Split Menus: Effectively Using Selection Frequency to Organize Menus

ANDREW SEARS

and

BEN SHNEIDERMAN
University of Maryland

When some items in a menu are selected more frequently than others, as is often the case, designers or individual users may be able to speed performance and improve preference ratings by placing several high-frequency items at the top of the menu. Design guidelines for *split menus* were developed and applied. Split menus were implemented and tested in two in situ usability studies and a controlled experiment. In the usability studies performance times were reduced by 17 to 58% depending on the site and menus. In the controlled experiment split menus were significantly faster than alphabetic menus and yielded significantly higher subjective preferences. A possible resolution to the continuing debate among cognitive theorists about predicting menu selection times is offered. We conjecture and offer evidence that, at least when selecting items from pull-down menus, a logarithmic model applies to familiar (high-frequency) items, and a linear model to unfamiliar (low-frequency) items.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms: Design, Human Factors

Additional Key Words and Phrases: Human-computer interaction, menus, selection frequency, split menus, user interface

1. INTRODUCTION

Menus are an increasingly popular method of interacting with computers, and therefore, designers are paying greater attention to menu organization so as to speed learning and performance. Designers must not only decide the

This was supported by NASA grant NGT-50762.
Authors' addresses: A. Sears, Department of Computer Science and Information Systems, DePaul University, 243 South Wabash Avenue, Chicago, IL 60604; email: sears@cs.depaul.edu; B. Shneiderman, Human–Computer Information Lab and Computer Science Department, Institute for Systems Research, University of Maryland, College Park, MD 20742; email: ben@cs.umd.edu.

overall organization of the menus for an application, they must also decide how to organize the items within each menu. Typical choices include alphabetical, logical, or categorical organizations.

When menus are relatively short, as they are in many commercial products, traditional organizations work well. However, situations exist where menus are long (e.g., control panels, font menus, b-board lists, and many custom applications), and alternative organizations may prove useful. When menus get longer and a small subset of items are selected more frequently then the remaining items alternate methods for selecting these high-frequency items may prove useful. One alternative is to assign special key combinations that can be used to select these frequently used items. This works well in some situations, but it can quickly become overwhelming. For instance, 41 command key combinations are defined for the editor used to write this article. A small subset of these key combinations proves very useful, but if users do not remember the necessary combination all benefits are lost. Reorganizing menu items will allow faster selections while maintaining the standard selection mechanism and eliminating the need for users to remember additional commands.

In everyday life, people often consider how frequently they use items when organizing them. Phone books often have an easily accessed section for frequently dialed numbers, and bookcases are often organized with a section for frequently used books. Several researchers have suggested organizing menu items by how frequently they are selected [Brown 1988; Norman 1991; Paap and Roske-Hofstrand 1988; Shneiderman 1992; Smith and Mosier 1986]. Often the advice is simply to place the most frequently selected items at the top of the menu. When more detailed advice is given, designers are instructed to place the most frequent item at the top and the least frequent item at the bottom, implying that all items should be ordered by how frequently they are selected. This can lead to faster selections, but users may feel that the menu has no obvious organization.

When giving guidelines for formatting data, Smith and Mosier [1986] suggested: "Where some data items are used more frequently than others, consider grouping those items at the top of the display." They do not indicate how to select these "high-frequency" items, how to organize them at the top of the display, or what to do with the remaining items. Refining this suggestion and applying it to menus results in *split menus* [Sears 1993a].

Split menus are created by splitting a menu into two sections. Designers or individual users place frequently selected items in the top section and infrequently selected items in the bottom section. Split menus should prove useful when a small subset of the menu items represent the majority of selections. By moving these frequently used items to the top of the menu, users should be able to locate and select them more rapidly. As the length of the menu increases, the potential benefits of split menus also increase.

This research investigates the efficacy of split menus and offers a theoretical model for predicting performance times. We suggest a resolution to the continuing debate among cognitive theorists about predicting menu selection times. We conjecture and provide evidence that a logarithmic model applies to familiar (high-frequency) items, and a linear model to unfamiliar (low-frequency) items.

This article discusses details of split menus including a model that predicts the benefits. Two *in situ usability studies* are described that demonstrate the potential of split menus in normal working conditions followed by a controlled experiment which provides evidence supporting the efficacy of split menus and the accuracy of our cognitive model. Finally, a guideline is proposed to assist designers in creating split menus.

2. RELATED RESEARCH

People often apply frequency of use to organize objects and thus make frequent tasks easier. Similarly, using the frequency of actions to organize user interfaces should result in faster performance as well as improved user preference ratings [Rubinstein and Hersh 1986; Sears 1993b].

Many menu organizations have been suggested such as alphabetical, logical, categorical, or frequency of use. Alphabetical organizations order the items based on the lexical order of item names and are one of the orderings referred to as 'traditional' throughout this article. Traditional organizations include alphabetical, numerical, and chronological (e.g., Monday, Tuesday, Wednesday, ...). Logical organizations order menu items by the logical relationships between the items (e.g., inches, feet, yards, ...). Determining what is a logical orgnaization can be a highly subjective task. A categorical organization can often be imposed on a group of items (e.g., ways of opening files, ways of closing/saving files, ways of exiting, ...). Categorizing items is also highly subjective. Frequency-based orderings typically refer to placing the most frequently used item at the top of the menu, followed by the next most frequent item. This continues until all items are placed in the menu. Many guidelines documents suggest using one or more of these organizations depending on the items being displayed [Apple Computer 1987; Dept. of Defense 1991; OSF 1990; Smith and Mosier 1986; Sun Microsystems 1990]. Additional methods of organizing items include: sequential (listed by sequence of use), functional (similar to logical), and importance (place critical items first) [Brown 1988; Norman 1991].

Several studies have investigated the effects of menu organizations on user performance. Somberg [1987] compared four menu organizations: alphabetic, probability of selection, random, and positionally constant. However, it is important to remember that for the alphabetic, probability, and random organizations, menu items changed positions between each selection making it impossible to learn the location of an item. Initially, alphabetic or probability ordering were fastest, but after practice, menus that maintained a constant position for each item proved fastest. Random organizations were the slowest throughout the study. These results indicate that keeping words in fixed locations is better than allowing the words to move within a menu. However, it does not provide a comparison between various methods of organizing items in a positionally constant menu. Card [1982] compared positionally constant alphabetical, categorical (called functional by Card), and random organizations. Alphabetically ordered menus were the fastest, and randomly ordered menus were the slowest. These results indicate that in addition to keeping menu items in fixed locations, a meaningful organization should also be used. While there is no simple answer to the question of which organization to use, it is clear that providing users with a stable menu that uses a known organization results in significant benefits.

Another alternative is to dynamically organize the menu based on the current frequency of selection. This could lead to a menu that changes automatically after users make selections, or to a system that is under user control and only changes when the user decides that a change would be beneficial. Mitchell and Shneiderman [1989] compared static menus and menus that were automatically reorganized based on the users' current pattern of selections and found that users preferred static menus. Additionally, when comparing the first exposure to the system, users were faster and made fewer errors with static menus. After practice there was no difference in performance, but users still preferred static menus. Greenberg and Witten [1985] investigated the benefits of organizing menus based on an a priori set of frequencies and updating the menu to reflect recent usage as users make selections. The results suggests that organizing items by frequency and recency of use may prove useful. These two studies indicate that automatically updating menus to reflect current usage patterns may be useful, but can also lead to problems.

Other attempts at speeding up menu selection have used nonlinear menus. Callahan et al. [1988] investigated the benefits of circular (pie) menus which make the distance to each item equal, while Walker and Smelcer [1990] explored the benefits of making menu items larger the farther down they were in a menu. Both of these research efforts focused on making the movement to a menu item easier, and both demonstrated that this can save users time.

3. SPLIT MENUS

The benefits of using split menus depend on two factors: how often each item is selected and where the frequently selected items are located in the traditional and split menus. If all items are selected with near-equal frequency, minimal benefits would be expected for split menus. However, if a few items represent the majority of selections, as is often the case with computer commands [Greenberg 1988], it is likely that split menus will improve performance. If the frequently selected items are located at the top of the traditional menu, minimal benefits would be expected, but if they are at the bottom of the traditional menu, split menus should prove beneficial. However, even in situations where users do not save time, they may prefer a menu organization that emphasizes frequently selected items.

Organizing Split Menus

To facilitate rapid scanning of both the low- and high-frequency sections, particularly if the menu is relatively large, both sections should be presented in the 'traditional' order the entire menu would have been presented in. For instance, if the menu contains names, it would be organized alphabetically. If it contains dates, chronological ordering would be used. Organizing both sections in a traditional order allows users to search each section, as they would normally search the entire menu, quickly locating the item of interest.

We developed preliminary guidelines which help decide which items should be placed in the high-frequency section of a split menu. These preliminary guidelines were developed strictly for use in our evaluation of split menus and are replaced later in this article by a guideline based on the results of the field studies and the controlled experiment.

Preliminary Guideline 1. Limit the number of items in the high-frequency section to four or less in most situations. This will allow users to quickly scan and remember the high-frequency items. A few more items may be placed in the high-frequency section to preserve some meaningful organization or categorical relationship.

Preliminary Guideline 2. Sort all items by selection frequency. Starting with the least frequently selected item, scan until the increase in frequency between two successive items is greater than the mean of the frequencies. Once this point is located, all items on the high-frequency side of this point are placed in the high-frequency section. If there are more than four items, only the four most frequently selected items are placed in the high-frequency section.

Applying these preliminary guidelines to the menu in Figure 1, using hypothetical selection frequencies, results in Figure 2.

Predicting the Benefits of Split Menus

If the benefits of using a new menu organization are uncertain, users may be hesitant to risk changing their system. However, if a reasonably accurate prediction can be made, users may be more willing to switch to a new organization. To predict the amount of time that will be saved we must understand what happens when users select items from a menu. First, there are at least two different types of selections that users make. When users know the name of the item they want, they must locate and select it. When users know the action they want to perform, but do not know the name of the menu item that performs that action, they must search for an item that they feel performs the desired action and select it. Our discussions will be limited to interactions where users know the name of the item they want to select. Split menus may also be beneficial when users are unsure of which item to choose because likely choices are displayed at the top of the menu.

Several models have been developed to predict the amount of time necessary to select an item from a menu. These models typically fall into one of two categories: linear or logarithmic [Landauer and Nachbar 1985; Lee and MacGregor 1985; Norman 1991; Paap and Roske-Hofstrand 1986]. Linear models predict that the amount of time to select an item is a linear function of the position of the item in the menu [Lee and MacGregor 1985]. These models typically assume that the majority of the selection time is spent

Fig. 1. Traditional menu.

searching the menu in a linear fashion for the desired item. Logarithmic models predict that the amount of time to select an item is a logarithmic function of the position of the item in the menu. Logarithmic models assume that users do not scan the menu for the item linearly, but use the order of the menu items to search more efficiently. For instance, users may move to the center of the menu and decide if the desired item is above or below the center. Using this information users move to the middle of the upper or lower half of the menu and repeat the process until the desired item is located. Fisher et al. [1990] developed a model that accounted for different menu items being selected with different frequencies and extended previous work with the linear model to a larger variety of menus.

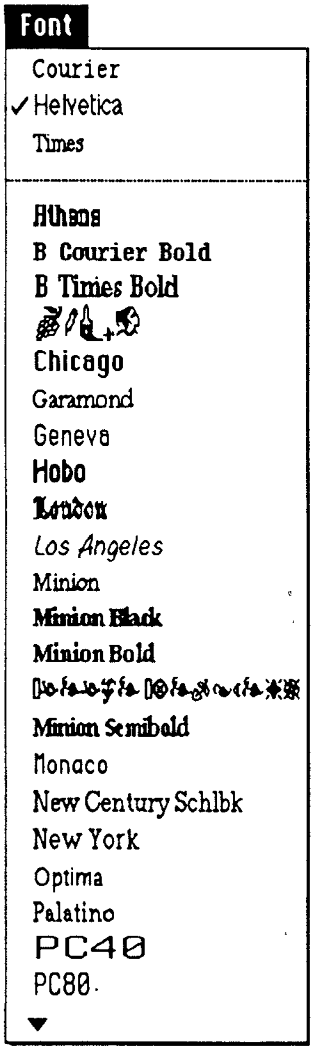

Fig. 2. Split menu.

The first limitation of these models is that they assume that users select all menu items using the same strategy. We believe that different strategies are used depending on how familiar users are with the item being selected. Second, each of these models assumes menu selection is performed by pressing a key on a keyboard. This is not accurate for many mouse-based computer menus which are widely available. Our model must deal with menu selections which are performed using a mouse and a pull-down menu. When selecting items from a pull-down menu users must move the cursor to the top of the menu, press a mouse button to select the menu, move the cursor to the desired item while holding the mouse button down, and release the mouse button to select the highlighted item. During the process of searching for the correct item, the mouse cursor acts as a visual anchor changing the way the search proceeds.

We suggest a cognitive model in which users are not equally familiar with all items in a menu. Users' familiarity with an item is assumed to be a function of how frequently they select the item. Items that are selected frequently will be more familiar than items selected infrequently.

Users quickly learn the location of frequently selected items allowing them to move the cursor directly to the item of interest. This indicates that the time to select a frequently selected menu item should be dominated by the time necessary to move the cursor to the correct location. Therefore, selection times for high-frequency items should be a logarithmic function of the location of the item in the menu, as would be predicted by Fitts' Law. Infrequently selected items do not become familiar to users; therefore users cannot automatically move to the correct location in the menu. This lack of familiarity forces users to scan the menu for the desired item. Since users must move the cursor down the menu to the item of interest, the cursor acts as a visual anchor guiding their search. This results in users scanning the first item, then the second, third, etc. until the desired item is located. Therefore, the selection time for an infrequently selected item should be a linear function of the position of the item in the menu. Selection times for items that are selected occasionally would be predicted by some intermediate model, but for the purpose of predicting the benefits of split menus, either a linear or logarithmic model will be used for each item in the menu.

Split menus divide menu items into two categories: high- and low-frequency items. Assuming that users are familiar with the high-frequency items, we use a logarithmic model to predict times for these items. Similarly, assuming users are not familiar with the low-frequency items, a linear model will be used to predict times for these items. Of course, users will be somewhat familiar with many of the low-frequency items, so predictions are expected to be less accurate for the low-frequency items than for the high-frequency items.

To predict the amount of time that will be saved, or lost, by switching from a traditional organization (alphabetical) to split menus the following values must be known:

—$f(x)$: how frequently item x is selected as compared to other menu items ($\Sigma f(x) = 1$).

—$L_T(x)$: the location of item x in the Traditional Menu (often organized alphabetically). In Figure 1 L_T(Courier) = 6, L_T(Helvetica) = 9, and L_T(New York) = 20.

—$L_S(x)$: the location of item x in the Split Menu. In Figure 2 L_S(Courier) = 1, L_S(Helvetica) = 2, and L_S(New York) = 22. $L(x)$ is equal to one plus the number of items above item x in the menu (including the dividing line for split menus).

—$t_T(x)$: the average amount of time to select item x in the Traditional Menu.

—S_{HF}: the slope of the regression equation for the high-frequency (HF) items. The regression equation for high-frequency items is based on $\log_2(L_T(x))$ and $t_T(x)$.

—S_{LF}: the slope of the regression equation for the low-frequency (LF) items. The regression equation or low-frequency items is based on $L_T(x)$ and $t_T(x)$.

Using these values the total amount of time we can expect to save, or lose, is computed using the following log-linear formula:

$$\text{Expected Benefits} = S_{HF} \cdot \sum_{\text{HF items}} f(x)^* [\log_2(L_T(x)) - \log_2(L_S(x))] + S_{LF} \cdot \sum_{\text{LF items}} f(x)^* [L_T(x) - L_S(x)]$$

The first part of this equation provides an estimate of the amount of time users will save when selecting the high-frequency items. The second part provides an estimate of the amount of time users will save (or lose) when selecting low-frequency items. The result, the Expected Benefits, is the average amount of time (in seconds) that users can expect to save when selecting items from a split menu as compared to a traditional (often alphabetical) menu.

Of course, if the menu has not been implemented, it is impossible to compute the necessary regression equations. Using the experience of the usability studies and the controlled experiment, reasonable values for the slopes of the regression equations will be presented. These results are presented in the section that discusses the accuracy of the expected-benefits formula.

4. TWO IN SITU USABILITY STUDIES

The first step in evaluating split menus was to install them in normal working environments. The goal of these in situ usability studies was to demonstrate the potential of split menus when real users were performing the tasks they normally do. We use the term *in situ usability studies* to describe data collection that takes place in the users' own environment rather than in a usability lab with artificial tasks. If split menus are successful, additional studies can be conducted to provide a better understanding of exactly what happens when users make selections from these pull-down menus.

A program that created split menus for the font menus in MacWrite and Microsoft Word was installed on Macintosh computers at two sites: the Computer Science Department at the University of Maryland and NASA–Goddard Space Flight Center. Data collection software ran for four weeks while users continued using the standard menus provided by their system (Phase One). At the end of four weeks, selection frequencies were analyzed to determine how the menus should be reorganized. This information was used to create split menus for both MacWrite and Microsoft Word. Data collection continued for an additional five weeks (Phase Two). Split menus were used for five weeks to allow users to adjust to the new menu organization. After the split menus were used for five weeks, the software was removed and the data analyzed.

There were six users of the four Macintosh computers at the University of Maryland. All four computers had the same font menus, allowing them to be analyzed as one distributed system with multiple users. The font menus contained 28 items. Eleven users at NASA, each with their own Macintosh and their own font menu, participated in the usability study. NASA font menus had between 6 and 18 items each with an average of 11. Each Macintosh at the NASA site was analyzed separately.

Once the data were collected they were filtered. Any menu item with three or fewer selections was eliminated due to the potentially large variability with such a small sample size. Next, any selection times which were more than twice as long as the next longest selection time for the same menu item were eliminated. A total of 5% of the selections were filtered before the data were analyzed. For Phase Two, only selections of items for which data were available from Phase One were used, and the first 20% of the selections were considered practice and were not used in the analysis.

Results

All data from Phase One were used to reorganize the menus into split menus. The data were then divided into two parts: MacWrite and Microsoft Word selections. A total of 277 menu selections were recorded at the University of Maryland (UMD) site. Twelve fonts which were selected an average of 1.2 times each were filtered before the data were analyzed.

Since each system at NASA had a unique font menu, each was reorganized based on selections on the system alone. A total of 232 menu selections were recorded at the NASA site. Seven fonts which were selected an average of 1.8 times each were filtered before the data were analyzed.

Table I contains the average selection time for individual fonts, with the exception of the NASA data, as well as the mean for UMD—Word, UMD—MacWrite, and NASA—Word. T-tests were used to compare the selection times for each entry in Table I. Split menus resulted in faster mean selection times for each menu. Split menus also resulted in faster selection times for several individual fonts.

At the end of the usability studies, participants were asked which of the two styles of menus they preferred. Of the 13 participants that answered, one preferred the alphabetical menus; nine preferred the split menus; and three expressed no preference. The Freidman test was used to determine the extent to which the subjects ranked the menu organizations in the same order. The results indicate that subjects consistently rated split menus better than the traditional alphabetic menus ($X_r^2 = 4.92$, $p < 0.05$).

Table I. Time (Standard Deviations in Parentheses) Necessary to Select Individual Fonts and the Average Time for All Fonts (in Seconds). Results from statistical tests are included; bold values indicate significant differences.

Menu	Item	Traditional Menus		Split Menus		% Savings	t values	p
UMD-Word	Mean	3.0	(1.77)	2.2	(1.09)	27	t(63)=2.94	**<0.05**
	Courier	1.7	(0.36)	1.4	(0.17)	18	t(16)=1.37	>0.05
	Helvetica	3.3	(1.76)	2.6	(1.21)	21	t(36)=1.22	>0.09
	Times	3.7	(1.93)	1.7	(0.40)	54	t(20)=2.03	**<0.05**
UMD-MacWrite	Mean	3.6	(1.88)	1.5	(0.87)	58	t(81)=6.38	**<0.001**
	Courier	3.0	(1.41)	1.3	(0.76)	57	t(17)=3.21	**<0.005**
	Times	3.8	(1.97)	1.6	(0.90)	58	t(62)=5.62	**<0.001**
NASA	Mean	2.9	(1.28)	2.4	(1.16)	17	t(137)=2.11	**<0.02**

Discussion

As expected, split menus reduced the average selection time for all menus. The results of the usability studies illustrate the practical potential of split menus. Even with users who were familiar with the traditional alphabetical ordering of font menus, split menus resulted in faster selections after limited practice. Additionally, 92% of the participants either had no preference or preferred the split menus.

Individual differences are important and must be accommodated whenever possible. At the University of Maryland site, multiple users shared multiple computers. Users were treated as a group, and average selection frequencies for the group were used to create split menus. This strategy worked well, and selection times were reduced by 27 to 58% depending on the editor being used. At the NASA site, each user had their own Macintosh, and each system was customized based on the individual user. Once again, this strategy proved effective, reducing selection times by an average of 17%. In situations where groups of users work with a standardized subset of the available options, using average selection frequencies for the group may prove effective. However, when individuals working on one system have distinctly different usage patterns, it may be more effective to develop a method of providing a customized split menu for each user or to continue using traditional menus. If users must identify themselves to the system before using it, customizing the menus for each individual is simple. Otherwise traditional menus may remain the most effective interface.

5. CONTROLLED EXPERIMENT

With the success of the usability studies, we proceeded to run a controlled experiment to develop a more precise understanding of how users select items from pull-down menus. The controlled experiment had several purposes. The first was to validate that split menus can result in faster selections. To demonstrate this, alphabetical, frequency-ordered, and split menus were all evaluated in the experiment. Frequency-ordered menus were included to demonstrate that using frequency information alone is not sufficient. The second purpose was to evaluate our cognitive model for pull-down menu selection and the expected-benefits formula. Since the benefits of split menus depend on how frequently each item is selected and where each item is located in the original (traditional) menu organization, three frequency distributions were explored. This resulted in a total of nine menu organization–frequency distribution combinations.

Subjects

Thirty-eight frequent computer users were recruited as subjects from the University of Maryland campus at College Park. Two subjects were not included in the analysis due to failure to follow instructions. Subjects were primarily graduate students in the Computer Science and Electrical Engineering Departments. Subjects were offered $10.00 to participate in the experiment.

Design and Procedure

The experiment used a within-subject design. Three menu organizations were explored: alphabetic, frequency, and split. Alphabetic menus were organized based on the name of each menu item. Frequency menus were organized with the most frequently selected item at the top and the least frequently selected item at the bottom. Split menus were arranged based on the preliminary guidelines. Three selection–frequency distributions were explored. These distributions assigned 60% of selections to the three high-frequency items while in the usability studies the two or three high-frequency items represent 70–90% of selections. Distribution One had the frequently selected items near the bottom of the alphabetic menu (Table II). Distribution Two had the frequently selected items near the middle of the alphabetic menu (Table III). Distribution Three had the frequently selected items near the top of the alphabetic menu (Table IV). Tables II–IV describe how frequently each menu item was selected for each of the three menu organizations for Distributions One, Two and Three respectively. Note that the alphabetic menus have all items in the traditional location; frequency-ordered menus have all items in order by selection frequency; and split menus place the frequently selected items at the top of the menu while maintaining the traditional ordering for both the high- and low-frequency items.

Every subject used each of the menu styles with each of the frequency distributions. Thirty-six evenly distributed permutations (using a Latin square design) were used to determine the order in which the menu–distribution combinations were used.

All menus contained 15 items. Menu items were selected from a filtered list of the 1000 most frequently used words in printed English [Thorndike and Lorge 1944]. Words that began with capital letters, contractions, and words that were less than four characters or more than eight characters were eliminated from the list of possible menu items. Every subject used nine different menus which contained a randomly selected set of 15 words from the remaining list. No word appeared in more than one menu for a given

Table II. Frequency of Selection for Each Menu Item for Distribution One in Each of the Three Menu Organizations

Alphabetic		Split		Frequency	
Item	f(x)	Item	f(x)	Item	f(x)
1	2	11	24	11	24
2	4	13	20	13	20
3	0	15	16	15	16
4	8	1	2	4	8
5	2	2	4	8	8
6	2	3	0	12	6
7	4	4	8	2	4
8	8	5	2	7	4
9	0	6	2	1	2
10	2	7	4	5	2
11	24	8	8	6	2
12	6	9	0	10	2
13	20	10	2	14	2
14	2	12	6	3	0
15	16	14	2	9	0

Table III. Frequency of Selection for Each Menu Item for Distribution Two in Each of the Three Menu Organizations

Alphabetic		Split		Frequency	
Item	f(x)	Item	f(x)	Item	f(x)
1	2	6	24	6	24
2	4	8	20	8	20
3	0	10	16	10	16
4	8	1	2	4	8
5	2	2	4	13	8
6	24	3	0	12	6
7	4	4	8	2	4
8	20	5	2	7	4
9	0	7	4	1	2
10	16	9	0	5	2
11	2	11	2	11	2
12	6	12	6	14	2
13	8	13	8	15	2
14	2	14	2	3	0
15	2	15	2	9	0

Table IV. Frequency of Selection for Each Menu Item for Distribution Three in Each of the Three Menu Organizations

Alphabetic		Split		Frequency	
Item	f(x)	Item	f(x)	Item	f(x)
1	24	1	24	1	24
2	4	3	20	3	20
3	20	5	16	5	16
4	8	2	4	4	8
5	16	4	8	8	8
6	2	6	2	12	6
7	4	7	4	2	4
8	8	8	8	7	4
9	0	9	0	6	2
10	2	10	2	10	2
11	2	11	2	11	2
12	6	12	6	14	2
13	0	13	0	15	2
14	2	14	2	9	0
15	2	15	2	13	0

subject. The order of the selections was randomized for each subject and each menu. Additionally the order of selections was controlled to create four balanced blocks. The first, second, third, and fourth block of 25 selections each accurately represented the overall selection frequencies for the distribution.

Subjects used a Macintosh computer to perform the menu selections. Subjects were instructed that a word would appear on the screen and that they were to select the same word from the single pull-down menu which was available. Subjects were instructed to select the item as rapidly as possible while maintaining high accuracy since they would be required to repeat selections until they were correct. Subjects took a short break between each of the nine menus to prevent fatigue. When a subject began the experiment the type of menu they would be using (alphabetic, frequency, or split) was presented on the screen. The menu item to be selected was then presented on the screen. Subjects had to move the cursor to the pull-down menu at the top of the screen, click and hold the mouse button (this is when we began recording selection times), drag the cursor to the correct item, and release the mouse button to select the highlighted item. If an error was made the subject was instructed to try again. Once the correct item was selected, there was a brief pause, and the next item to be selected was presented to the subject. Subjects made 100 selections with each of the nine menus. Selection times and the number of errors were recorded for every selection. Once they completed the 100 selections for a given menu subjects took a short break before continuing with the next menu. When the subject completed all nine menus, they were asked to rank the menu organizations in order of preference (1 = most preferred menu, 3 = least preferred menu).

Hypotheses

We expected different results for each of the three frequency distributions. Distribution One has the frequently selected items near the bottom of the alphabetical menu. Split menus should be faster than both alphabetic and frequency-ordered menus for this frequency distribution. Frequency-ordered menus should provide small benefits when compared to alphabetic menus for Distribution One. Split menus should still be faster than alphabetic menus for Distribution Two which has the frequently selected items near the middle of the alphabetic menu, but the benefits will be smaller. Frequency-ordered menus should provide even smaller benefits when compared to alphabetic menus for Distribution Two. Distribution Three, which has the frequently selected items near the top of the alphabetic menu, should result in no differences between the alphabetic and split menus. Frequency-ordered menus should result in a small negative impact when compared to alphabetic menus for Distribution Three. Overall, we expected users to prefer split menus due to the ease of accessing the frequently selected items and the alphabetical ordering which is useful when selecting low-frequency items. Due to the difficulty of understanding the organization, users are expected to rate the frequency-ordered menu as the worst of the three. Although the frequency-ordered menu may be faster than the alphabetic menu for some frequency distributions, the apparent random organization of the frequency-ordered menu will lead to lower preference rankings.

Results

Mean preference rankings are presented in Table V. The Freidman test was used to determine the extent to which the subjects ranked the menu organizations in the same order. The results indicate that subjects consistently rated split menus the best, alphabetic second best, and frequency-ordered menus worst ($X_r^2 = 25.68$, $p < 0.001$).

Since performance at both the first exposure to the menu and after limited practice are important, mean selection times for the first and last blocks are presented in Table VI. Two 3×3 ANOVAs with repeated measures were performed for the first and last blocks separately. The ANOVA for the first block showed significant main effects for menu organization and frequency distribution ($F(2, 70) = 11.39$, $p < 0.001$; $F(2, 70) = 18.05$, $p < 0.001$ respectively). There was also a significant interaction between menu organization and frequency distribution ($F(4, 140) = 4.17$, $p < 0.005$).

The ANOVA for the last block showed significant main effects for menu organization and frequency distribution ($F(2, 70) = 20.87$, $p < 0.001$; $F(2, 70) = 17.16$, $p < 0.001$ respectively). There was also a significant interaction between menu organization and frequency distribution ($F(4, 140) = 9.57$, $p < 0.001$). Contrast matrices were used to perform individual comparisons between each of the three menu organizations for the first and last blocks for each distribution. We felt it was critical to explicitly compare the various menu organizations for each frequency distribution. Since multiple comparisons were being made, the Bonferroni technique was used to set the overall significance level at $p < 0.05$. Although this technique is usually reserved for post hoc situations, we wanted to be conservative in reporting significant differences. Significant differences are summarized in Table VII.

For Distribution One, the individual comparisons indicated that split menus were significantly faster than the alphabetic menu during both the first and last blocks. Frequency menus were also faster than alphabetic menus during the last block (see Figure 3). The individual comparisons indicated that there were no significant differences for Distribution Two. Although split menus were faster than alphabetic menus during both the first and last blocks the differences were not significant (see Figure 4).

For Distribution Three, alphabetic menus were significantly faster than frequency menus during both the first and last blocks, and split menus were significantly faster than frequency menus during the last block (see Figure 5). There were no significant differences between alphabetic and split menus.

Mean error rates for the first and last blocks are presented in Table VIII. Two 3×3 ANOVAs with repeated measures were performed for the first and last blocks separately. The ANOVA for the first block showed no significant main effects or interactions. The ANOVA for the last block also showed no significant main effects or interactions. Contrast matrices were used to make

Table V. Mean Preference Ranking for Alphabetic, Split, and Frequency Menu Organizations (Standard Deviations in Parentheses), 1 = Best, 3 = Worst

	Alphabetic	Split	Frequency
Mean	2.00	1.40	2.60
SD	(0.60)	(0.73)	(0.41)

Table VI. Mean Selection Times in Seconds (Standard Deviations in Parentheses)

	Block					
	First			Last		
	Menu Organization			Menu Organization		
Distribution	Alphabetic	Split	Frequency	Alphabetic	Split	Frequency
One	1 80	1 64	1 76	1 67	1 43	1 52
	(0 44)	(0 41)	(0 50)	(0 44)	(0 34)	(0 44)
Two	1 79	1 67	1 84	1 66	1 52	1 55
	(0 45)	(0 46)	(0 55)	(0 48)	(0 40)	(0 44)
Three	1 47	1 56	1 72	1 38	1 39	1 55
	(0 36)	(0 40)	(0 41)	(0 32)	(0 37)	(0 42)

Table VII. Summary of Significant Difference from Post Hoc Tests

Distribution	**Block**	**Significant differences**
1	First	Split < Alphabetic
1	Last	Split < Alphabetic
1	Last	Frequency < Alphabetic
2		None
3	First	Alphabetic < Frequency
3	Last	Alphabetic < Frequency
3	Last	Split < Frequency

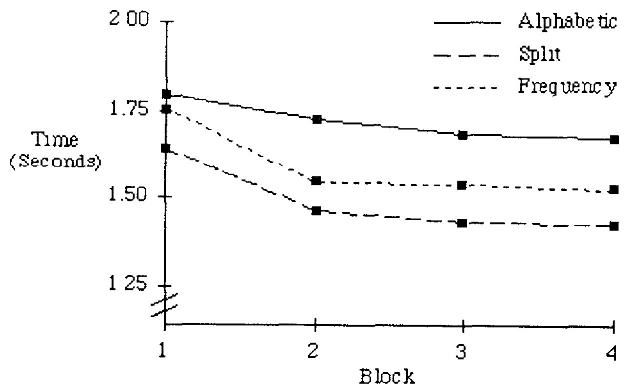

Fig. 3. Average selection time vs. block for three menu organizations for Distribution One.

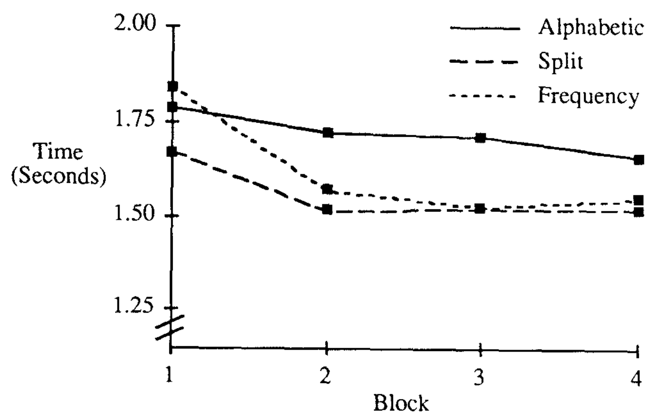

Fig. 4. Average selection times vs. block for three menu organizations for Distribution Two.

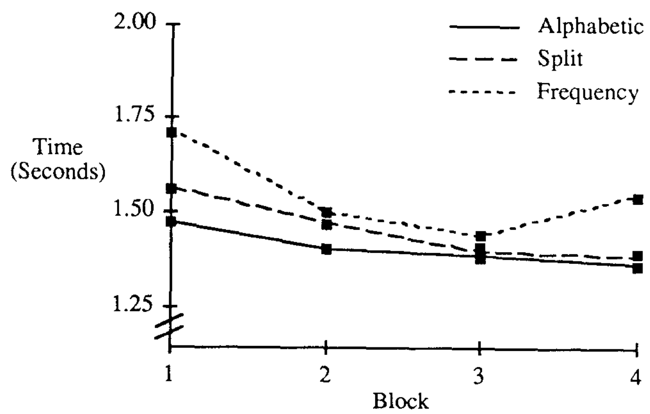

Fig. 5. Average selection time vs. block for three menu organizations for Distribution Three.

Table VIII. Mean Errors per Selection (Standard Deviations in Parentheses)

	Block					
	First			Last		
	Menu Organization			Menu Organization		
Distribution	Alphabetic	Split	Frequency	Alphabetic	Split	Frequency
One	0 012 (0 025)	0 006 (0 014)	0 006 (0.014)	0 014 (0 027)	0 006 (0 014)	0 006 (0 017)
Two	0 012 (0 023)	0 010 (0 020)	0 011 (0 025)	0 016 (0 032)	0 014 (0 032)	0.010 (0 020)
Three	0 007 (0 018)	0 009 (0 024)	0.003 (0 011)	0 007 (0 018)	0 011 (0 021)	0 007 (0 015)

individual comparisons similar to those done for the selection time data. No significant differences were found.

Accuracy of the Expected-Benefits Formula

Assuming the menu would normally be organized alphabetically, the benefits that can be expected by switching from an alphabetical menu to a split menu were calculated. The following computations used the average selection time for each menu item and combined the three alphabetic menus to generate the two necessary regression equations. Expected benefits were then computed for each item in each of the three split menus, as well as for the three distributions. These values were compared to the actual time savings.

Four additional models, different from our log-linear model, for describing the savings users could expect were evaluated. Each model used different combinations of linear and logarithmic equations to describe user performance:

(1) A single linear equation modeled all menu items.
(2) One linear equation modeled the low-frequency menu items, and a second linear equation modeled the high-frequency menu items.
(3) A single logarithmic equation modeled all menu items.
(4) One logarithmic equation modeled the low-frequency menu items, and a second logarithmic equation modeled the high-frequency menu items.

Although statistical comparisons between these models and our log-linear model indicated that no single model described selection times more accurately than any other model, we felt that our theoretical basis for selecting a log-linear model was sound. It is possible that one of these other models may describe selection times as well, or better than, our log-linear model. However, we believe that the distinction between high- and low-frequency menu items is important. As a result, our combined log-linear model was evaluated.

The regression equation, based on the logarithm of the position in the menu, for the high-frequency items from the three alphabetic menus is: $T = 0.199^* \log_2(L_T(x)) + 0.948$ with $r^2 = 0.96$, $p < 0.001$. This supports our conjecture that the logarithmic model describes performance for the high-frequency items.

The regression equation, based on the position in the menu, for the low-frequency items from the three alphabetic menus is: $T = 0.042^* L_T(x) + 1.339$ with $r^2 = 0.76$, $p < 0.001$. Although the correlation is lower than for the high-frequency items, as was expected, it is still relatively high. This supports our second conjecture that the linear model describes performance for the low-frequency items. Although our initial conjectures are confirmed by these results, the correlation between the predicted savings and the actual savings is far more important. An analysis of these results follows.

Using the slopes from the two regression equations ($S_{\text{HF}} = 0.199$ and $S_{\text{LF}} = 0.042$) the expected benefits can be computed for each individual menu item, as well as the three distributions as a whole. The expected benefit for an individual menu item is weighted by how frequently the item is selected. Figure 6 presents the expected versus actual time saved for individual menu

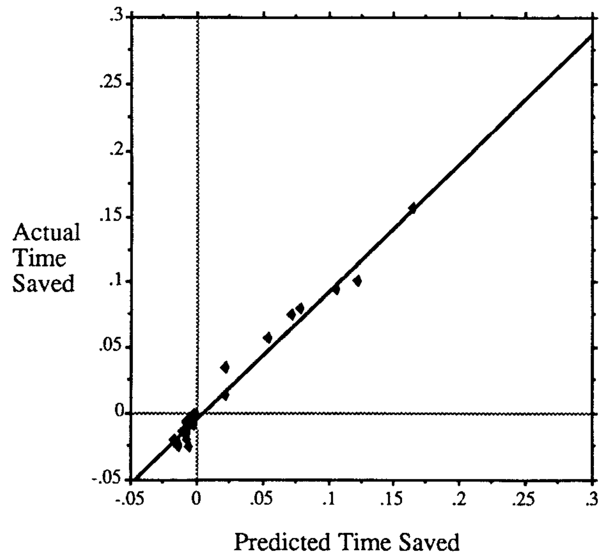

Fig. 6. Predicted vs. actual time saved (in seconds) for individual items.

items. The regression equation for individual items is Actual = 0.97* Predicted + 0.00 and $r^2 = 0.98$, $p < 0.001$. This is very close to the perfect equation of Actual = 1.00* Predicted + 0.00. Similarly, Figure 7 presents the overall results for the three distributions. The regression equation is Actual = 0.88* Predicted − 0.03 and $r^2 = 0.97$, $p < 0.09$. This divergence from the "perfect" equation is probably due to the small but consistent overestimate of the benefits for the high-frequency items and small but consistent underestimate of the negative effects of split menus on low-frequency items.

If a menu has not been implemented, but expected frequency data are available, it may still be possible to estimate the benefits of split menus. Based on the limited data from the usability studies and the controlled experiment, the slope of the high-frequency items can be estimated to be 0.20, and the slope of the low-frequency items can be estimated as 0.06. Using these values for S_{HF} and S_{LF} respectively should provide conservative estimates of the benefits that can be expected by using split menus. Additional data must be collected before more accurate default values can be provided.

Discussion

As expected, there were no significant differences in the error rates. Subjects made between 0.3 and 1.6% errors depending on the menu organization, frequency distribution, and block.

Although users do become familiar with menu organizations, split menus would likely represent a small percentage of all menus on a computer system. Additionally, different users may have different split menus depending on their particular usage patterns. Users who occasionally use another person's

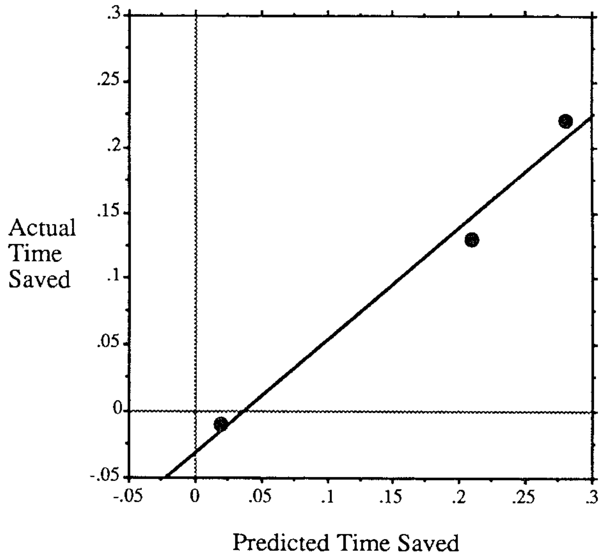

Fig. 7. Predicted vs. actual time saved (in seconds) for three distributions.

computer, or rarely use a split menu, must still be able to make selections rapidly. Of course, performance after practice is important for regular users. Therefore, the results for both the first and last blocks were analyzed.

The results for Distribution One indicated that split menus were significantly faster than alphabetic menus in both the first and last blocks. Frequency-ordered menus were significantly faster than alphabetic menus by the last block. These results indicate that using the frequency of selection resulted in a significant improvement over a purely alphabetical organization for this distribution. Combining the traditional organization with selection frequencies, as is done in split menus, appears to result in even larger benefits. On average, subjects took approximately 0.25 seconds, or 17%, longer per selection when using the alphabetic menus than they did when using split menus.

Distribution Two was an intermediate distribution. It was selected to provide a middle point between Distribution One (most favorable for split menus) and Distribution Three (least favorable for split menus). There were no significant differences between the three menu organizations for this distribution. However, if you compare the results with those of Distributions One and Three it becomes clear that the results are what should be expected. Alphabetic menus took an average of 0.14 seconds, or 9%, longer than split menus for this distribution.

Distribution Three was chosen to represent one of the least favorable situations for split menus. Although distributions where all items are selected with near-equal frequencies would be less favorable for split menus, the preliminary guidelines would not have recommended split menus in that situation. Alphabetic menus were significantly faster than frequency menus in both the first and last blocks. Split menus were significantly faster than frequency menus by the last block. There were no significant differences between alphabetic and split menus. These results indicate that considering only the selection frequency results in slower selections for this distribution. However, if selection frequencies and the traditional organization are considered there is no additional cost over using the traditional organization alone. Interestingly, the proposed guideline presented in the next section would not have suggested using split means for this frequency distribution.

Users preferred the split menus over both the alphabetic and frequency menus. Even though alphabetic menus were slower than frequency menus for two of the three frequency distributions, users did not like the "random" appearance of the frequency-ordered menus. These results provide support for the use of split menus. Regardless of the distribution used, split menus were always as fast or faster than alphabetic menus. Even in the least favorable situation, where split menus may be expected to be slower than alphabetic menus, selection times were equal. Additionally, users preferred split menus, suggesting that even in situations where they are not very beneficial, users may prefer menu organizations that place the frequently selected items where they are easier to access.

Although the advantage was just a few tenths of a second in our experiment, this speedup could make a substantial difference in high-volume (e.g., airline reservation) or stressful (e.g., air traffic control) applications. This advantage will increase with menu length and with more skewed distributions.

6. PROPOSED GUIDELINE

Preliminary guidelines have been presented to assist designers and individual users in deciding if split menus should be used and how to organize items within the menu. Given the traditional menu organization and how frequently each item is selected, these guidelines could be automated, and a split menu could be proposed if it would be beneficial. Users could then view the traditional and split menus to decide which they prefer to use. This section presents a revised guideline, based on the expected-benefits formula, which should provide more helpful advice to designers. The following guideline uses the expected-benefits formula to determine the number of items to be placed in the high-frequency section. This guideline maintains the requirement that the items in the high-frequency section be selected more frequently than those in the low-frequency section.

$EB(x)$ is the expected benefit from moving the x most frequently selected items to the top of the split menu, $EB(0) = 0$. If multiple items are selected with the same frequency, the item closest to the bottom of the menu should be moved to the top of the split menu first. The requirement that all items in the high-frequency section be selected more frequently than items in the low-frequency section is easily maintained using this process.

ΔEB_{1-2} is the difference between $EB(1)$ and $EB(2)$. In other words, ΔEB_{1-2} is the additional improvement that can be expected if two items are moved to the top of the menu instead of one.

ε is the minimum improvement that is necessary before additional items will be moved to the top of the split menu. Therefore, ΔEB_{1-2} must be greater than or equal to ε ($\Delta EB_{1-2} \geq \varepsilon$) before two items will be moved to the top of the menu.

To determine the number of items to move to the high-frequency section of the split menu begin with ΔEB_{0-1}. If $\Delta EB_{0-1} \geq \varepsilon$ then move at least one item to the high-frequency section. If $\Delta EB_{0-1} \geq \varepsilon$ and $\Delta EB_{1-2} \geq \varepsilon$ then move the two most frequently selected items to the top of the split menu. This process repeats until $\Delta EB_{x-y} < \varepsilon$. This indicates that X items would be moved to the top of the split menu.

Of course, to calculate $EB(x)$ the slope of regression equations for both the high- and low-frequency items must be known. Since the high-frequency items have not been specified, and multiple items must be specified for each section of the menu before the regression equations can be computed, we must estimate both slopes. As discussed near the end of Section 5, the slope for the high-frequency items can be estimated to be 0.20, and the slope for the low-frequency items can be estimated to be 0.06. Using these values, and an ε of 0.05 seconds, the three frequency distributions from the controlled experiment were analyzed. The results appear in Table IX. The three most frequently selected items would be moved to the top of the split menu for Distributions One and Two. Split menus would not be recommended for Distribution Three. These recommendations match the results of the experiment well. Of course, additional research may refine the values for the slopes for both the high- and low-frequency items as well as ε (a lower value for ε, such as 0.03 would probably work well for creating split menus).

7. FUTURE DIRECTIONS

Several important questions must still be answered. When a system is first being used selection frequencies can be estimated until sufficient data can be collected. The question is: how much data is necessary before it is sufficient? A variant of this question becomes important once split menus are being used. Usage patterns may change over time, so the question becomes: how do we identify changing selection patterns which lead to suggesting changes to the split menus? These two questions are essentially the same: how do we identify when usage patterns have stabilized? Since selection frequencies will change with time, it is clear that we must place more emphasis on recent selections. This could be done by weighing recent selections more heavily [Greenberg and Witten 1985], or by using only the most recent selections. Strategies must be developed that can monitor selection frequencies and visually alert users to consider reordering a menu.

An interesting alternative to the current version of split menus would be to leave the high-frequency items in the bottom of the split menu and duplicate

Table IX. Proposed Guidelines Applied to the Three Frequency Distributions from the Controlled Experiment (Bold Items Represent Items that Would Be Placed in the Top of the Split Menu)

# HF items (x)	Expected Benefits Distribution One	Distribution Two	Distribution Three
1	**0.10**	**0.07**	-0.05
2	**0.19**	**0.15**	----
3	**0.25**	**0.20**	----
4	0.25	0.21	----
5	----	----	----

them in the high-frequency section. This would ensure that users would locate the high-frequency items regardless of the section of the menu they searched, while maintaining the benefits of placing the items at the top of the menu. However, the results of the controlled experiment indicate that users scan the high-frequency list regardless of the item they are searching for, indicating that this organization would not prove beneficial.

8. CONCLUSIONS

The usability studies demonstrated the potential of split menus in normal work conditions. Even users who had used the alphabetic menus for years saved time when using split menus (between 17 and 58% depending on the usability study site and menu). Twelve of thirteen users either expressed no preference or preferred the split menus.

The controlled experiment provided valuable support for the use of split menus and the refined cognitive model for pull-down menu selection. The controlled experiment also confirmed the potential of the expected-benefits formula we proposed. The predictions for individual items closely approximated actual values, while the predictions for the three distributions provided valuable information concerning the expected impact of split menus. Actual benefits for split menus may be larger than in these studies if menus are longer or if the skewness of the frequency distribution is larger, as appears to be the case in many realistic situations.

The results of the controlled experiment not only demonstrated the time savings and higher preference ratings split menus create, but also demonstrated the value of the proposed guideline for creating split menus. The proposed guideline would have suggested the three most frequently selected items be moved to the top of the split menus for Distributions One and Two, and that split menus were not appropriate for Distribution Three. These recommendations match the results of the experiment well.

Split menus provide the benefits of both frequency-ordered and alphabetical menus. Frequently selected items are moved to the top of the menu making them easy to locate and select, while both sections of the menu maintain a traditional (alphabetical) organization allowing users to quickly scan the menu to locate the item of interest. We encourage interface designers to consider incorporating the necessary software to support split menus. Monitoring selection frequencies, and providing a dialog box which allows users to view the menu both alphabetically and as a split menu and to choose which one to use, can result in faster selections and higher user preference ratings.

ACKNOWLEDGMENTS

We would like to thank David Kieras and the anonymous reviewers for their thoughtful comments. Their detailed reviews led to many improvements. We appreciate Richard Potter's efforts in creating the data collection and menu organization software used in the usability studies. We would also like to thank all of the participants in both the usability studies and the controlled experiment.

REFERENCES

APPLE COMPUTER INC. 1987. *Human Interface Guidelines: The Apple Desktop Interface*. Addison-Wesley, Reading, Mass.

BROWN, C. M. 1988. *Human-Computer Interface Design Guidelines*. Ablex, Norwood, N.J.

CARD, S. K. 1982. User perceptual mechanisms in the search of computer command menus. In *Human Factors in Computer Systems Proceedings*. ACM, New York, 190–196.

CALLAHAN, J., HOPKINS, D., WEISER, M., AND SHNEIDERMAN, B. 1988. An empirical comparison of pie versus linear menus. In *Proceedings of Human Factors in Computing Systems '88*. ACM, New York, 95–100.

DEPARTMENT OF DEFENSE. 1991. Military Standard—Human engineering design criteria for military systems, equipment and facilities. MIL-STD-1472D, Dept. of Defense, Washington, D.C.

FISHER, D., YUNGKURTH, E., AND MOSS, S. 1990. Optimal menu hierarchy design: Syntax and semantics. *Hum. Fact. 32*, 6, 665–683.

GREENBERG, S. AND WITTEN, I. H. 1985. Adaptive personalized interfaces: A question of viability. *Behav. Inf. Tech. 4*, 1, 31–45.

GREENBERG, S. 1988. Tool use, reuse, and organization in command-driven interfaces. Ph.D. dissertation, Dept. of Computer Science, The Univ. of Calgary, Alberta, Calgary, 28–33.

LANDAUER, T. K. AND NACHBAR, D. W. 1985. Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width. In *Proceedings of Human Factors in Computing Systems*. ACM, New York, 73–78.

LEE, E. AND MACGREGOR, J. 1985. Minimizing user search time in menu retrieval systems. *Hum. Fact. 27*, 2, 157–162.

MITCHELL, J. AND SHNEIDERMAN, B. 1989. Dynamic versus static menus: An exploratory comparison. *SIGCHI Bull. 20*, 4, 33–37.

NORMAN, K. 1991. *The Psychology of Menu Selection*. Ablex, Norwood, N.J.

OPEN SOFTWARE FOUNDATION. 1990. *OSF/Motif Style Guide*. Prentice-Hall, Englewood Cliffs, N.J.

PAAP, K. R. AND ROSKE-HOFSTRAND, R. J. 1988. Design of menus. In *Handbook of Human-Computer Interaction*. Elsevier Science Publishers, Amsterdam, 205–235.

PAAP, K. R. AND ROSKE-HOFSTRAND, R. J. 1986. The optimal number of menu options per panel. *Hum. Fact. 28*, 4, 377–385.

RUBINSTEIN, R. AND HERSH, H. 1986. *The Human Factor: Designing Computer Systems for People*. Digital Press, Burlington, Mass.

SEARS, A. 1993a. Layout appropriateness: Guiding user interface design with simple task descriptions. Ph.D. dissertation, Computer Science Dept., The Univ. of Maryland, College Park, Md.

SEARS, A. 1993b. Layout appropriateness: A metric for evaluating user inferface widget layout. *IEEE Trans. Softw. Eng. 19*, 7, 707–719.

SHNEIDERMAN, B. 1992. *Designing the User Interface*. 2nd ed. Addison-Wesley, Reading, Mass.

SMITH, S. L. AND MOSIER, J. N. 1986. Guidelines for designing user interface software. Rep. 7 MTR-10090, Mitre Corporation, Bedford, Mass.

SOMBERG, B. L. 1987. A comparison of rule-based and positionally constant arrangements of computer menu items. In *Proceedings of CHI and GI '87*. ACM, New York, 255–260.

SUN MICROSYSTEMS, INC. 1990. *OPEN LOOK Graphical User Interface Application Style Guidelines*. Addison-Wesley, Reading, Mass.

THORNDIKE, E. L. AND LORGE, I. 1944. *The Teachers' Word Book of* 30,000 *Words*. Columbia University Press, New York.

WALKER, N. AND SMELCER, J. 1990. A comparison of selection times from walking and pull-down menus. In *Proceedings of Human Factors in Computing Systems '90*. ACM, New York, 221–225.

Received December 1992; revised January 1994; accepted January 1994