Using predictors to partition menu selection times

JOCHEN MÜSSELER

Max Planck-Institut für psychologische Forschung, Leopoldstraße 24, D-80802 München, Federal Republic of Germany; e-mail: muessele@mpipf-muenchen.mpg.de

Behaviour & Information Technology
Publisher: Taylor & Francis Group
Issue: Volume 13, Number 6 (1994)
Pages: 362 – 372
URL: http://pdfserve.informaworld.com/862002_777306414_773190235.pdf
DOI: 10.1080/01449299408914617

Abstract. Selection times of drop-down menus are in many ways influenced by cognitive and motor processes of the user and by design variables of the menu. Since the number of these variables is too large, the contribution of individual variables to selection time cannot be assessed by using factorial designs. Multiple regression is introduced to solve this problem. The technique uses selection times as criterions and a set of general menu characteristics as predictors. The non-standardized slopes β report the increase (or decrease) in selection time which can be assessed for each predictor. In a first experiment, the validity of the technique was demonstrated replicating various well-known effects in a mouse-driven editor. For example, the selection times increased with the number of subordinate menu items or atypical items. Further, due to motor components of the mouse movement, selection times depended on the spatial position of an item within the menu. In a second experiment, mouse selection was replaced by key selection to stress cognitive processes contributing to response times. The technique yielded results that were sensitive to this variation. Limitations of the technique are discussed.

1. Introduction

Menu selection systems are widely accepted in modern computer systems. Menus can eliminate training and memorization of complex command sequences, and this has proved to be particularly advantageous for unpractised or occasional users. These advantages are due to the recognition principle and the prestructured dialogue sequence that characterize menu selection systems.

Special variants of menu selection systems are so-called *pull-down* menus, in which the submenu can be made visible by clicking a corresponding menu item with the computer mouse, and so-called *drop-down* menus, in which the menu drops down as soon as the mouse cursor touches the respective item in the menu bar. Everyday interaction with pull-down or drop-down menus is characterized by how frequently the menu bar has to be scanned and how long the selection time is for a particular item. The selection time is the sum of the sequence of predefined dialogue steps, that is, selection time within the menu bar (main menu) and that for the subsequently selected item in the submenu. If the dialogue structure is hierarchical and contains several submenu levels, selection time can be partitioned correspondingly.

Hence, selection times and the number of scanned submenus are important parameters for judging the user performance with such an interface (Lee and MacGregor 1985). Both variables are influenced by the cognitive and motor processes of the user. They are decisively dependent on the menu design and the variables it contains, for example, the depth/breadth of the tree structure, the number of items in the menu, and the typicality of an item for its superordinate menu category (for a review, see Shneiderman 1986a, b). The selection time for each item is determined by several of these variables. For more complex selection systems, as found in many applications software packages, it is possible to list a great number of potential variables. However, the number of these variables is too large to allow for experimentation with complete factorial designs and thus to evaluate the total impact of the various variables in a complete menu that is already in use.

The present article proposes an evaluation technique that draws on *multiple regression* (for a review, see, e.g., Knight 1984). Multiple regression can provide information on how selection times (the criterion variables) are determined *quantitatively* by a set of general menu characteristics (the predictor variables). The measure for the relative impact of the predictors on the criterions is the respective slope β. In its non-standardized form, β reports the increase (or decrease) in selection time in units of the predictor variables. In contrast to traditional applications of multiple regression analysis, we are less concerned with detecting the 'best' fit with the lowest possible number of predictor variables, as performed in stepwise regression, than with determining and quantitatively partitioning the impact of individual variables on selection time (or the number of submenus scanned). In this context, quality of prediction is the prerequisite of the technique. A further difference between this more traditional applications of multiple regression analysis is that the predictors are set before the experiment is run; only criterion variables are recorded empirically.

In recent years, the technique applied here has become increasingly important in cognitive psychology, for example, in partitioning reading times (see Haberlandt and Graesser 1985, Knight 1984, Lorch and Myers 1990, Rickheit *et al.* 1992). In this field, multiple regression is used to determine the variables underlying speed at which words are read in a text. A typical outcome is that the time it takes to read a specific word is determined by its length, its position in the sentence, its semantic content, etc. A menu selection system can be treated in a similar way. Every item can be classified on a number of variables, for example, its position in the menu hierarchy, its length, or the number of available response alternatives. Multiple regression provides information on the individual contribution of these predictors to selection time. The partitioning thus gained can then be used to evaluate the user interface.

The goal of the first experiment, in which menu selection was performed with a mouse, was to use multiple regression to partition mouse selection times on the basis of different predictors. This should also contribute to the acceptance of this technique. An editor interface was chosen to demonstrate the technique, although the results should also be generalizable for other menu-driven software (e.g., graphic programs, CAD systems, databases). The second experiment tested whether the technique was sensitive to the exclusion of individual variables. Item selection with a mouse was replaced by selection with key combinations. In our context, this variation should reduce the importance of motor variables elicited by the mouse in favor of more cognitive components.

2. General procedure

The present studies used a drop-down menu as can be found in a 'normal' editor[1]. The menu contained two hierarchical levels: the *menu bar* (main menu) displayed continuously at the top of the screen, and six *submenus*, which could be activated with the mouse. The number of items in the submenus varied between two and seven. Figure 1 presents all 27 items in the sequence found in the menu.

As mentioned above, each item is characterized by a set of menu variables. The first step was to specify each item on predictors. Then the independence (collinearity) of the predictors was tested to confirm whether the coefficient's β can be interpreted at all.

2.1. *Predictor variables*

Three possible groups of predictor variables can be

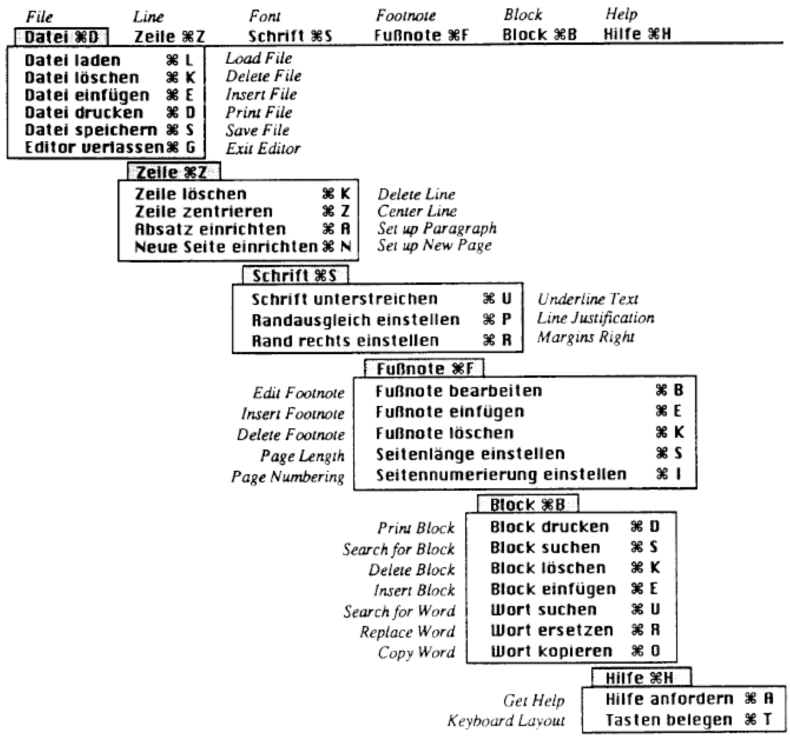

Figure 1. The drop-down menu. Moving the mouse cursor in the menu bar (first line) displays the subordinate submenu. English translations printed in italics.

Table 1. Coding scheme for the editor items (excerpt from the data matrix).

Item / Predictor variable	PM	PS	NS	AF	OF	SMG	OC_1	OC_2[1]	IL	LMB	WSM
Datei laden (Load File)	1	1	6	1	2	2	1	1	11	5	16
Zeile löschen (Delete Line)	2	1	4	2	2	3	1	2	13	5	21
Schrift unterstreichen (Underline Text)	3	1	3	...							
...											

[1]Additional predictor in experiment 2.

discriminated: (1) a group of variables determining the external layout of the menu bar and the submenu; (2) potential cognitive structuring within one menu and among menus; and (3) features on the individual item level. The analysis of the menu used in the present studies focuses on variables that can be assumed to influence selection time. These variables will be specified in the following:

(1) External layout includes the position of the item in the menu bar (PM), that is, the arrangement in the drop-down menu from left to right. In addition, each item has a particular position in the submenu (PS). Differences in mouse selection times can be anticipated solely on the basis of this variation in spatial positions (Walker, Smelcer, and Nilsen 1991). Additionally, studies of menu selections using joysticks and touchscreens have shown that positioning time depends lawfully on target distance and target size (Card, English, and Burr 1978, Landauer and Nachbar 1985, Radwin, Vanderheiden, and Lin 1990; cf. Fitts' law, Fitts 1954).

Another variable is the total number of items in the submenu in which an item is embedded (number of submenu items, NS). Although this variable determines the external layout of the menu it has clearly cognitive components. As reaction times are longer when several response alternatives are available, this should also affect menu-driven selection (see the Hick–Hyman law, Hyman 1953, with reference to human–computer interaction, see Arend *et al.* 1987, Kiger 1984, Landauer and Nachbar 1985, Shinar and Stern 1987).

(2) As the submenu's 'cognitive structuring' includes fanning (cf. Anderson 1983: 107–121), it is used here as a predictor. Its effect on human computer interaction has been demonstrated recently (Arend *et al.* 1987, Heydemann 1989: 82–104). In commands such as '*delete* file', '*delete* footnote', and '*delete* line', the repetition of the verb in our menu elicits an action fanning between submenus (AF). In contrast, commands such as 'load *file*', 'delete *file*', and 'print *file*' reveal an object fanning within one submenu (OF). Both fanning variables were assessed dichotomously, that is, we assessed whether an item was fanned or not.

A further variable in this context is the typicality of the item. When an item is typical for its submenu, selection tasks are performed more efficiently than when it is atypical (Verwey *et al.* 1988, but see Hayhoe 1990, for the problems with categorization). Software programs frequently contain items that directly refer to the entry in the menu bar (e.g., 'load file' in the submenu 'file'), while other commands cannot be classified so well (e.g., 'exit editor' in the same submenu). If the menu bar does not indicate the presence of a submenu item, the item is interpreted as being 'covert' (e.g., 'search for word' in the menu 'block'), in contrast to 'overt' (e.g., 'delete file' in the menu 'file'). Both codings lead to the variable 'overt vs. covert' (OC_1).

Another form of typicality results from the possible ways of grouping items. It can be assumed that the number of groupings within the submenu (SMG, e.g., four 'block' items and three 'word' items are two groupings) also influences selection time; the more segmented a submenu, the more difficult it should be to select an item.

(3) On the individual item level, reading time variables should have the strongest impact. It has been demonstrated, for example, that the time it takes to read a word within a text correlates with the number of letters it contains (Rickheit *et al.* 1992). In our user interface, we expect to find effects of variables such as item length in letters (IL), the corresponding length of the menu bar item (LMB), and the width of the submenu (WSM). The latter variable corresponds to the maximum length of an item (IL) within a drop-down menu.

Each item has a specific value on each variable. The results of the coding scheme are presented in table 1.

2.2. *Collinearity between predictor variables*

An unequivocal interpretation of regression coefficients ideally requires independence between predictor variables (Knight 1984). Otherwise it is impossible to estimate what predictor variables contribute to the criterion variable. High correlations (of about 0.8 or more) between two variables are viewed as a sign of collinearity problems. Accordingly, the bivariate correlations between predictor variables are shown in table 2. Three out of the total of 55 correlations were extremely high (between 0.727 and 0.888); these were between variables on the item level (IL, LMB, WSM).

One possible way of encountering the collinearity problem was to drop variables from the regression analysis (Cohen and Cohen 1975, Knight 1984). From the three problematic variables, the length of individual items (IL) was retained, as it was most likely to have an impact on selection times. Hence, in view of the high proportion of unexplained

Table 2. Intercorrelations between predictors.

		PM	PS	NS	AF	OF	SMG	OC_1	OC_2[1]	IL	LMB	WSM
External layout												
Position in menu bar	PM	1.000										
Position in submenu	PS	0.004	1.000									
Number of submenu items	NS	0.009	0.459	1.000								
Cognitive menu structuring												
Horizontal fanning	AF	− 0.059	− 0.128	0.097	1.000							
Vertical fanning	OF	− 0.166	0.139	0.501	0.402	1.000						
Submenu grouping	SMG	− 0.331	− 0.139	− 0.303	− 0.193	0.174	1.000					
Overt vs. covert	OC_1	0.150	0.578	− 0.173	− 0.196	− 0.079	0.079	1.000				
Overt vs. covert	OC_2	0.076	0.413	0.266	0.182	0.104	− 0.104	0.301	1.000			
Item level												
Item length	IL	− 0.056	0.038	− 0.507	0.196	− 0.061	0.061	0.428	− 0.020	1.000		
Item length in menu bar	LMB	0.113	− 0.169	− 0.369	0.349	0.042	− 0.271	0.122	0.005	0.727	1.000	
Item length in submenu	WSM	− 0.097	− 0.210	− 0.457	0.289	0.139	0.137	0.102	− 0.050	0.729	0.888	1.000

[1]Additional predictor in experiment 2.

variance between the remaining variables, collinearity problems could be disregarded.

2.3. *Pairwise interactions*

Apart from testing main effects for individual predictor variables, regression analysis also permits us to test interactions among variables (Kerlinger and Pedhazur 1973: 415). The presence of an interaction can be judged by calculating the product of two predictor variables and entering this into the regression analysis as an additional variable.

Even with just eight predictor variables of experiment 1, 28 first-order interactions are possible—in principle, even higher-order interactions are testable—and thus a statistical problem arises from the number of analyses to be calculated. Therefore, the following experiments initially tested only those interactions for which directional hypotheses could be formulated.

3. Experiment 1

Experiment 1 assessed three criterion variables. First, selection times within the menu bar and, second, selection times within submenus. Only those selection times where subjects had directly selected the item were included in the analyses; otherwise, time profiles would have been distorted by previously scanned submenus and attendant uncontrollable effects. To take into account multiple scanning, the number of scanned submenus was entered into the analyses as the third criterion variable.

Predictions for the predictor variables have already been given in the preceding section. In addition, an interaction was anticipated between the 'number of submenu items' (NS) and the 'overt vs. covert' (OC_1) variable. In general, selection times should increase with the number of alternative items presented. However, covert items are positioned at the bottom of the menu, this parallels real-life user interfaces, where this position is frequently reserved for nontypical items. They are difficult but nevertheless salient items within one submenu. Thus, it is predicted that the selection of covert items is not affected by NS and, therefore, an interaction between the two variables was anticipated.

3.1. *Method*

3.1.1. *Subjects:* Four female and six male graduate college students with a mean age of 29.9 years took part in the experiment. As we are interested in effects of everyday interaction with the menu all subjects had extensive experience in working with various editors, so that it was not necessary to explain drop-down menus as such and the function of the individual items within a real editor.

3.1.2. *Apparatus:* The experiment was carried out on a PC with a monochrome display. The program for generating menus was based on GEM[2] routines, so that a menu configuration such as that in figure 1 was displayed. Each mouse movement along the menu bar and the activation of the corresponding submenu that it triggered was registered. A logfile was set up for the selections (selected submenu or clicked item) and their times. Selection times were recorded in milliseconds.

3.1.3. *Procedure:* Subjects first had the opportunity to get familiar with the menu for 10 min; then they were informed about the task. They were instructed to search in the menu for an item, constantly displayed at the bottom center of the screen, as quickly and accurately as possible and to click a mouse key in order to select it. This identity matching task was chosen to stress mere selection times of skilled users. If they clicked the wrong item, there was a short auditory signal (1000 Hz for 50 ms) and a correction was requested. If selection was correct, positive written feedback was given and the mouse cursor had to be reset at the center of the display. A 2 s pause was then followed by a longer auditory signal (440 Hz for 200 ms) and a new item.

The experiment was organized into four blocks. In each block, all 27 items were displayed in a random sequence. A short break could be taken between blocks. To minimize learning effects, only data from the third and fourth blocks were analysed.

3.2. *Results and discussion*

3.2.1. *Testing linearity:* The multiple regression analyses calculated here required a linear relationship between each of the predictor variables and the criterion variables. This assumption could only be tested meaningfully for the four nondichotomous predictors (PM, PS, NS, IL). Orthogonal contrasts were used to perform trend analyses on these predictors. As linear, quadratic, and cubic trends were tested for each of the three criterion variables, this resulted in a total of 36 trend analyses. Critical quadratic or cubic trends were found in five analyses only. 'Item length' (IL) showed both nonlinear trends to selection time on the menu bar and to the number of scanned submenus; the 'number of submenu items' (NS) showed a quadratic trend to the number of scanned submenus. However, as a linear trend could also be confirmed in each of these cases, the assumption of linearity could not be dismissed.

3.2.2. *Main effects and interactions:* As each subject was confronted with all 27 editor items in a repeated measurements design, the three criterion variables were first entered separately into multiple linear regression analyses for each subject, as recommended by Lorch and Myers (1990). In addition, an analysis for the predicted interaction was calculated. Every regression analysis based on 54 observations, thus a total of 1620 observations across all subjects and

Table 3. Summary of multiple regression analyses on selection times for menu bar, submenu, and number of scanned submenus (β = mean slope coefficients, SE = standard error, t = t-value, r_c = criterion correlation).

Variable		Selection time menu bar				Number of scanned items				Selection time submenu			
		β	SE	t	r_c	β	SE	t	r_c	β	SE	t	r_c
External layout													
Position in menu bar	PM	67.76	12.18	5.56**	0.178	0.001	0.065	0.05	−0.086	1.44	11.08	0.13	−0.061
Position in submenu	PS	8.90	40.88	0.69	0.308	−0.025	0.070	−1.14	0.199	125.48	16.34	7.68**	0.554
Number of submenu items	NS	125.71	23.58	5.33**	0.155	0.111	0.185	1.90	−0.180	−25.10	13.19	−1.90	0.312
										(84.85	28.34	2.99)[1]	
Cognitive menu structuring													
Horizontal fanning	AF	122.99	43.20	2.85	0.188	0.053	0.213	0.79	−0.007	33.31	36.70	0.91	0.182
Vertical fanning	OF	−114.58	58.09	−1.97	0.088	−0.090	0.560	−0.51	0.085	299.24	72.09	4.15*	0.361
Submenu grouping	SMG	594.91	92.32	6.44**	0.232	0.558	0.287	6.16**	0.306	−100.79	41.96	−2.40	−0.120
Overt vs. covert	OC_1	426.20	87.18	4.89**	0.582	0.525	0.517	3.21	0.474	−136.44	34.08	−4.00*	0.233
										(191.86	63.28	3.03)[1]	
Item level													
Item length	IL	92.55	15.01	6.17**	0.629	0.072	0.062	3.62*	0.480	15.05	4.87	3.09	0.210
Mean multiple R		0.831 (0.739–0.899)				0.752 (0.552–0.853)				0.788 (0.621–0.892)			
Mean multiple R^2		0.691				0.566				0.621			

[1] including the NS × OC_1-interaction: $\beta = -76.68$, SE = 17.08, $t = -4.49$, $p < 0.01$

* df = 9, $p < 0.01$ (2-tailed)

** df = 9, $p < 0.001$ (2-tailed)

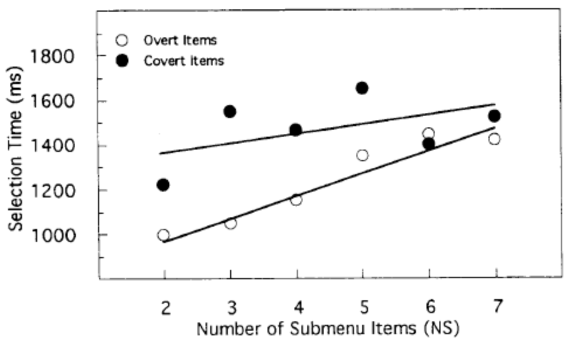

Figure 2. NS × OC_1 interaction. Submenu selection times were summarized across all remaining variables.

criterion variables entered in the analyses. Inferential t-tests were used to check whether the (descriptive) regression coefficients averaged across subjects showed a significant slope or not (cf. Lorch and Myers 1990).[3]

Table 3 presents the results for all three criterion variables. The mean multiple correlations were notably high, ranging from $R = 0.752$ to $R = 0.831$, so that 57% to 70% of the variance, depending on the criterion variable, could be explained by the eight predictors.

The β-slopes were estimations of the selection times or the number of scanned menus in units of the corresponding variables. For example, selection time in the menu bar depended on the item's 'position in the menu bar' (PM). A β-value of approximately 68 ms meant that selection time increased as a multiple of this amount, the further the submenu to be selected was located toward the right in the menu bar, $t = 5.56$, $p < 0.001$. This effect indicated a preference for cognitively scanning the menu bar from left to right.

In contrast, selection time in the menu bar did not depend on the position of the item within the submenu (PS, $t = 0.69$, n.s.); i.e., the spatial position of the item within the submenu is not involved in the decision regarding which item from the menu bar was finally selected. This position only became important when the actual item within the submenu had to be clicked. Here, the selection times rose by 125 ms for each descending position ($t = 7.68$, $p < 0.001$), as could be anticipated owing to motor components of the mouse movement. Conversely, it was assumed that the position of the preselected submenu in the menu bar should not exert any influence on this decision, which was also the case ($t = 0.13$, n.s.).

The impact of the other predictor variables should have more to do with factors influencing cognition. For example, selection in the menu bar was determined by the 'number of submenu items' (NS, $\beta = 125.71$, $t = 5.33$, $p < 0.001$) and by the 'overt vs. covert' variable (OC_1, $\beta = 426.20$, $t = 4.89$, $p < 0.001$).

In selection within the submenu, the main effects of NS and OC_1 were characterized by a seemingly inexplicable relationship between the criterion correlations and the slope coefficients. While every slope coefficient was negative, each correlation was positive. This result indicated the presence of the anticipated interaction effect. As can be seen from figure 2, it took the predicted direction ($\beta = -76.68$, SE = 17.08, $t = -4.49$, $p < 0.01$)[4] and the main effects could be corrected to a clearly positive relationship. An increase in submenu selection time could be seen as a function of the number of items, an effect that was less steep for the covert items. The latter result was anticipated because—as in a normal editor—covert items were located at the bottom of the submenu. This property was relatively independent of the number of submenu items.

Selection time on the menu bar and submenu was not affected by action fanning (AF); however, object-fanned items (OF, e.g., 'load *file*', 'delete *file*', 'save *file*'), interfered within the submenu ($t = 4.15$, $p < 0.01$). When, for example, it was necessary to select 'save file' from several other 'file' items, selection time was 299 ms longer than for items with no object fanning.

The number of 'groupings within a submenu' (SMG) had the following impact on menu bar selection time and the number of previously scanned menus: With increasing item variety within a submenu, selection time slowed down ($t = 6.44$, $p < 0.001$), and the prior scanning of other submenus occurred more frequently ($t = 6.16$, $p < 0.001$).

Item length (IL) showed an almost identical pattern of effects across all three criterion variables. Selection time on the menu bar increased by approximately 93 ms per letter ($t = 6.17$, $p < 0.001$). Although the mean number of previously scanned submenus increased only slightly, this was statistically significant ($t = 3.62$, $p < 0.01$). Third, selection time for items within the submenu increased marginally by 15 ms ($t = 3.09$, $p < 0.05$). The latter value is approximately the same as that found in studies on reading (12 ms, see Rickheit *et al.* 1992).

In sum, the analyses technique employed here shows the expected dependence of mouse selection times on motor and cognitive components. Moreover, the results indicate the point of time in the selection process when the specific variable is important. For example, position in the menu bar (PM), the number of items in the submenu (NS) and the overt vs. covert variable (OC_1) affect the time needed to choose the submenu from the menu bar, whereas position in the submenu (PS) and object fanning (OF) produce significant slopes in choosing the item from the submenu. The latter result could be due to visual scanning of similar and therefore conflicting items.

4. Experiment 2

Experiment 2 differed from experiment 1 in that items were activated with key combinations (for a comparison between these methods, see Card *et al.* 1978, Karat *et al.* 1986). One key was used to make the selection on the menu bar; a second one to select within the submenu. This should exclude, or at least reduce, those pure motor components of selection time that are only due to mouse control, and particularly involve the variables PM (position on the menu bar), PS (position in submenu), and, to some extent, NS (number of submenu items). Just as in normal editor programs, cognitive components were further emphasized by the fact that the appropriate drop-down submenu did not appear when the first key was pressed. This changed the search and reading task into a pure recall task.

Activating the item through key combinations led to a modification of the predictors. The 'overt vs. covert' variable OC_1 in experiment 1 now reported whether the initial letter of the item was identical with the first key that had to be pressed. This corresponded to the selection from the menu bar. For example, in the item 'delete *b*lock' (in German: '*B*lock löschen'), the first key was determined by the initial letter, and was subsequently 'overt' for the block menu (key combination *B* and K). The second key did not agree in this case, so that it could accordingly be considered 'covert'. Therefore, a further variable (OC_2) was introduced that referred to the second key to be pressed and corresponded to selection within the submenu. This variable represents stimulus–response compatibility vs. incompatibility.

4.1. *Method*

4.1.1. *Subjects.* Fifteen graduate college students (7 female, 8 male) with a mean age of 28.1 years participated in the experiment. As in experiment 1 all subjects were skilled users of various editors.

4.1.2. *Apparatus and procedure*: The design was based on experiment 1. Instead of using a mouse, key combinations had to be entered with a German DIN keyboard. Together with the Ctrl-key, which was pressed with the left hand, each item required two keys to be pressed in succession with the right hand. The corresponding key combinations were displayed in the drop-down menu (e.g., 'load file' through Ctrl D and L, see figure 1). In this way, they could be learned during the experimental blocks with the mouse. In case several errors were made and the correct key combination could not be recalled, the drop-down menu could be reactivated with the mouse in order to solve the task.

A practice phase again consisted of three blocks of all 27 items that had to be activated with a mouse. During this phase, subjects were not informed that the key combinations would have to be recalled later on. After the practice phase subjects were instructed to activate items with key combinations and were allowed to use the mouse if necessary. However, this was registered as an error, also the entering of an incorrect key combination. The experiment ended when the learning criterion—less than three errors of both types within one block—was achieved.

Subjects were instructed to respond as quickly and accurately as possible. In addition, they were told to enter the first key even if they were still uncertain about the second key. Mean scores on the last two blocks before attaining the learning criterion were analysed. On average, subjects required 4.9 blocks to achieve criterion.

4.2. *Results and discussion*

4.2.1. *Testing linearity:* As in experiment 1, selection times for key combinations mostly had a linear relationship with the predictor variables. Latencies up to the first key press showed a significant linear and quadratic trend in the NS variable. Latencies up to the second key press had both a linear and a cubic relationship to the IL variable. A threat to the linearity assumption in this criterion variable was a demonstrable quadratic trend to the NS variable ($F\ 1{,}399 = 7.34$, $p < 0.01$), with no significant linear trend. To avoid a violation of assumptions, the NS variable was subjected to a logarithmic transformation. This adjustment led to both a linear trend ($F\ 1{,}399 = 5.00$, $p < 0.05$) and a quadratic trend, ($F\ 1{,}399 = 7.61$, $p < 0.01$).

4.2.2. *Main effects and interactions:* Table 4 presents mean findings from regression analyses. Multiple correlations showed a high coefficient of $R = 0.850$ for the first key and a somewhat lower $R = 0.716$ for the second key; thus, between 50% and 73% of the variance could be explained.

In contrast to experiment 1, object fanning was mirrored in the time for the first key selection (OF, $t = 4.29$, $p < 0.001$). As the first key was identical for all object-fanned items and fanning was produced by the second key, this indicated that the two key presses were no independent processes. Rather, it suggests that both keys had already been prepared in an action plan before the first response. This is a possible explanation for the occurrence of the fan effect at this position.

As in experiment 1, the 'overt vs. covert' variable had a significant impact (OC_1, $t = 6.66$, $p < 0.001$): The first key press was performed, on average, 1500 ms later when selecting covert items. Probably this effect was that strong because subjects initially tried to classify the item to the menu bar. However, the menu bar offered no hints for covert items. Although 'overt vs. covert' referred to the first key in the OC_1 variable, it still had an effect on the second key press ($t = 5.43$, $p < 0.001$). A marked increase in the slope coefficients was also still indicated there.

Item length also significantly affected both criterion variables (IL, $t = 4.29$, $p < 0.01$ respectively $t = 3.42$, $p < 0.01$). The slopes corresponded to those in experiment 1, namely 99 ms per letter for the first key and 20 ms per letter for the second key.

Table 4. Summary of multiple regression analyses on selection times for first key (menu bar) and second key (submenu) (β = mean slope coefficients, SE = standard error, t = t-value, r_c = criterion correlation).

Variable		First key: menu bar β	SE	t	r_c	Second key: submenu β	SE	t	r_c
External layout									
Position in menu bar	PM	78.58	27.30	2.88	0.082	6.95	14.42	0.48	0.07
Position in submenu	PS	− 17.58	41.86	− 0.42	0.206	− 73.01	20.04	− 3.64*	0.170
						(90.88	58.23	1.56)[1]	
Number of submenu items	NS	− 242.03	163.48	− 1.48	− 0.211	196.84	121.68	1.61	− 0.021
Cognitive menu structuring									
Horizontal fanning	AF	− 29.90	120.25	− 0.25	− 0.023	− 103.61	87.59	− 1.18	− 0.009
Vertical fanning	OF	812.06	189.46	4.29**	0.068	114.99	158.36	0.73	0.046
Submenu grouping	SMG	425.99	201.21	2.12	0.178	− 177.67	112.76	− 1.58	− 0.046
Overt vs. covert	OC_1	− 1524.13	228.84	6.66**	0.565	514.71	94.72	5 43**	0.312
Overt vs. covert	OC_2	− 272.53	107.54	− 2.54	0.038	188.53	66.12	2.85	0.190
						(580.70	174.67	3.32*)[1]	
Item level									
Item length	IL	99.06	23.09	4.29**	0.486	20.93	6.12	3.42*	0.184
Mean multiple R		0.850 (0.713–0.929)				0.716 (0.497–0.890)			
Mean multiple R^2		0.723				0.513			

[1] including the NS × OC_1-interaction: $\beta = -120.48$, SE = 43.64, $t = 2.76$, $p < 0.01$
* df = 14, $p < 0.01$ (2-tailed)
** df = 14, $p < 0.001$ (2-tailed)

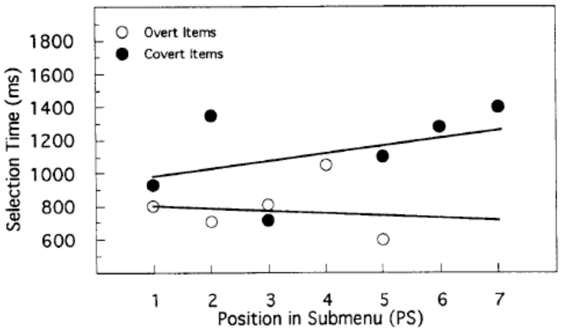

Figure 3. PS × OC_2 interaction. Mean selection times on the second key were summarized across all remaining variables.

As anticipated, variables that were assumed to be motor-dependent had no effects in this experiment, with only one exception: the PS variable (position in submenu) in selection time for the second key. However, this variable showed a negative slope and a positive criterion correlation ($t = -3.64, p < 0.01$), which, once more, could be evaluated as indicating the presence of an interaction. As no specific interaction predictions were formulated, all eight first-order interactions with PS were calculated in post hoc analyses. One interaction, namely, that with the OC_2 variable, proved to be marginally significant ($\beta = -120.48$, SE $= 43.64$, $t = 2.76, p < 0.05$; see figure 3). The regression line for overt items runs almost parallel to the abscissa, and that for covert items shows an increase as a function of position in the submenu. They seemed to be harder to recall the lower they were located in the submenu.

In this context, it is important to note that the main effect of the PS variable disappeared when the interaction was taken into account, and, instead, the anticipated effect of the OC_2 variable increased ($t = 3.32, p < 0.01$). When the second key did not have the same initial letter as the item and, accordingly, could be considered covert, selection time increased by an additional 580 ms ($t = 3.32$, $p < 0.01$).

5. General discussion and limits of the technique

If the procedure employed here is to be used for evaluation purposes, regression coefficients should mirror quantitatively the amount of each variable in selection time. The results meet this requirement.

The dependence of mouse selection times on motor components is revealed in the slopes, as the items within the menu bar and the submenu affect item selection. As predicted, interferences are elicited by both the number of items in the submenu and object fanning. This applies to both mouse and key selection. Heydemann (1989: 82–104) has reported a comparable fan effect in key sequences.

The length and the typicality of an item (overt vs. covert) proves to be relatively independent from the various subtasks (selecting menu bar and submenu with mouse or key). Both variables affect nearly all criterion variables in both experiments. Item length also shows a consistent effect across experiments 1 and 2. However, it has to be remembered that this variable may not only have been determined by the length of the items in letters, but possibly also by the different demands expressed in the complexity of an item. An item such as 'load file' has a short and concise meaning in a text editor, while items such as 'set up new page' require a shift from the editor level to the printable layout, and they are used less frequently.

Generally, the analysis technique used here can be recommended. When designing menus, it is impossible to avoid a number of factors, such as position and fan effects. Multiple regression, or the partitioning of selection times that it enables, is a tool that can be used for a quantitative evaluation of individual variables and that can be taken into account when revising menu designs.

However, the number of analyses that need to be calculated is immense. For the present designs with repeated measurements, slope coefficients had to be determined for each subject and each criterion variable, so that no less than 60 regression analyses were calculated for both experiments (without testing interactions and assumptions!). According to Lorch and Myers (1990), these calculations are needed for the inferential statistics of the slopes across subjects. The effort necessary is acceptable only because of the statistical programs available today.

When evaluating any particular (e.g., already existing) selection system, it may well be necessary to make compromises. Trying to keep collinearity at a minimum will probably create the greatest difficulties, especially when one is interested in a specific variable that correlates highly with others. The unavoidable exclusion of variables from the analysis may be painful; one would have to be satisfied with the information that can be drawn from correlations among these variables.

Interpretation becomes particularly problematic when possible interactions between variables are overlooked. The present example shows that regression coefficients can change drastically when an interaction is added. However, it is almost impossible to control all interactions in such a multivariate approach. This is especially risky when no directional hypotheses are available for either main effects or interactions. Fortunately, the results provide indications of possible interactions. In these cases, one has to be careful, as the validity of the regression coefficients is endangered.

The linearity assumption is less problematic. If a nonlinear relationship is indicated, this can be dealt with by a corresponding scale transformation. However, even with relatively reliable nonlinear relationships, linear approximations often fit as well, as our results demonstrate. Owing to the restricted possibilities of grading a predictor variable within a selection system, it will, in any case, often not be possible to decide on the type of relationship. In this respect, the technique applied here should be viewed basically as a supplement to existing techniques and as being less suitable for demonstrating general psychological laws.

A related problem originates from the fact that the predictors are arbitarily chosen and are not from random samples of different menus. Of course, it is possible that in another menu structure some effects reported here play no part in selection times, whereas other unnoticed effects and variables are important. Thus, the results do not justify the conclusion that the reported effects are found in general and in every menu. We restricted ourselves to demonstrate the technique with a specific menu and to examine those variables that can be found in many editors.

Hence, the use of the present technique should be limited to obtaining a general overview of the selection time parameters of a specific menu. By also including qualitative findings, interface designers then have the possibility jointly of considering 'on-line' and 'off-line' data and combining them optimally in new interface designs. How far the technique can also be applied to other types of dialogues will have to be assessed by further research.

Acknowledgements

I wish to thank Gisa Aschersleben, Wolfgang Prinz, Ben Shneiderman, and, in particular, Martina Hielscher for their helpful comments on a previous version of this contribution. Thanks are due to Jonathan Harrow and Heidi John for many insightful stylistic suggestions.

Notes

[1]The used menu was the result of a pilot study with 7 existing drop-down editors. In this study each item of the editors was classified on a number of variables. We tried to realize those variables for the present menu that were found in each of these editors.

[2]GEM (Graphics Environment Manager) is a registered trademark of Digital Research Inc.

[3]The significance level was set at $p < 0.01$ as the data analysis was based on three criterion variables obtained in one experiment, this giving rise to the risk of an alpha error.

[4]Including the interaction also changed the remaining slope coefficients. However, these changes were very small and were therefore disregarded.

References

ANDERSON, J. R. 1983, *The Architecture of Cognition* (Harvard University Press, Cambridge, MA).

AREND, U., MUTHIG, K. P. and WANDMACHER, J. 1987, Evidence for global feature superiority in menu selection by icons, *Behaviour and Information Technology*, **6**, 411–426.

CARD, S. K., ENGLISH, W. K. and BURR, B. J. 1978, Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT, *Ergonomics*, **21**, 601–613.

COHEN, J. and COHEN, P. 1975, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Erlbaum, Hillsdale, NJ).

FITTS, P. M. 1954, The information capacity of the human motor system in controlling amplitude of movement, *Journal of Experimental Psychology*, **47**, 381–391.

HABERLANDT, K. F. and GRAESSER, A. C. 1985, Component processes in text comprehension and some of their interactions, *Journal of Experimental Psychology: General*, **114**, 357–374.

HAYHOE, D. 1990, Sorting-based menu categories, *International Journal of Man–Machine Studies*, **33**, 677–705.

HEYDEMANN, M. 1989, *Der Fächerungseffekt* [the fan-effect], unpublished doctoral thesis, Institute of Psychology, Technical University of Darmstadt, Germany.

HYMAN, R. 1953, Stimulus information as a determinant of reaction time, *Journal of Experimental Psychology*, **45**, 188–196.

KARAT, J., MCDONALD, J. E. and ANDERSON, M. 1986, A comparison of menu selection techniques: touch panel, mouse and keyboard, *International Journal of Man–Machine Studies*, **25**, 73–88.

KERLINGER, F. N. and PEDHAZUR, E. J. 1973, *Multiple Regression in Behavioral Research* (Holt, Rinehart and Winston, New York).

KIGER, J. I. 1984, The depth/breadth trade-off in the design of menu-driven user interfaces, *International Journal of Man–Machine Studies*, **20**, 201–213.

KNIGHT, G. P. 1984, A survey of some important techniques and issues in multiple regression, in D. E. Kieras and M. A. Just (eds.), *New Methods in Reading Comprehension Research* (Erlbaum, Hillsdale, NJ), 13–29.

LANDAUER, T. K. and NACHBAR, D. W. 1985, Selection from alphabetic and numeric menu trees using a touch screen: breadth, depth, and width, *Proceedings of CHI '85, Human Factors in Computing Systems* (ACM, New York), 73–78.

LEE, E. and MACGREGOR, J. 1985, Minimizing user search time in menu retrieval systems, *Human Factors*, **27**, 157–162.

LORCH, R. F. and MYERS, J. L. 1990, Regression analyses of repeated measures data in cognitive research, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 149–157.

RADWIN, R. G., VANDERHEIDEN, G. C. and LIN, M.-L. 1990, A method for evaluating head-controlled computer input devices using Fitts's law, *Human Factors*, **32**, 423–438.

RICKHEIT, G., GÜNTHER, U. and SICHELSCHMIDT, L. 1992, Coherence and coordination in written text: reading time studies, in D. Stein (ed.), *Co-operating with Written Texts. The Pragmatics and Comprehension of Written texts* (De Gruyter, Berlin), 103–127.

SHINAR, D. and STERN, H. 1987, Alternative option selection methods in menu-driven computer programs, *Human Factors*, **29**, 453–459.

SHNEIDERMAN, B. 1986a, *Designing the User Interface: Strategies for Effective Human–Computer Interaction* (Addison-Wesley, Reading, MA).

SHNEIDERMAN, B. 1986b, Designing menu selection systems, *Journal of the American Society for Information Science*, **37**, 57–70.

WALKER, N., SMELCER, J. B. and NILSEN, E. 1991, Optimizing speed and accuracy of menu selection: a comparison of walking and pull-down menus, *International Journal of Man–Machine Studies*, **35**, 871–890.

VERWEY, W., GLASER, K., GRASS, A. and EBERLEH, E. 1988, *Zum Einfluß von Typikalität und semantischer Distanz von Menü-Items auf Suchzeiten und -genauigkeiten in hierarchischen Menüs* [the influence of typicality and semantic distance on response speed and accuracy in hierarchical menus], Technical Report No.1–45 (ACCEPT): Institute of Psychology, Technical University of Aachen, Germany.