems.* Morgan Kaufmann Publishers, Inc., San Francisco CA , 1998.
2. Holtzblatt K. and Jones, S. Contextual Inquiry: Principles and Practice. *Participatory Design: Principles and Practice.* Lawrence Earlbaum, New York NY, 1993.
3. Logitech TrackMan Live! Description at http://www.logitech.com/us/trackballs/tb22_100.html
4. Mackay, W.E. and Tatar, D.G. Introduction to the Special Issue on Video as a Research and Design Tool. *ACM SIGCHI Bulletin* 21(2), 48-50.
5. Mablekos, C. *Presentations that Work.* Institute of Electrical and Electronics Engineers, New York NY, 1991.
6. Mandel, S. *Technical Presentation Skills.* Crisp Publications, Los Altos CA, 1998
7. MindPath PocketPoint Description at http://www.mindpath.com/Products/pocketpointer.
8. Nelson, L., Ichimura, S., and Pedersen, E.R. Palette: A Paper Interface for Giving Presentations. *CHI'99,* 354-361.
9. Nunamaker, J.F., Dennis, A.R., Valacich, J.S. et. al. Electronic Meeting Systems to Support Group Work. *Commun ACM 34(7)*, 40-61.
10. Pebbles SlideShow Commander. Available at http://www.cs.cmu.edu/~pebbles/
11. Wixon,D. and Ramsey, J. *Field Methods Casebook for Software Design.* Wiley Computer Pub., New York NY, 1996