An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents

RICHARD CHIMERA and BEN SHNEIDERMAN
University of Maryland

Three different interfaces were used to browse a large (1296 items) table of contents. A fully expanded stable interface, expand/contract interface, and multi-pane interface were studied in a between-groups experiment with 41 novice participants. Nine timed fact retrieval tasks were performed, each task is analyzed and discussed separately. We found that both the expand/contract and multi-pane interfaces produced significantly faster times than the stable interface for many tasks using this large hierarchy; other advantages of the expand/contract and multi-pane interfaces over the stable interface are discussed. The animation characteristics of the expand/contract interface appear to play a major role. Refinements to the multi-pane and expand/contract interfaces are suggested. A predictive model for measuring navigation effort of each interface is presented.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces--*screen design, evaluation/methodology*; H.1.2 [**Models and Principles**]: User/Machine Systems--*human information processing*

General Terms**:** Human Factors, Experimentation, Documentation

Keywords and Phrases**:** browsing, hierarchies, table of contents, user interfaces

Authors' Addresses: Richard Chimera, Human-Computer Interaction Laboratory, A.V. Williams Building, University of Maryland, College Park, MD 20742-3255. carm@cs.umd.edu
Ben Shneiderman, Computer Science Department, Human-Computer Interaction Laboratory, Institute for Systems Research, University of Maryland, College Park, MD 20742-3255. ben@cs.umd.edu

1 Introduction

Although only a few studies have shown the superiority of online texts when compared to paper documents, the increasing requirement for electronic access to voluminous data dictates improved methods for browsing online information. Several studies have shown that a hierarchical representation with associated browser is an effective approach [1, 19]. Recent studies have shown that browsing through a table of contents is a preferred method over more analytical methods such as query formulation [2, 4, 14]. Benefits of online information include geographically distributed high speed access to data; faster, more complete updates to information [20]; the capability for string searching; and inclusion of animation [16], sound, and video.

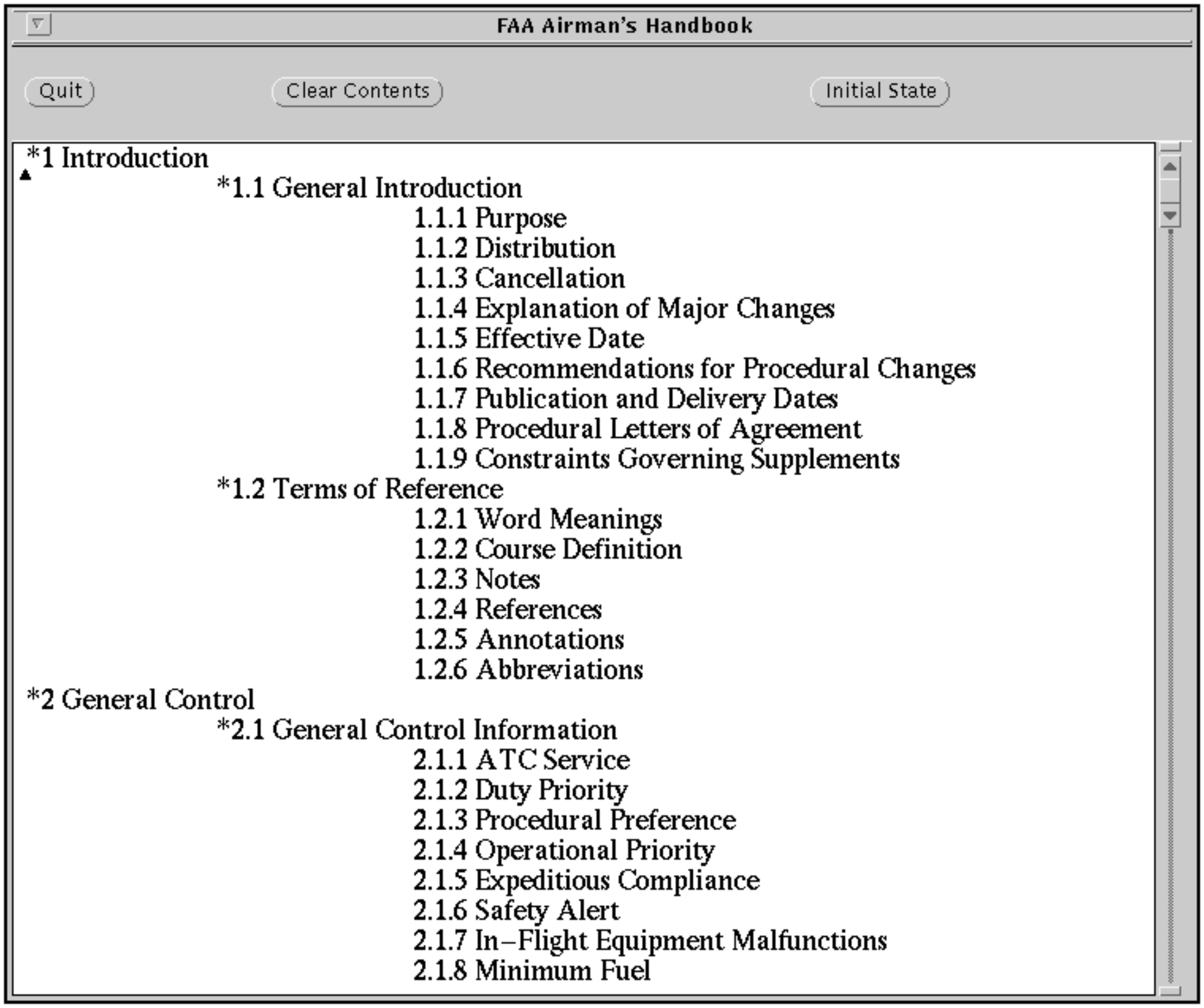

Figure 1a. Stable interface in initial state.

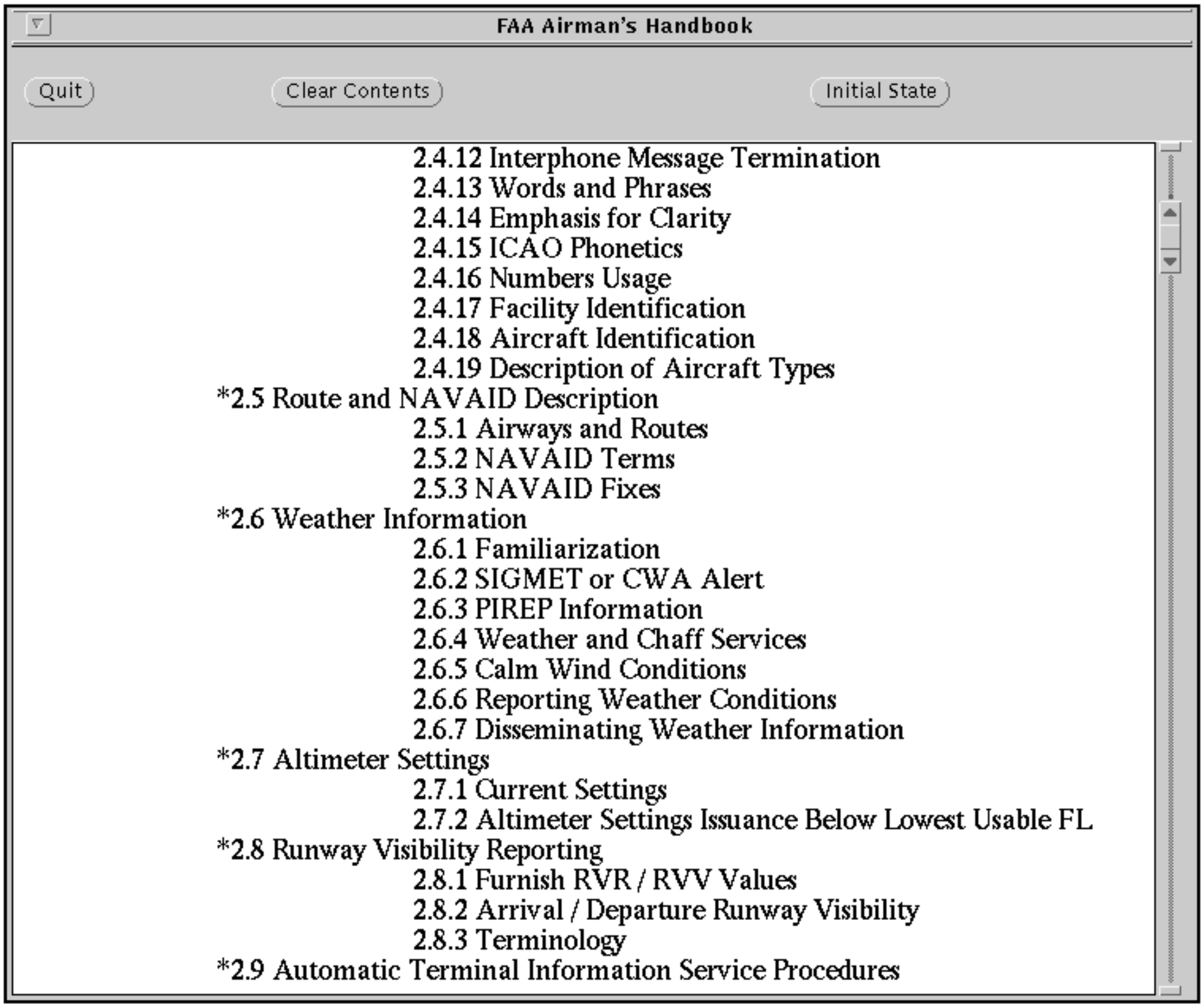

Figure 1b. Stable interface viewing section 2.6. Scrolling is the only navigation technique.

There are three generally used interface methods for browsing online, hierarchical information. One type of interface which we call *stable* (Figures 1a-1b) uses a fully expanded, stable screen layout, much like a table of contents for a printed book. These interfaces are quite widely used for viewing online hierarchies [11]. Another interface, *expand/contract* (Figures 2a-2c), is employed by many outline programs available for PCs (e.g., PC Outline, GrandView, etc.). A third interface, *multi-pane* (Figures 3a-3c), usually consists of a predetermined number of separate panes, each pane containing contents for a separate level of the hierarchy (e.g., Smalltalk browser). There are variations to what the behavior can be when the number of levels of the hierarchy is more than the number of panes.

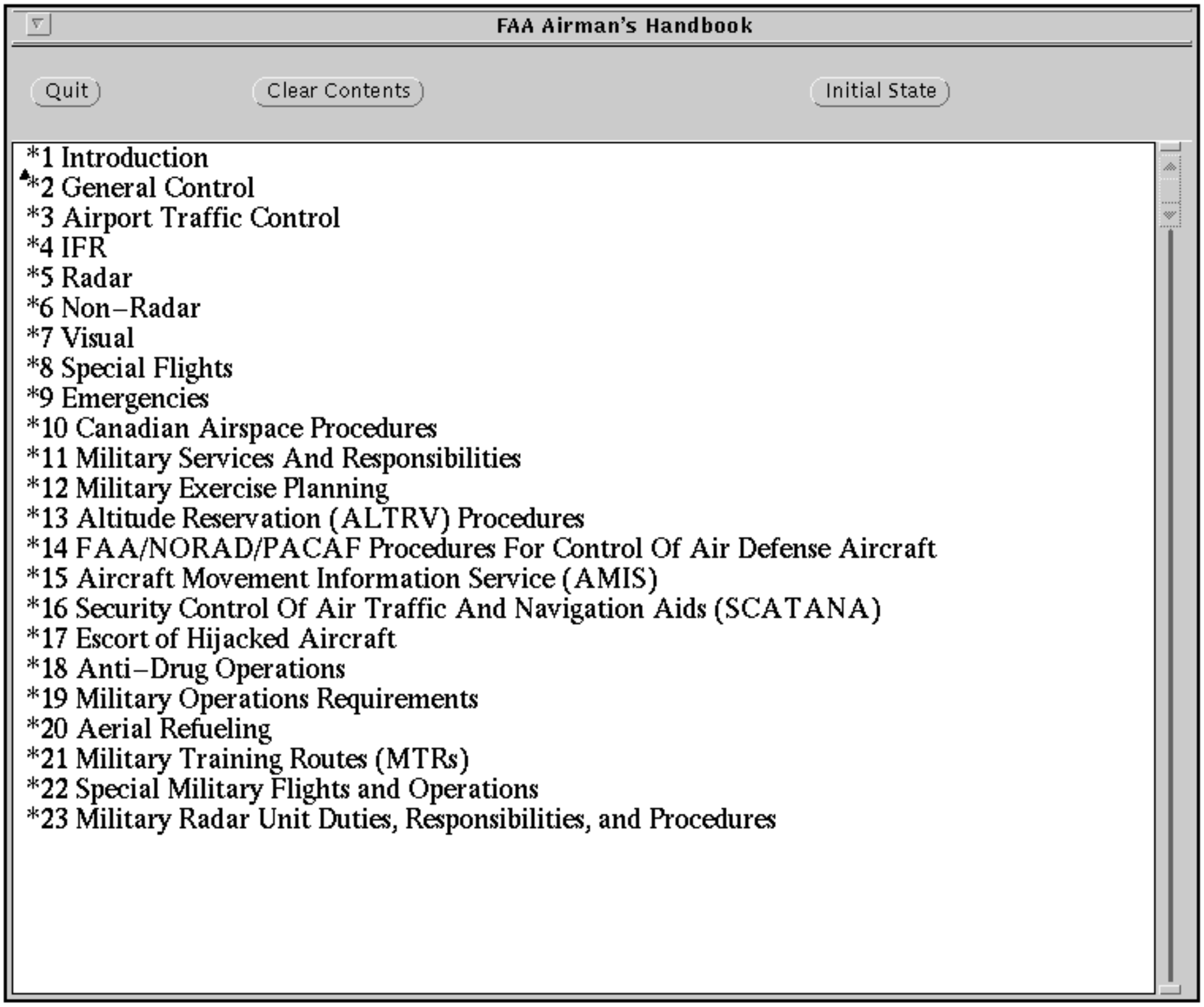

Figure 2a. Expand/Contract interface in initial state. Only highest level (chapter) items are visible.

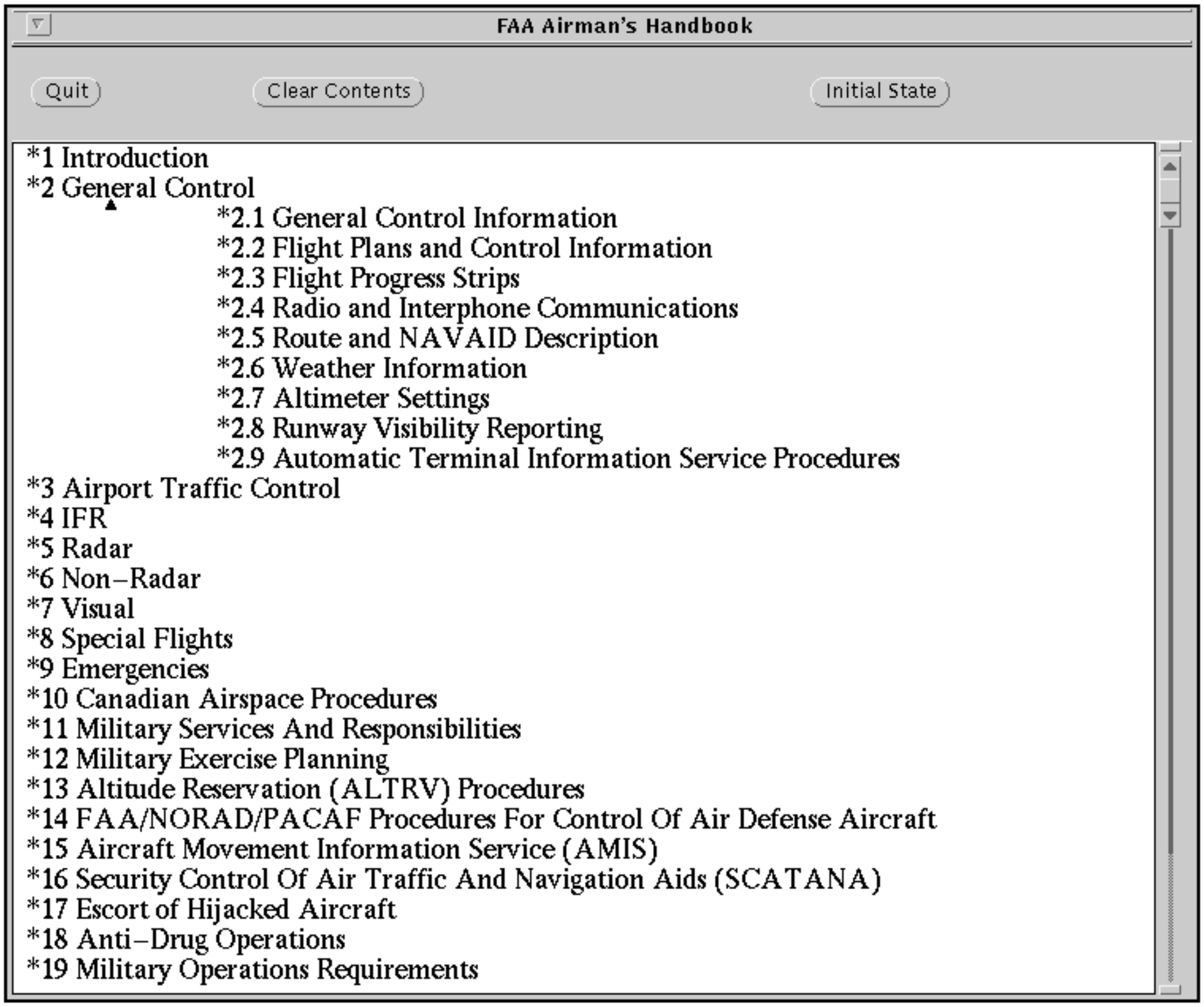

Figure 2b. Expand/Contract interface viewing chapter 2, produced by clicking the mouse on the title. Only items in the immediately next lower level (sections) appear via line-at-a-time animation.

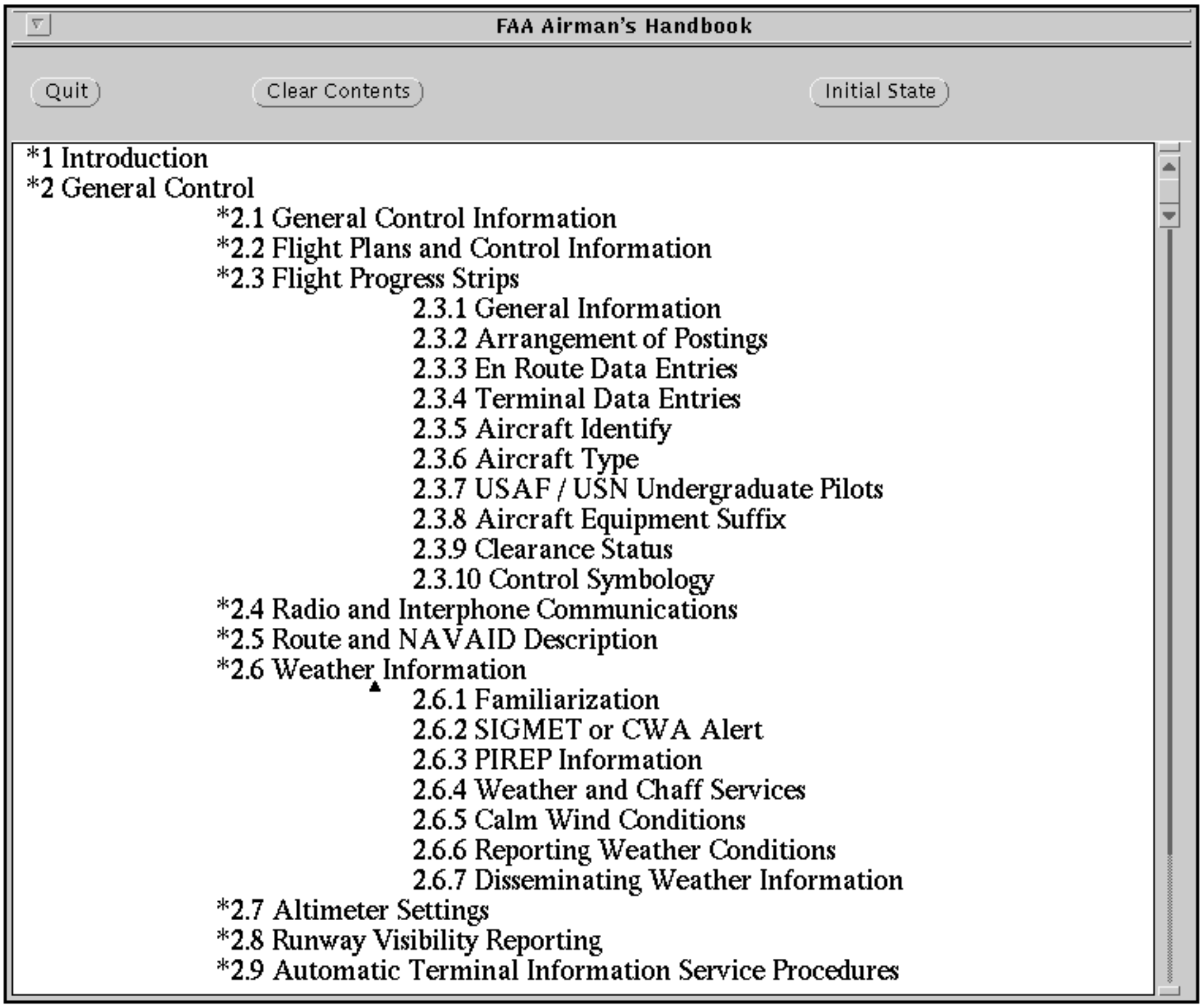

Figure 2c. Expand/Contract interface viewing section 2.6 and section 2.3.

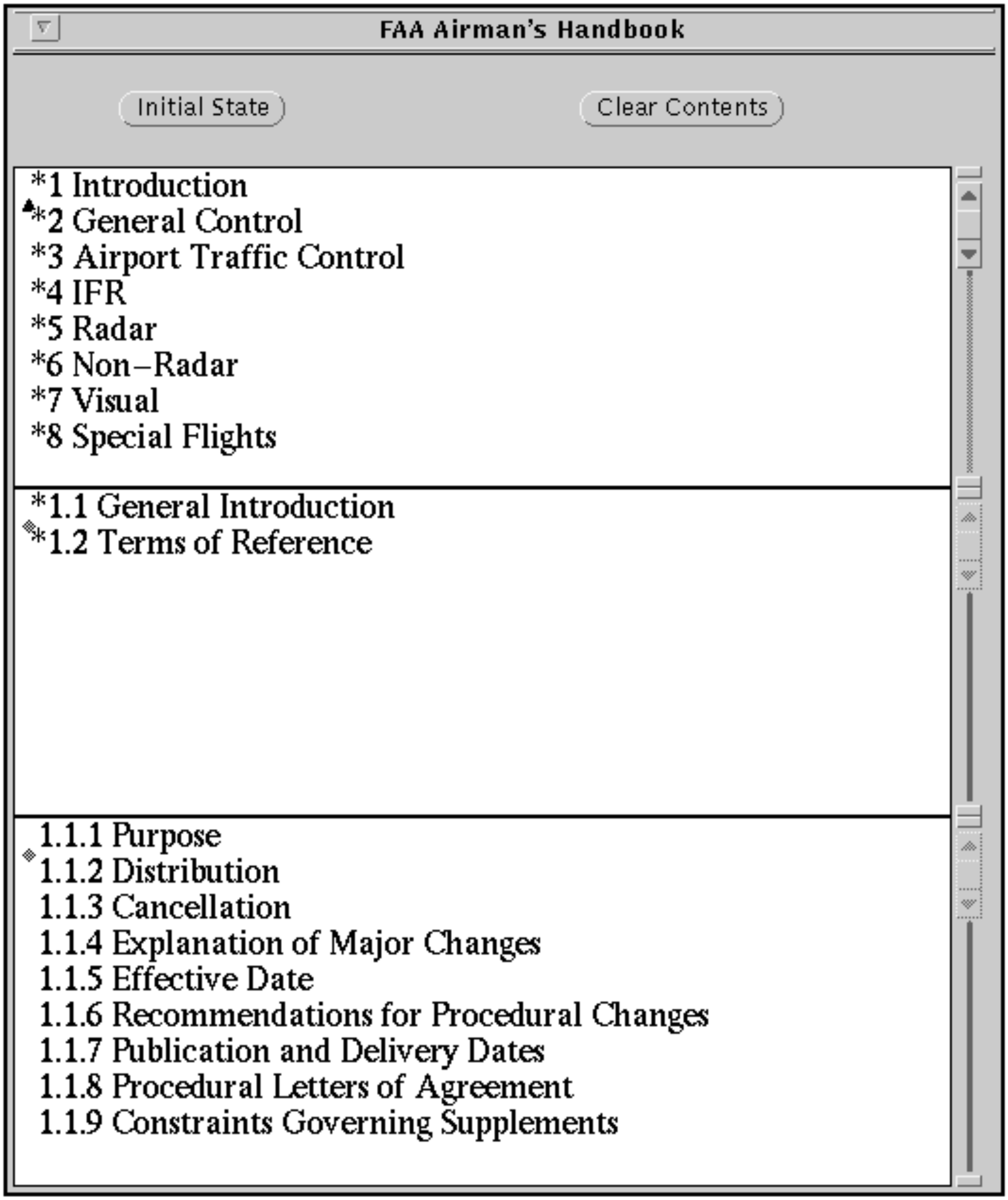

Figure 3a. Multi-pane interface in initial state. Chapter items appear in the top pane, section items in the middle pane, and subsection items appear in the bottom pane.

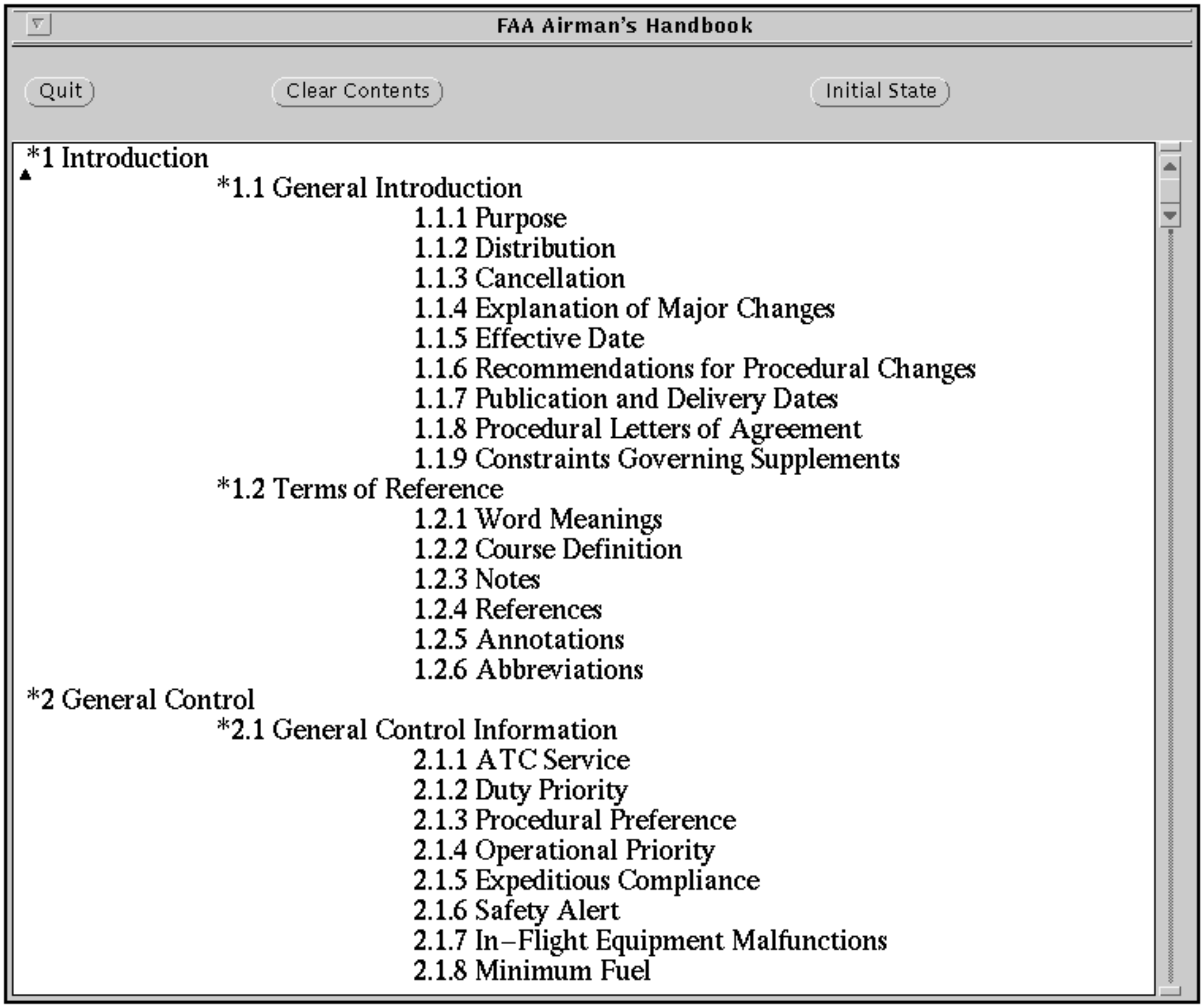

Figure 3b. Multi-pane interface viewing chapter 2, produced by clicking the mouse on its title. Notice the bottom pane automatically displays the subsections of section 2.1 (the subordinates of the middle pane's firstmost item).

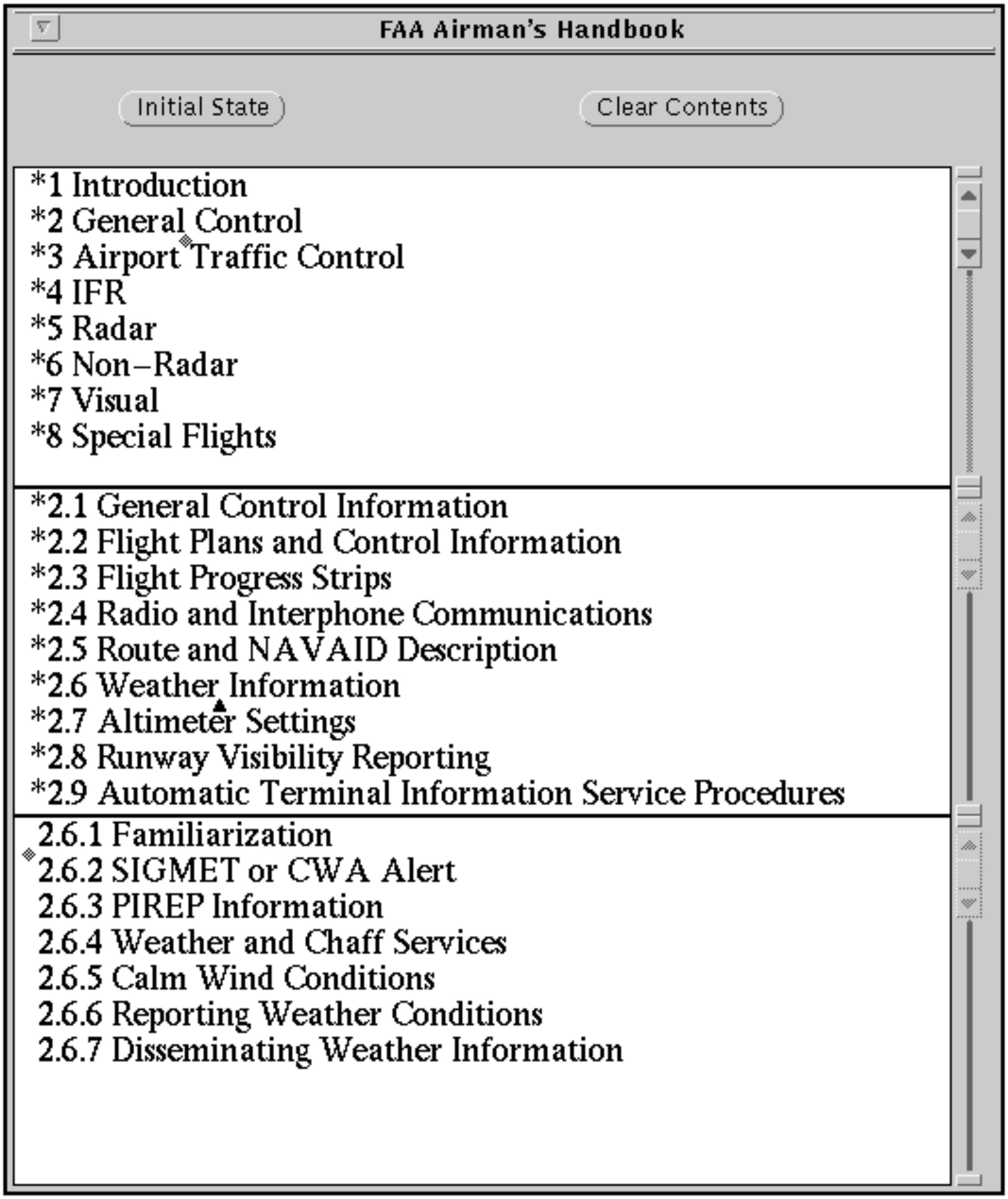

Figure 3c. Multi-pane interface viewing section 2.6, produced by clicking the mouse on its title in the middle pane.

The major drawback to a stable interface is that users have difficulty in perceiving the global hierarchical breakdown of the text. Because section and subsection titles are permanently displayed in-line, many screens/pages would have to be scrolled to show enough chapter titles to comprehend the major divisions of the text. The expand/contract and multi-pane interfaces attempt to overcome the hierarchy breakdown problem by displaying the high level information contiguously and giving users the choice of viewing specific section and subsection levels on demand. The motivation behind these two interfaces is analogous to the fisheye view supported by Furnas in which the fisheye strategy "is to provide a balance of local detail and global context."[11]

There are inherent problems when using online textual documents when compared to the customary leafing through a paper document. When using an online replica of a printed document, e.g., viewing the document through the stable interface, users must perform a considerable amount of scrolling and can get lost in a large document. Also, since the online version usually displays fewer items per screen as compared to a printed page, users suffer from limited conveyance of informational context [17]. The resultant disorientation from viewing a large document can be caused by conceptual confusion concerning the logical structure of the document [4]. The expand/contract interface preserves more of the logical structure and context than the stable interface since the chapter titles preceding the one chosen for expansion often are visible on the screen. The multi-pane interface overcomes the conceptual deficiency by constantly displaying the high level chapter and section information.

Another aspect of disorientation is actually getting lost while navigating through the document [4]. The only navigational aid offered in the stable interface is the scroll bar, although the inclusion of chapter and section numbers can help. When viewing a section that is longer than one screen, users can quickly forget what part of the document is being viewed. In contrast, by directly interacting with a TOC item to be viewed, the multi-pane and expand/contract interfaces provide an additional reinforcement and anchor point (users are cognizant of the TOC item and hopefully its location because they clicked on it) which contributes to users' awareness of their location within the document.

The SuperBook text browser [8, 15] incorporates a similar type of interface format to our expand/contract interface. The SuperBook version of the interface is called "the fisheye viewer" and it allows users to browse lower level information at the same time as higher level information is displayed. SuperBook incorporates a table of contents window that allows for the expansion and contraction of the chapter and section levels. An experiment with SuperBook showed that the online browser combined with text search capabilities proved to be better in search tasks when compared to conventional paper documentation [9].

The studies reviewed showed a performance improvement for a fisheye view when compared to either a paper document or to a stable interface. This exploratory study builds upon the previous studies by validating that both the expand/contract and multi-pane interfaces (two different implementations of a fisheye viewer) are superior to the stable interface. We also attempt to reveal the performance differences between these two interfaces for various tasks. The major difference between the two interfaces is that expand/contract could have large expansions such that other high level items are pushed off the screen; but in the multi-pane interface, an item's siblings are always kept together contiguously on the screen when its subordinates are viewed.

Another goal of this exploratory evaluation was to create and validate a predictive model of user effort for the three interface designs. User effort is defined in terms of the average number of mouse clicks needed to make a page of text visible. The predictive model is for the compound task of locating a topic of interest in the TOC and getting that page of the document onto the screen; the last action is not part of the experiment that we discuss in this paper but makes the predictive model more complete.

2 Experiment

2.1 Introduction

This exploratory experiment attempted to determine which of the three browser interface designs would provide the most efficient access to data in a hierarchy. We chose a table of contents (TOC) as the source document for browsing because of the increasing need to find online information from electronic books, journals, and manuals. We felt providing only manual browsing without additional functionality such as text searching would shed light on how one of the most basic interface characteristics affects locating information online. A between-groups design was employed to compare browsing online tables of contents with three interface styles: stable, expand/contract, and multi-pane; each of three groups used only one of the different interfaces for browsing the TOC. 41 participants were given nine questions to answer using a large TOC that contained 1296 items[1] . Since the differences would become more clear as the size of the TOC grows, we compared the interfaces with a large TOC. We define the size of a TOC relative to the fully expanded view of the stable interface. In a small TOC all items fit on one display screen. For a medium TOC between two and fifteen screens are needed to display all items. A large TOC requires more than fifteen screens to display all items.

A medium TOC of 174 items requiring seven screens to view was used as training, and a large TOC of 1296 items requiring 48 screens to view was used as the TOC size on which to collect interaction data for this experiment. A small TOC (of less than 28 items) was excluded because we felt the results would have been quite straightforward (an advantage for the stable interface because all information would be visible on the screen requiring no user activity other than reading). TOCs come in many shapes as well as many sizes. To reduce variance in performance due to TOC structure, we chose TOCs that were "well formed" down to the third level. Well formed means that each of the two upper levels' items always had subordinates, and no item at the third level had any subordinates. We acknowledge that the majority of TOCs in

[1] The TOC can be obtained via anonymous ftp on the machine bedrock.cs.umd.edu in the file pub/toc/large.text.

existence are not well-formed, but for the purposes of our empirical study, we felt we needed to be rigid in our design of the TOC to better account for performance differences attributable to interface design. Redesign of the interfaces, especially the multi-pane interface, may be necessary to optimize performance with non-well formed TOCs.

2.2 Description of Interfaces

All interfaces were built using the Developer's Guide user interface development package in the OpenWindows environment on a Sun Microsystems Sparcstation 1+ workstation with a 17 inch color monitor and optical three-button mouse. An 18 point TimesRoman font was used, and the height of the windows used for all three interfaces was kept constant at 6.375 inches (16.2 cm) for the text area. 28 lines of text could be displayed on one screen in the stable and expand/contract interfaces, only 27 lines of text (8, 9, and 10 for respective panes) could be displayed in the multi-pane interface because some height was consumed by the horizontal lines separating the three text panes. The widths were kept at the minimum needed to display the longest line of text on one line, 6 inches (15.2 cm) for the multi-pane interface and almost 9 inches (22.9 cm) for the stable and expand/contract interfaces since subordinate items were indented. The text was displayed as black on a white background.

Scrolling methods were the same for all interfaces, providing the ability to scroll forward or backward by one line at a time, one screen at a time, go directly to the end or beginning of the TOC, or go to any part of the TOC via gross scrolling by sliding an elevator in the scroll bar. All interfaces displayed only the TOC itself and no pages of actual online text. This was done so that we could more readily attribute differences in user performance to interface features rather than the possibility of participants reading the full text of the document in varying degrees. Again we justify this simplification for more accurate attributions of cause and effect directly to interface features rather than to complex design and interaction issues.

The implementation of the three interfaces was planned such that none of the interfaces used interface features *unnecessarily* unique to that interface. The only unique features that distinguished the three interfaces were what happened when participants clicked the mouse in a text pane and where and how new or different information was displayed. We tried to make each interface have the most "obvious" interaction and display characteristics for that interface, but nothing that made it strikingly different from the other interfaces. For example, it would have made sense for the multi-pane interface to have highlighted the entire line of the TOC just clicked so that it was readily apparent which item had been opened up in the pane below. However, this would have made that interface's display "too" different from the other two in the sense that the highlighted region may be harder to read and attract attention. Such confounding factors would be difficult to account for and the highlighting behavior is not necessary for the multi-pane interface to operate in a natural and straightforward manner.

The stable interface presents a view of the data just like that in a book, every item's subordinates were indented and permanently listed directly underneath that item (Figures 1a-1b)[2] . The only interaction and navigation available is to scroll through the TOC.

The expand/contract interface initially presents only chapter titles (Figure 2a) and users must click the mouse on TOC items to have them expand in place (Figures 2b-2c). If *all* TOC items were expanded, this interface's view would be exactly that of the stable interface. Clicking on an expanded TOC item would contract all levels of its subordinates. The animation characteristics of the expansion and contraction are critical to the ease of use of this interface. We chose to paint (erase) rapidly one item at a time so that smooth animation made it clear where the new information appeared. We conjecture that if an expansion (contraction) was performed with an entire screen repaint, it would require complete reorientation to identify where the new information appeared. It took approximately one second to expand an item with 16 subordinates, approximately 1.25 seconds to contract an item with 16 subordinates.

The multi-pane interface used three separate text panes, one pane to view only chapter information, one to view only section information, and the third to view only subsection information (Figures 3a-3c). When a TOC item is clicked with the mouse, the text pane directly below it would clear the old information and display the subordinates of the item clicked; it operates similarly to the Smalltalk object browser. If a chapter pane receives a mouse click, not only does the section pane display that chapter's sections, but the subsection pane would change its information to display the subordinates of the first item in the section pane (Figure 3b). An item's subordinates are displayed very quickly, all the text in a section and/or subsection pane of 16 items appears on the screen in under 0.2 seconds.

2.3 Participants

41 undergraduates from the Psychology Department freshman subject pool received course credit for participating in the experiment. We did not require participants to have computer experience, over half considered themselves novice users of either a PC or Macintosh, three claimed intermediate experience, none were expert users, and the rest had no computer experience. No participants had used the workstation specific to our experiment. The most dominant interaction feature of our experiment, the mouse, had been used by 33 of the 41 participants. Since the subject matter for our medium and large TOCs were in the areas of abnormal psychology and air traffic, respectively, we also asked about experience level in these areas and found none were experts in either field. For air traffic, one reported intermediate knowledge, two reported themselves at the novice level, and the rest had no knowledge of the field. For abnormal psychology, two had intermediate knowledge, nine reported novice level knowledge, and the rest had no knowledge of that field.

[2] A video has been produced [5] to show the interfaces in action, which gives the viewer a much better understanding than text descriptions and snapshots of each interface's interaction characteristics.

2.4 Materials

The independent variable was type of interface format: stable, expand/contract, or multi-pane. The dependent variable was the time to find the correct response to fact retrieval questions about the TOC contents via browsing.

Three online TOCs were used in this experiment. The first one, used for the initial training phase of the experiment, was the TOC from the first edition of Designing the User Interface [18]. It had 118 items, similar in length to the medium TOC size of 174 items. The medium TOC, used for practice tasks similar to those for which experimental data was collected, was taken from the textbook Abnormal Psychology: Experiences, Origins, and Interventions [12]. For the large TOC, we put together two TOCs from the Air Traffic Control Handbook and the Military Handbook [10]. The introduction chapters were removed from the latter TOC, in our opinion there was no discontinuity in context or flow resulting from appending the two TOCs together; the resulting TOC had a total of 1296 items.

We created a set of tasks that we felt were representative of simple and complex browsing tasks. Each of the following nine timed tasks was typed on a separate page:

1. What is the title of section 1.1?
2. What is the title of subsection 8.2.3?
3. What chapter, section, or subsection number, if any, has the title, "Radar Departure"?
4. What chapter, section, or subsection number, if any, has the title, "Aerial Refueling Fire Emergencies"?
5. Name the table of contents item that deals with all of the topics of approach lights, runway lights, and taxiway lights.
6. What is the number of the chapter that appears first in this document for the chapter titles: "Anti-Drug Operations" "Military Training Routes"
7. In chapter 2, General Control, what is the number of the section that appears first: "Weather Information" "Radio and Interphone Information"
8. In section 2.1, General Control Information, what is the number of the subsection that appears first: "Traffic Advisories" "Operational Requests"
9. Does this table of contents contain more information about: radar emergencies

Tasks 1 and 2 were simple navigation tasks, like the real-world task of being given a reference to a particular numbered item in a TOC. Task 1 was designed to be very easy to give participants initial confidence; task two was designed to be of average difficulty, its location was about one third of the way through the TOC. Tasks 3 and 4 were text search oriented, typical of what one would do to determine if a certain topic is covered in a TOC/book. The title given in task four did not exist in the TOC (a nonexistent title task was also given in the training tasks), revealing how well an interface facilitates the discovery that a topic wasn't covered in the TOC. Task 5 was a high cognitive load exploration task to determine if and where there was topic coverage. Tasks 6, 7, and 8 had some navigation aspects but mostly were designed to be sibling comparison tasks, each task focused on a different level of the TOC. The interfaces treated items' siblings and children differently, these tasks helped to pinpoint where and how these differences revealed themselves.

Task 9 was another form of a high cognitive load exploration and topic coverage task similar to task five, but concentrated on the higher levels of the TOC.

In creating these tasks we wanted both short, low cognitive load tasks and long, high cognitive load tasks to be performed to represent the large variety of browsing tasks. We offer a working definition of browsing in a computer environment to be interactively navigating, scanning, and interpreting information to determine if it is useful or interesting.

2.5 Procedures

Each session lasted approximately one hour and all participants worked on the same physical workstation under similar operating conditions. A typical session consisted of the phases:

1. Introduction and Training: Participants read a description of the experiment and signed an informed consent form. For training consistency, a document was used as a guide to explain the mouse and scrolling features as well as any specific features of the interface participants were using. For example, the training document for the expand/contract interface explains how to expand information and contract it again by positioning and clicking the mouse pointer on the line of the topic of which there is interest in its subordinates. Each training document also explained how to scroll the text continuously, line-at-a-time, page-at-a-time, and jump to any part of the TOC.
2. Practice tasks: To be sure the participants understood all the required functions of the interface, eight practice tasks were given that tested mastery of the interface training issues. Participants were encouraged to ask any questions they may have about how the interface worked. Then a complete set of practice tasks was administered with the medium TOC, its tasks were similar to the timed tasks participants would perform in the experiment. During this phase the tasks were timed and the experimenter could not explain anything the participants did not understand, emulating the conditions in the data collection phase.
3. Timed tasks: The participants were required to answer nine questions about information in the large TOC. Participants read and understood the question, they clicked an "Initial State" button to display the TOC in its initial state, and started work towards answering the question. The administrator started a digital stopwatch when the initial state button was clicked. Participants gave answers in the form of verbal responses. After each correct response the administrator stopped the stopwatch and recorded the time, the screen was cleared by participants clicking on a "Clear" button. If an incorrect response was made, participants were informed that the response was incorrect and asked to continue with the same question. The maximum time for each task was four minutes; if participants did not finish a task within this time limit, the time recorded was the maximum time limit.

Timing

Some factors involved in online search strategy include the cognitive abilities of the participant when selecting methods for browsing and the reading speed at which users read or scan the material [13]. To try to minimize the difference in reading and comprehension speed, the timing of each browsing task was begun once the participants fully understood the question and signified that they were beginning the search task by clicking on an “Initial State” button to display the TOC in the window. The TOC was erased from the display between tasks and not displayed again until the “Initial State” button was pressed; each task started from the same initial state. This method reduced variability of completion times as well as disorientation effects some participants might have suffered due to the screen being left at the completion position of the previous task; “extra” viewing of the TOC between tasks was also eliminated. It was decided that a digital stopwatch controlled by the experiment administrator would provide the most accurate timing to the correct response. Having participants click on start and stop buttons would increase their cognitive load, increase completion time variance due to remembering and successfully clicking on the stop button, and make it more difficult to continue timing for incorrect responses. Administrator error on the order of a fraction of a second in stopping the stopwatch seemed a minimal price considering that completion times are in the tens and hundreds of seconds.

2.6 Hypothesis

Our hypothesis was that the multi-pane interface would speed performance with a large TOC for several reasons. First, it provided a useful fisheye view of the information which we have come to call a MIMI view (Most Important Multiple Items) that keeps high level items always visible and together. This is in contrast to the expand/contract interface where the higher level items get separated from their siblings during an expansion and may even be pushed off the screen. Also, the multi-pane doesn’t require any undoing of operations to retain a global view, but the expand/contract requires contracting what is expanded to regain the global view. We predicted the stable interface would generate the worst performance for a large TOC because of the excessive amount of scrolling necessary to access later portions of the TOC, and the complete loss of a compact view of the high level items because low level items permanently separate them.

2.7 Model of Interaction Effort

A simple predictive model, in the spirit of the keystroke-level model [3], was created for estimating the relative effort of bringing into view a specified page within a specified subsection (lowest level of TOC) of an online manual. The model estimates users' effort by counting the number of mouse clicks needed to scroll screen-by-screen within a window and/or to select chapters, sections, subsections, and ultimately selecting to view a page of documentation (note the last mouse click/activity is not present in our experiment). The model assumes users would always scroll by one screen at a time until the target item was visible; models incorporating gross scrolling would involve very sophisticated probabilistic formulas. This "clicks model" has parameters for the number of text lines in the window, number of levels (depth) in the TOC, and uniform fanout (breadth) at each level. The model does not predict speed, overall comprehension, nor capacity to compare meaningful content in the TOC, these issues and their relationship with the three interfaces are discussed elsewhere in this paper. However, overall user effort would seem proportional to the number of clicks predicted by this model.

For a 30-line window and three levels in the hierarchy, as in our experiment, the clicks model yields average effort levels to view a page for the three interfaces (Table 1).

Fanout	Number of items	Stable	Expand/ Contract	Multi-pane
3	39	1.30	3.00	2.80
5	155	3.15	3.00	2.80
10	1110	19.07	3.00	2.80
15	3615	60.82	3.20	3.80
20	8420	140.90	3.54	4.30
30	27930	466.07	4.03	5.80

Table 1. Predictive model for the average number of mouse clicks needed to scroll and select TOC items.

The size of the TOC grows exponentially with the fanout (breadth) at each level--the size of the TOC (number of items) is equal to $x + x^2 + x^3 + \ldots + x^n$ for an n-level TOC, where x is the fanout at every level. To arrive at this value, consider the number of items in a level starting from the highest level. There are x items in the chapter level. Considering each item has x subordinates we now have logical groups consisting of the item and its subordinate, $1 + x$ items, occurring x times for a total of $x(1 + x)$ items. Going down one more level to the subsection level, each of the $x(1 + x)$ items have themselves and x subordinates, or $x(1 + x(1 + x))$ items, or $x + x^2 + x^3$ items for a 3-level TOC. The table shows how the effort involved in the stable interface grows linearly with the size of the TOC (the size of the TOC itself growing exponentially). It also shows how both the expand/contract and multi-pane interfaces increase very modestly the amount of effort needed as TOC size grows. The point at which the expand/contract interface outperforms the multi-pane interface is *right after* the point at which the fanout is equal to the number of items that are visible in one pane in the multi-pane interface. The maximum for this number (some space may be taken by the dividing space between panes) is arrived at by dividing the total number of lines in the window by the number of panes (for this table, 30÷3=10).

3 Results

We ran repeated measures analysis of variance using MANOVAs and the Wilks' λ statistic to determine if there were statistically significant differences in speed for tasks, interfaces, and interactions between task and interface. Task effect ($F = 121.71$), interface effect ($F = 24.49$), and task by interface interaction effect ($F = 18.04$) were all significant at the $p < .0001$ level. For corroboration, we ran univariate ANOVAs using the adjusted Greenhouse-Geisser F-values, all of which reported similar results significant at the $p < .0001$ level.

Based on these findings, we then ran separate one-way ANOVAs to analyze each timed task for differences in completion times across the three interfaces. Tukey's post-hoc analysis was used to determine specifically which interface(s) had the statistically significant advantage. The results are presented both in Table 2 and Figure 4. Data for Task 1 showed that both the multi-pane and stable interfaces had significantly faster mean times than the expand/contract interface at the $p < .01$ level. The data for Tasks 2, 3, 6, and 7 showed that both the multi-pane and expand/contract interfaces had significantly faster mean times than the stable interface at the $p < .01$ level. The data for Task 4 showed that the expand/contract interface had a significantly faster mean time than the stable interface at the $p < .05$ level. The data for Task 9 showed that the multi-pane interface had a significantly faster mean time than the expand/contract interface at the $p < .05$ level.

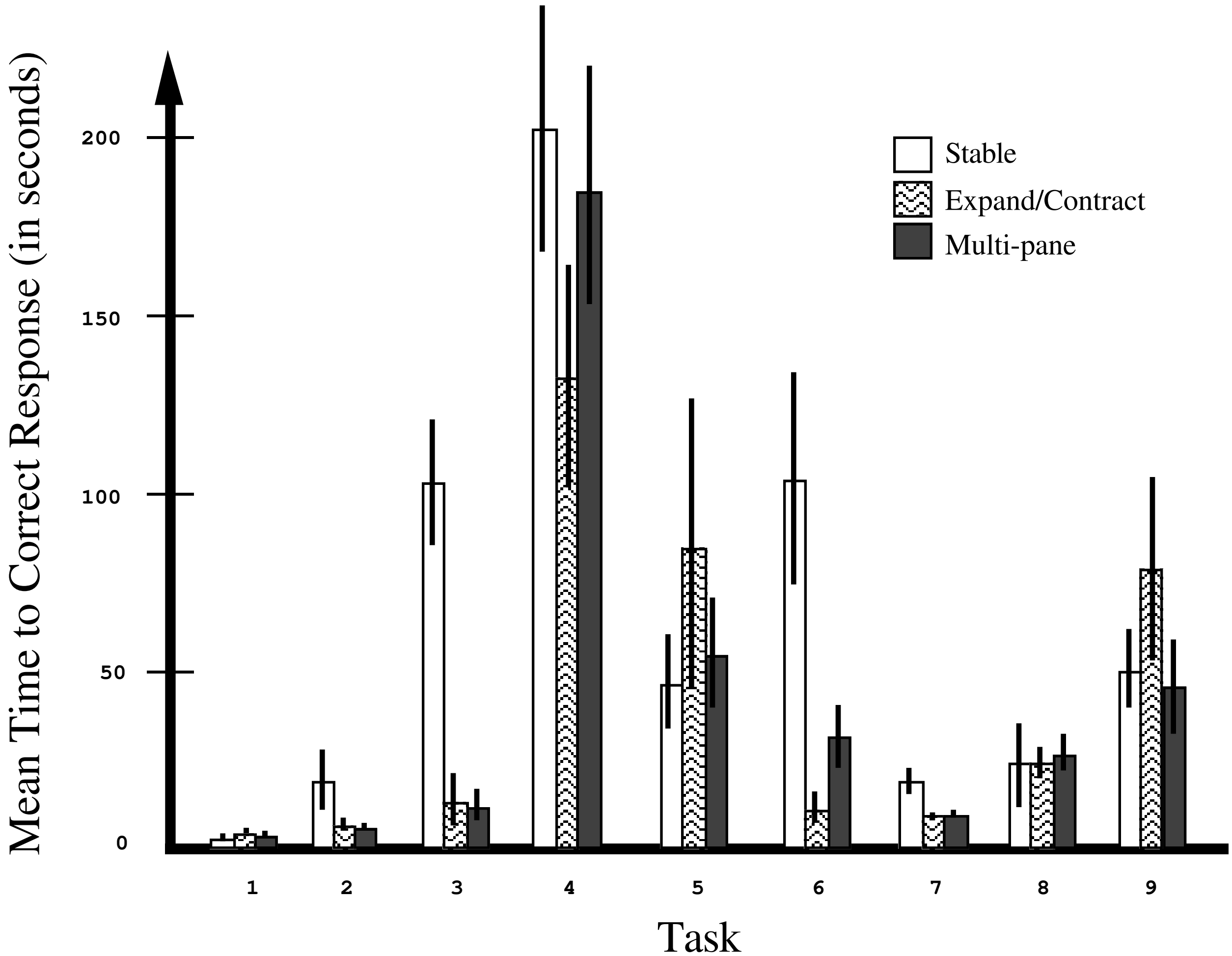

Figure 4. Mean time in seconds to complete correctly each task. The thick lines within each bar represent one standard deviation centered about the mean. These data are provided textually in Table 2.

Task #		Stable		Expand/Contract		Multi-pane	*F*-value
1	++	2.0 (0.9)	–	4.3 (1.3)	++	2.4 (1.0)	17.92
2	–	19.4 (16.3)	++	7.4 (3.1)	++	6.4 (1.6)	7.87
3	–	103.3 (33.6)	++	13.9 (15.6)	++	11.6 (8.1)	79.91
4	–	202.0 (69.0)	+	131.3 (61.9)		184.1 (67.0)	4.34
5		46.1 (27.1)		85.9 (82.5)		55.0 (30.7)	2.15
6	–	105.1 (59.5)	++	11.7 (8.0)	++	31.8 (16.0)	26.24
7	–	19.3 (7.6)	++	9.7 (2.6)	++	10.0 (2.5)	17.58
8		23.3 (22.0)		23.1 (7.3)		26.8 (10.1)	0.28
9		49.9 (21.7)	–	77.8 (51.2)	+	45.9 (27.4)	3.29

Table 2. Mean time in seconds to complete correctly each task with standard deviation in parentheses. Two plus signs (+ +) or one plus sign (+) indicate statistically significant differences ($p < .01$, $p < .05$, respectively) favoring that interface(s) over the time with a minus sign (-) for that task as determined by the Tukey post-hoc test. These data are provided graphically in Figure 4.

4 Discussion

As expected, both the expand/contract and multi-pane interfaces produced faster performance times than the stable interface. The differences predicted between the expand/contract and multi-pane interfaces across tasks favored the multi-pane interface for task 1 and task 9.

Another major factor in the relatively good performance of the expand/contract interface may be the animation characteristics of the expansion and contraction of TOC items. We have seen other systems that use the expand/contract feature yet the animation characteristics were poor, simply a complete or partial window refresh with all the new information present (or hidden). Our animation used revealing (or hiding) rapidly *one new item at a time* in the proper order until all new items were displayed (or hidden). We feel that this helped users locate where the new information was being put (or taken away) than would the paint-all-at-once non-animation characteristics of other systems. Section 5.2 describes different ways the animation characteristics can be performed with their tradeoffs.

4.1 Observed General Browsing Characteristics

This section describes the observed browsing characteristics of the participants in our experiment for each of the three interfaces. The observations are supported by collected user interaction data which logged mouse events for scrolls and TOC item clicks.

Stable Interface

Participants often read or scanned carefully all the items on one screen before scrolling to the next screen of information. Participants almost never scrolled quickly just looking for the flushleft chapter titles to guide them through the TOC as they might when using a physical book. Participants rarely took advantage of the gross scrolling feature of the scroll bar, but rather scrolled most often screen-by-screen or continuously by a single line at a time. This behavior seems to be due more to the fact that they were unfamiliar with the TOC structure and size rather than being unfamiliar or uncomfortable with the different features of the scroll bar; for those participants that attempted gross scrolling often did not arrive close to their target location and went back to the safe approach of scrolling screen-by-screen.

Expand/Contract Interface

Users of this interface used chapter titles to guide their browsing through the TOC--they had no other choice. Participants usually left items expanded when doing most of the tasks. The exception to this was during heavy exploration tasks (tasks five and nine); during that time, participants more often contracted TOC items after viewing their subordinates. It seems that when the purpose was to find one thing, participants would leave their “trail” visible, perhaps to remind them where they had been and/or of the information in those TOC items, even if their first target locations were not correct. But when the purpose was to view a large portion of the TOC, they would systematically expand and then contract TOC items so as not to waste valuable screen space with unneeded, expanded items.

Many participants reacted when an item expanded into more than about a dozen subordinates. Often participants physically moved away from the screen and/or chuckled when such an expansion happened. Perhaps they were overwhelmed by the amount of new data made visible in the TOC; perhaps the animation characteristics take on a humorous look when so many items are inserted one item at a time.

Multi-pane Interface

Users of this interface used chapter titles to guide their browsing through the TOC. The information changed very quickly to the new information to be displayed, and thus exhaustive exploration was a popular interaction style. However, it seemed that some participants forgot that the chapter and section panes could be scrolled, and often they explored only the TOC items visible in those two panes; yet participants did not seem to forget the subsection pane could be scrolled. This could be because the training TOC almost always had all the chapter and section items visible and scrolling them was not necessary, but scrolling was still necessary in the subsection pane. Thus when the large TOC was encountered, it seemed that some participants forgot scrolling was a navigation technique for the two upper-level panes, even though they had scrollbars and there was no whitespace at the bottom of those text panes which would indicate that there was no further information in the pane. When trying their first large-scale search, many participants made a remark that they were lost. Initially, they were visibly shaken or disappointed in themselves for getting lost, but they quickly realized (within ten seconds) that the navigation features of this interface made being lost an easy state from which to recover, they could just click on a chapter item to display its contents and anchor the context again.

4.2 Individual Timed Tasks

Most of the tasks that favored the expand/contract and multi-pane interfaces over the stable interface can be explained by the fact that the stable interface required **excessive** scrolling and more reading to locate proper TOC items to complete tasks. This was borne out in our pilot studies and some attempts were made to lessen this frustration in the large TOC by locating target items towards the beginning more than would happen on average. In contrast, the expand/contract and multi-pane interfaces allowed participants to explore much more easily alternate paths toward answers without much penalty to backtrack or change paths. Contracting an explored item or simply selecting another item to view in the other panes is much faster and easier than trying to scroll to distant parts of the TOC as would be done in the stable interface.

For many tasks, most notably Task 9, performance differences between the expand/contract and multi-pane interfaces may be explained as tradeoffs in design choice for the animation speed of the expand/contract interface. If it is too fast, then users may become disoriented as to where new information appeared. If it is too slow, then users are not able to perform actions as fast as they may want (since user action is prohibited until the animation is complete). The multi-pane interface displayed new information very rapidly and did not require a contraction step to keep information from becoming fragmented or overwhelming, thus allowing exhaustive exploration to occur faster than with the expand/contract interface. Because the new information always appeared in the same place(s) in the window in the multi-pane interface, participants would be less likely to become disoriented as to where new information was to be found. Thus this method of rapid display of new information posed little problem with disorientation.

Task 1: What is the title of section 1.1?
This was a very simple navigation task that was created to be very easy to accomplish in order to build confidence in participants' minds. Both the multi-pane and stable interfaces allowed users to perform significantly faster than the expand/contract interface because both those interfaces had the answer displayed on the initial screen. The expand/contract interface's initial display did not have the answer on the screen; users had to click on chapter one to expand its subordinates onto the screen. The data matched our expectations of performance among the interfaces.

Task 2: What is the title of subsection 8.2.3?
This task was created as a natural successor to task one, it was the same type of task but more effort was needed to accomplish it. Both the multi-pane and expand/contract interfaces allowed users to perform significantly faster than the stable interface. The two fisheye interfaces needed only two mouse clicks and no scrolls to reveal the answer. However, the stable interface needed twenty full-screen scrolls with much scanning of information to determine if the answer was present. The other two interfaces required little reading since with each TOC item click users knew they were homing in directly on the answer. Users of the stable interface could have used gross scrolling techniques rather than full-screen scrolls, but luck or great estimation skills (quite difficult without knowledge of the TOC structure) would have been needed to arrive at the correct spot in the TOC.

Task 3: What chapter, section, or subsection number, if any, has the title, "Radar Departure"?
This was a textual searching task where participants had to find a TOC item given the name (the reverse of tasks one and two); the form of the response was to report the section number of the title. This task was made purposely straightforward by giving a title whose location in the TOC would be clear from the chapter titles. We acknowledge that this task would be performed better by a text search function in the interface, which any good browsing interface should have. But that does not test the differences in interfaces due to their unique features which is this experiment's goal. Both the multi-pane and expand/contract interfaces allowed users to perform significantly

faster than the stable interface. In both those interfaces, the appropriate chapter title was in view in the initial display and only one mouse click was needed to reveal the answer, thus they had similar completion times. The stable interface once again required much reading of every screen of information, and screen-by-screen scrolls would have required thirteen scrolls to find the answer. Gross scrolling would have been completely ineffective since they were searching for a specific text string and not just scanning for numbers as in Task 2.

<u>Task 4:</u> What chapter, section, or subsection number, if any, has the title, "Aerial Refueling Fire Emergencies"?
This task was similar to Task 3, however, the title given to participants <u>did not exist</u> in the TOC; tricky, yes, but a real world task that occurs often. This type of task was also given during the training TOC. The expand/contract interface allowed users to perform significantly faster than the stable interface. We created a title that we felt would logically seem to be in either of two chapters in the TOC. We hoped participants would search the two candidate chapters exhaustively, then state that the item did not exist in the TOC. Indeed, this browsing model occurred most often in the expand/contract interface and least often in the stable interface; thus the statistically significant result. Multi-pane users did more exhaustive searching than did expand/contract users; we feel this occurred because of the quicker response time for showing new information provided by the multi-pane interface. This incurred less of a penalty for exploring TOC items (that may not have so obviously contained the target title) since it was quicker to do so than in the expand/contract interface (which takes more time due to animation characteristics to show new information via expansion and "doubly that much more time" if those items were contracted), so multi-pane users explored more than the other interface users. This result did not match our expectations of performance between the expand/contract interface and the multi-pane interface; we felt participants would behave the same with these two interfaces. Instead, the mean time for the multi-pane interface was almost fifty percent longer than the mean time for the expand/contract interface, though this difference was not statistically significant due to high variance. It is slightly ironic that the interface with the smaller penalty for more exhaustive browsing would produce longer overall task completion times than the interface with the larger penalty. Perhaps this is analogous to setting lofty goals enabled by improved technology, e.g., the possibility of high achievement in, say, using a word processor with rich functionality. One might strive for perfect formatting which could take more time than using a crude tool where the formatting goal is much more modest.

<u>Task 5:</u> Name the table of contents item that deals with all of the topics of approach lights, runway lights, and taxiway lights.
This task asked for the single, most specific TOC item that contained information on all of three given topics which may have involved the highest cognitive load for participants. The form of the response was to state the TOC item's number. Not only did this task require considerable browsing, but two kinds of mental effort: a) understanding the given topics and creating or

recognizing a higher order concept in the TOC, and b) the lower cognitive load task of verifying all given topics are covered by a chosen TOC item. There were no significant differences across the three interfaces for this task. What was surprising about the mean times for the interfaces was that the stable interface was fastest (stable 46.1, expand/contract 85.9, multi-pane 55.0). Actually, the mean times may be explained by our design to have the target located near the beginning of the TOC so as not to completely frustrate the stable interface users which apparently negated obtaining statistically significant differences. The expand/contract and multi-pane interface users took a goal directed browsing approach and carefully chose which chapters to explore. This was not so fruitful because the chapter title that contained the target TOC item was not straightforwardly indicative of its contents, though the section title that contained the target TOC item was indicative of its contents.

Task 6: What is the number of the chapter that appears first in this document for the chapter titles:
"Anti-Drug Operations"   "Military Training Routes"

This task was to identify which of two given chapter titles appeared first in the TOC, a sibling comparison task at the chapter level. The form of the response was to state the chapter number. The multi-pane and expand/contract interfaces allowed users to perform significantly faster than the stable interface. The reason the stable interface fared poorly was that it needed to be scrolled many times to compare chapter titles. This mostly matched our expectations of performance among the interfaces, though we thought the expand/contract interface would allow for statistically significantly faster performance than the multi-pane interface. This is because the expand/contract interface had all the information on the screen with the initial display, but the multi-pane interface had to have its chapter pane scrolled (by task design) to determine the answer. There was almost a factor of three difference between the mean times of these two interfaces (11.7s for expand/contract, 31.8s for multi-pane) which does support our hypothesis, but this difference was not statistically significant due to high variance in mean times.

Task 7: In chapter 2, General Control, what is the number of the section that appears first:
"Weather Information"   "Radio and Interphone Information"

The task was very similar to Task 6 except sibling comparison was at the section level within a specified chapter. We created this task to see if the performance would be different when the targets were in the section level of the TOC rather than the chapter level. The multi-pane and expand/contract interfaces allowed users to perform significantly faster than the stable interface. The similar completion times for expand/contract and multi-pane users is due to the need for both to click on the given chapter title to see its sections, and from there sibling comparison time was equivalent. Stable interface users had subsection titles separating section titles, thus scrolling was still needed to compare section titles which took longer to perform.

Task 8: In section 2.1, General Control Information, what is the number of the subsection that appears first:

"Traffic Advisories" "Operational Requests"

This task was very similar to Tasks 6 and 7, except that sibling comparison was at the subsection level within a specified section. There was no significant difference in speed among the interfaces for this task. Stable interface users would have performed much worse, but the target location was in chapter two (by design) so that stable interface users were able to scroll there quickly. Once there, subsection titles to be compared were all in view, an uncommon occurrence in the stable interface that only happens at the subsection level within one section. Multi-pane users had to scroll the subsection pane when the proper section had its subordinates displayed, thereby slowing their performance enough to allow stable interface users to perform as well. The animation characteristics of the expand/contract interface (two expansions were needed) slowed their performance enough to allow stable interface users to perform as well. However, notice the large variance for stable interface users but the low variance for expand/contract and multi-pane users.

Task 9: Does this table of contents contain more information about:

radar

emergencies

The task was to identify which of two topics was covered more thoroughly in the TOC, which involved a large amount of exploratory browsing. The form of the response was to state which topic was covered more thoroughly and why (e.g. the fact that radar had three times more TOC items than emergencies). The multi-pane interface allowed users to perform significantly faster than the expand/contract interface. The multi-pane users were able to do exhaustive exploration of the two chapters quicker than the expand/contract users. The difference in times for the multi-pane and expand/contract interfaces did match our expectations because multi-pane has the quicker display of new information upon receiving a mouse click, which allows for more browsing to take place in the same amount of time than with the expand/contract interface. We conjecture that the reason the stable interface users did as well as they did was because by this point in the experiment they were familiar with the structure of chapters in the TOC; this task used topics encountered in other tasks. Thus some participants attempted a gross scroll, or did screen-by-screen scrolls but with much less reading and more quick scanning than in previous tasks, to get to the correct locations within the TOC to do the necessary brief analysis. This result did not match our expectations for the stable interface. These users performed better than anticipated due to the fact that our task designs were created so as to ensure stable interface users *had a chance* at performing well. In our pilot studies they were constantly and visibly frustrated with scrolling being their only navigation technique.

4.3 Refined Theories

The clicks model prediction that the expand/contract and multi-pane interfaces would produce far faster performance than the stable interface was confirmed. Since the clicks model deals only with navigation, and not scanning or interpreting, it is more accurate in predicting performance on tasks such as Task 2. The prediction of approximately six times the effort for the stable model is near to what we found, if adjustment is made for the target location. A more accurate model would account for the varying effort preceding different clicks (repeated clicks on the scroll bar require little intervening cognitive effort), tasks requiring more scanning and interpretation to match phrases or concepts, familiarity with the TOC contents, and non-well-formed TOCs. While these refinements might improve the accuracy of predictions, they would detract from the inherent simplicity, objectivity, and intuitive appeal of the current clicks model.

5 Conclusion

Our exploratory evaluation used minimally functional implementations with unique view presentation strategies. We studied the impact of view presentation on navigation and effective hierarchy browsing.

5.1 Practitioner's Summary

Designing an effective browser for an online, hierarchical document requires an understanding of supported tasks plus several navigation and display strategies among which to choose. Online browsing can be facilitated by interface designs that present appropriate information in appropriate places. Designers must take into account dynamic and animation characteristics, font attributes, position attributes, consistency, and metaphors used by the system.

We found evidence that the expand/contract and multi-pane views reduce browsing times compared to a fully expanded, stable view for large hierarchies for just about every task. We suggest a stable interface be replaced with an expand/contract interfaace because it is similar enough and provides many user benefits. Multi-pane is better than expand/contract for heavy exploration tasks due to higher speed, less clutter, and no "housekeeping." These two interfaces performed equally well for our tasks involving sibling comparison at any level. Finally, multi-pane always performed statistically as well as expand/contract. We recommend a multi-pane interface design as a primary strategy unless other factors intervene.

5.2 Future Research

We recommend that future research concentrate on evaluating fisheye views and other enhancements to information presentation rather than considering full, stable views of hierarchies.

Future research could explore the animation characteristics of the expand/contract interface. As noted many times, there is a tradeoff inherent in deciding on the speed of expansion/contraction animation. It would also be interesting to find out why users physically move away from the screen during a large expansion. Does this affect performance? Should the system only present some of the many items being inserted, with a "more" button or action that inserts more of the items? Some possible ways to alter the animation characteristics of the expand/contract interface are:

- The contract operation could be a one time paint operation.
- Use display markers or other effects (highlighting, underlining, fonts and font size, etc.) rather than animation characteristics to show where the new information is displayed.
- Display only some of an item's subordinates for the initial expansion and have users request/confirm for more.
- If an expansion requires more screen space than is available below the item clicked, do not scroll or move the item clicked.
- Allow users to adjust the speed of inserting one item at a time until the most appropriate speed is determined.
- Automatically adjust the speed of inserting one item at a time *within* an implementation, displaying the first several items slowly (to allow users to begin reading/scanning) and speeding up the insertion of the remaining items.

More research could investigate features the multi-pane interface can support. One feature could be to allow users to adjust the percentage of screen space devoted to a particular pane or to a particular level of the hierarchy. These two cases are different if a TOC has more levels than there are panes, thus a particular level may not always be contained in the same pane. Study the effect of using a hierarchy that has more levels than there are panes. Natural implementations might be to increase the number of panes or have levels move upwards in the column of panes to accommodate deeper levels. How can this migration be clearly communicated to users? What effect would many migrations have on overall comprehension of the hierarchy?

A display feature that could be studied in both the multi-pane and expand/contract interfaces is the use of markers for showing users that an item has been explored in this session, across sessions, and/or at user discretion (i.e. to reset the visited state). Will it prevent needless, time-consuming expansions in the expand/contract interface? Will it hinder performance because users could forget when they visited an item--was an item visited before the "current" task was begun or not?

Another experimental method could be used to study the cognitive actions of participants when carrying out a browsing strategy--did times vary because one participant carefully scanned the high level items for an appropriate item to expand where another participant systematically expanded the items one by one until an answer was found? Asking the participants to describe verbally their search processes or analyzing more completely logged mouse events could shed light on some of the differences in task completion times.

Our participants were novice computer users with low domain knowledge of the TOCs. This audience may not be suitable for many applications that have more computer and task domain knowledgeable users. Performing an experiment that has participants use the same TOC many times, or has participants browse hierarchies they themselves have created may find that differences in views and interaction features have a more significant effect on browsing under these conditions.

Acknowledgements

We appreciate the support of Sun Microsystems, Inc., on this project. Specifically, we thank Kate Ehrlich for being the liaison between Sun and the Human-Computer Interaction Laboratory and for contributing much insight to guide our research directions. We thank Sharon Mark and Kay Wolman for assisting in the administration of the experiment and contributing to preliminary versions of this paper. The pilot subjects and participants in the experiment deserve thanks. We also appreciate Peter Polson's devoted efforts and wise counsel with respect to the statistical results plus the other reviewers' recommendations.

References

1. Borgman, C. L., Gallagher, A. L., Krieger, D., & Bower, J. Children's use of an interactive catalog of science materials. In *Proceedings of the 53rd American Society for Information Science Annual Meeting* (Toronto, Ontario, Canada, Nov. 4-8, 1990). Learned Information, Medford, NJ, 1990, pp. 55-68.

2. Campagnoni, F. R. & Ehrlich, K. Information retrieval using a hypertext-based help system. *ACM Transactions on Information Systems 7*, 3 (1989), pp. 271-291.

3. Card, S. K., Moran, T. P., & Newell, A. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, NJ, 1983.

4. Carey, T. T. , Hunt, W. T., & Lopez-Suarez, A. Roles for tables of contents as hypertext overviews. In *Proceedings of Human Computer Interaction--INTERACT '90* (Cambridge, U.K., 27-31 Aug., 1990). Elsevier , North Holland, 1990, pp. 581-586.

5. Chimera, R., & Shneiderman, B. Three interfaces for browsing hierarchical tables of contents. In *Human-Computer Interaction Laboratory Open House '91 Videotape,* Plaisant, C., Ed. Center for Automation Research, University of Maryland, College Park, Maryland 20742-3255. 1991.

6. Chimera, R., Wolman, K., Mark, S., Shneiderman, B. (1991). *Evaluation of three interfaces for browsing hierarchical tables of contents.* Technical report CS-TR-2620, CAR-TR-539, Center for Automation Research, University of Maryland, College Park, Maryland 20742-3255. 1991.

7. Chin, J. P., Diehl, V. A., & Norman, K. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of CHI '88 Conference on Human Factors in Computing Systems* (Washington, D.C., May 15-19, 1988). ACM Press, New York, 1988, pp. 213-218.

8. Egan, D.E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., & Lochbaum, C. C. Formative design evaluation of SuperBook. *ACM Transactions on Information Systems 7,* 1 (1989), pp. 30-57.

9. Egan, D.E., Remde, J. R., Landauer, T. K., Lochbaum, C. C., & Gomez, L. M. Behavioral

evaluation and analysis of a hypertext browser. In *Proceedings of CHI '89 Conference on Human Factors in Computing Systems*, (Austin, TX, Apr 30 - May 4, 1989). ACM Press, New York, 1989, pp. 205-210.

10. Federal Aviation Administration. *Air Traffic Control Handbook* [7110.65] and *Military Handbook* [7610.4]. 1991.

11. Furnas, G. W. Generalized fisheye views. In *Proceedings of CHI '86 Conference on Human Factors in Computing Systems* (Boston, Mass., Apr. 13-17, 1986). ACM Press, New York, 1986, pp. 16-21.

12. Goldstein, M. J., Baker, B. L., and Jamison, K. R. *Abnormal Psychology: Experiences, Origins, and Interventions*. Little, Brown, and Company, Boston, 1986.

13. Lee, E., & MacGregor, J. Minimizing user search time in menu retrieval systems. *Human Factors, 27*, 2 (1985), pp. 157-162.

14. Marchionini, G. An invitation to browse: designing full-text systems for novice users. *Canadian Journal of Information Science, 12*, 3/4 (1987), pp. 69-79.

15. Remde, J. R., Gomez, L. M., & Landauer, T. K. SuperBook: An automatic tool for information exploration - hypertext? In *Proceedings of Hypertext '87 Conference* (Chapel Hill, N.C., Nov. 13-15, 1987). ACM Press, New York, 1987, pp. 175-188.

16. Robertson, G. G., Mackinlay, J. D., and Card, S. K. Cone trees: Animated 3D visualizations of hierarchical information. In *Proceedings of CHI '91 Confernece on Human Factors in Computing Systems* (New Orleans, La., Apr. 27 - May 2, 1991). ACM Press, New York, pp. 189-194.

17. Shneiderman, B. *Designing the User Interface*, 2nd edition. Addison-Wesley Publishing, Reading, MA, 1992.

18. Shneiderman, B. *Designing the User Interface*. Addison-Wesley Publishing, Reading, MA, 1987.

19. Thompson, D. A. Interface design for an interactive information retrieval system: A literature survey and a research system description. *Journal of the American Society for Information Science, 22*, 6 (1971), pp. 361-373.

20. Walker, J. H., Young, E., & Mannes, S. A case study of using a manual online. *Machine-Mediated Learning, 3*, 3 (1989), pp. 227-241.