Int. J. Human-Computer Studies (2001) **55,** 845–860
doi:10.1006/ijhc.2001.0506
Available online at http://www.idealibrary.com on IDEAL®

Task structure and the apparent duration of hierarchical search

NOAM TRACTINSKY

Department of Information Systems Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer Sheva 84105, Israel. email: noamt@bgumail.bgu.ac.il

JOACHIM MEYER

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, P.O. Box 653, Beer Sheva 84105, Israel. email: joachim@bgumail.bgu.ac.il

(Received 9 August 2000, accepted 7 August 2001)

Research in the area of human–computer interaction (HCI) suggests that long or variable system delays lower user satisfaction with the interaction and the system in general. Designers cannot always control the delays in a system's responses (e.g. when accessing remote servers), but it is possible to design human-computer interactions so that the *apparent duration* of intervals will seem minimal. One way of achieving this goal is to structure tasks so that their apparent duration is reduced, partly by altering the number of choices and actions required for performing the task. Two laboratory experiments assessed the effects of the number of choices and the number of ballistic (simple) steps in a menu search on the apparent duration of the search. Results showed that the apparent duration increased with an increasing number of ballistic steps, while the number of choices had no effect on estimates. However, apparent durations were the shortest when the ratio of choices to ballistic steps was maximized. The implications of these findings for interface design are discussed.

KEYWORDS: apparent duration; duration estimation; time perception; psychological time, interaction design; system delays; menu search.

1. Introduction

Over the last two decades, computers have become remarkably faster. At the same time, however, the increasing use of networks and communication lines have led to frequent slowing in computer response (often referred to as system response time) and increased variability in response times. Delays in system responses and high variability of response times have had detrimental effects on users' satisfaction with computer systems (e.g. Brown, 1988; Guynes, 1988; Shneiderman, 1998). Traditionally, effective designs of computer systems were measured by minimizing the time required for users to perform a sequence of operations. In today's Internet-based systems, however, these times are often negligible compared to the delay times imposed by the overloaded communication lines and servers. Moreover, because of the complexity of the systems involved, the length of system delays can hardly be controlled or predicted. Some authors consider system delays to be a major detriment to the prosperity of the Internet (Nielsen, 1999) and of electronic commerce. Sliwa (1999) cites a Zona Research Inc. report, which claimed that an estimated $4.35 billion per year in electronic commerce sales may be at risk because of unacceptable download times of web pages. Consequently, numerous attempts have been made to shorten as well as to stabilize response times so as to mitigate users' annoyance with the system.

Negative responses to system delays are particularly likely to occur when delays exceed users' expectations (Shneiderman, 1998). These negative effects can be alleviated to some extent by informing users about when to anticipate long response times and by providing appropriate feedback for lengthy tasks. Accordingly, traditional design guidelines recommend the display of accurate system status information in order to reduce the uncertainty involved with long and variable delays (Brown, 1988; Dix, Finlay, Abowd & Beale, 1993; Shneiderman, 1998).

Although the actual length of the delays may be given, it should be possible to design the interaction with the computer so that the apparent duration will appear shorter. This approach is particularly well suited for applications in which response times cannot be accurately predicted (e.g. internet browsing). Two aspects of interface design are relevant here. The first aspect is the decision as to what information or stimuli to display during the delay. For instance, it is possible to present static or dynamic displays, or displays that show some progress as opposed to displays that show only that a process is currently under way. A series of experiments has demonstrated that different displays can indeed affect the apparent duration of intervals (Meyer, Bitan & Shinar, 1995; Meyer, Shinar, Bitan & Leiser, 1996).

A second aspect of the interface design that may affect the apparent duration of delays is related to the organization of the task. The basic question here is whether an interaction period during which a number of minor actions are performed will seem shorter or longer, compared to an interaction period during which fewer actions take place. However, this issue has received relatively little attention to date. Our goal is to investigate whether it is possible to design the human–computer interaction (HCI) so that, given a certain completion time of a task, the users' perceptions of that time will be minimal. The use of design principles to alter people's estimates of time delays is not new. For example, fast-food chains and theme parks use various line systems that are designed to display signs of line progress, which will, in turn, reduce psychological time (Solomon, 1996). We believe that in the field of HCI, similar principles can be defined that can help to shorten the apparent duration of a time interval. In particular, actions should be structured, so that the apparent duration of a sequence will appear as short as possible. For this purpose, we distinguish between two types of user actions during a human–computer interaction. These types are presented in the next section, followed by a discussion of some relevant methodological issues in the study of duration estimation. We then report two laboratory experiments that examined the effects of different interface designs on duration evaluation.

2. Variables impacting duration estimation

2.1. CHOICES AND BALLISTIC STEPS

For the purpose of this study we distinguish between two types of user actions that are performed during human–computer interaction—choices and ballistic steps. Choices or decisions are made whenever the user confronts two or more alternative options and has to choose one of them in order to proceed with the task. A choice involves some thought process that utilizes the user's cognitive resources (although the extent of such processes may vary considerably). After a choice has been made, users may need to execute a series of actions—named here "ballistic steps". These actions are a necessary consequence of the decision, are prescribed by the system, and are performed without additional deliberation. An example of ballistic steps is repeated presses on OK buttons to choose some default setting in a frequently used program. We choose the term "ballistic" because these steps are the necessary outcome of a previous decision. They are, of course, not actual ballistic movements in the sense for which the term is used in the context of motor behavior. Rather, they are similar to the notion of "operators" in GOMS models (Card, Moran & Newell, 1983), only that we assume that no mental preparations (which are optional operators in GOMS models) are needed for ballistic actions.

The distinction between choices and ballistic steps is supported by findings regarding two key variables in duration estimation research. One variable that has been demonstrated to affect duration estimation is the pace of events that occur during the estimated interval (e.g. Zakay, Nitzan & Glicksohn, 1983). It has been consistently shown that a faster pace of events leads to longer duration estimates (e.g. Zakay *et al*., 1983). Previous studies on duration estimation in human–computer interaction (Meyer *et al*., 1995, 1996) demonstrated that faster blink rates or more rapid advances of a graphic symbol led to increased duration estimates. This phenomenon suggests that more frequent low-level actions, such as ballistic steps, may lengthen the apparent duration of an interval. Thus, we propose that the number of "ballistic steps" during a time interval will be positively related to users' duration estimates of that interval.

2.2. REQUIRED MENTAL RESOURCES AND METHODOLOGICAL CONSIDERATIONS

Another variable that has been identified as pertinent to duration estimation is the mental resources that are required to complete a task. Findings and predictions regarding the effects of the cognitive complexity of a task on duration estimates are equivocal. Some studies showed an increase in the apparent duration for tasks that require more cognitive resources (e.g. Ornstein, 1969), while others showed a shortening of apparent durations for higher cognitive demands (e.g. Brown, 1985; Zakay *et al*., 1983). Zakay (1990, 1993; Zakay & Block, 1997) suggested that these conflicting results can be explained by differences in the methodology of duration estimation measurement. In particular, the response mode and the participants' awareness of the need to estimate the duration of the interval might moderate the effect of the cognitive complexity.

2.2.1. Response mode. Clausen (1950) distinguished between two response modes—one in which duration estimates can be absolute, i.e. the participant in the experiment uses some external absolute measure, such as seconds, to indicate the duration of an interval, and another in which duration estimates can be comparative. In the latter, a standard interval is presented to the participant who, in turn, compares the target interval to this standard. For example, an object may be presented for 10 s. This duration is defined as "100 interval points" and the duration of subsequent stimuli are then evaluated relative to the standard. The evaluator assigns a value greater than 100 if the apparent duration of a stimulus seems longer than the standard interval, or a value smaller than 100 if the stimulus' apparent duration seems shorter. There are two main reasons for using the comparative response mode. First, the use of standard estimates discourages the counting of seconds while performing the task, whereas direct numeric estimates of time intervals in seconds may induce the use of counting. In addition, the notion of a second is often not quite clear, and great interpersonal variance may exist in the apparent duration of seconds. The drawback of using the comparative method is that it requires the individual to actually refer to and compare between two time intervals, while the researcher is usually interested only in the estimation of one of these intervals.

2.2.2. Estimation method. The other methodological issue refers to the point in time when the participants in a duration estimation experiment become aware of the fact that they have to estimate the duration of the interval. With "prospective duration estimation" the participants know that they will have to estimate the duration of an interval before it begins. In contrast, a "retrospective duration estimation" paradigm requires participants to estimate the duration of an interval after it has ended. It is not clear what duration estimation methodology is closest to the conditions that exist during the actual use of a system. When an unexpected delay in system response occurs, duration estimation is similar to the estimation in the retrospective paradigm since the evaluation of the interval occurs only after it has ended. However, when users expect delays, they may attend to the delay's duration in advance, as in the prospective paradigm.

Zakay (1990, 1993; Zakay & Block, 1997) suggested that the effect of a task's cognitive complexity on its duration estimate depends on the combination of the estimation paradigm (prospective or retrospective) and the method of duration estimation (absolute or comparative). In retrospective duration estimates, increased mental demands that arise from a more difficult task will lead to a lengthening of the apparent duration of the interval. This finding is accounted for by models that deal with the memory trace of the interval as the basis for estimation (Ornstein, 1969). In contrast, when prospective time estimates are performed, the mental demands tend to be inversely related to the apparent duration, i.e. more demanding tasks seem to have shorter durations. This finding is in line with the predictions of models of duration estimation that maintain that the estimation process is based on the use of a mental timer. The high cognitive demands limit the allocation of resources to the timer, leading to the shorter apparent duration of the interval (Hicks, Miller, Gaes & Bierman, 1977; Thomas & Weaver, 1975; Zakay, 1989). Thus, duration estimation models suggest that more choices—i.e. cognitively more demanding tasks (Einhorn & Hogarth, 1981)—should lower duration estimates under prospective time estimation, but should raise duration estimates under retrospective duration estimation. Table 1 summarizes the predictions about the effect of the number of choices on duration estimates for these two elicitation methods.

Zakay (1990, 1993; Zakay & Block, 1997) also suggests that the estimation method moderates the effects of different variables. Comparative estimates must rely on the memory trace of the duration of the standard interval (or of the target interval if the standard is presented after the target) and will therefore be affected by variables that affect the memory trace. Absolute estimates rely less on memory and are less affected by these variables. This implies that choices are more likely to affect comparative duration

TABLE 1
Predictions of the effects of the number of choices on duration estimates for prospective and retrospective estimates and absolute and comparative judgments. Overall predictions regarding the two estimation methods and the two response modes appear in capital italics. The combined predictions are presented in the four lower-right cells

		Response mode	
		Absolute Measurement	Comparative Measurement
Estimation method	*Predicted effect of choices*	*No prediction*	*Choices lengthen duration estimates (memory trace model)*
Prospective method	*Choices shorten duration estimates (timer model)*	No prediction	Conflicting predictions
Retrospective method	*Choices lengthen duration estimates (memory trace model)*	Increased duration estimation	Increased duration estimation

estimates in a manner commensurate with the memory trace model. That is, in comparative estimates the presence of choices during a time interval causes an increase in the apparent duration. Table 1 also summarizes the predictions about the effect of the number of choices on duration estimates for the two measurement modes and for the four possible combinations of the two elicitation methods and the two measurement modes.

2.3. THE CURRENT STUDY

The particular context in which duration estimates were assessed in this study is the search in hierarchical menu systems. This context was chosen for two reasons. First, it provides a natural setting for the manipulation and measurement of the study's variables. Second, it allows for strict control over the participants' actions and thus increases the study's internal validity. Clearly, the emphasis on internal validity may come at the expense of the generalizability of the study's findings. However, our main goal is to establish the effect of the task structure on the users' duration estimates. These findings can then be applied to various HCI domains such as the structuring of hypertext hierarchies and the design of e-commerce and other web sites. We will return to this issue in the concluding section of the paper.

Menu trees are widely used in HCI, and their properties have been thoroughly researched (e.g. Norman, 1991; Shneiderman, 1998). One characteristic of menu search that has received much attention is the breadth vs. depth trade-off. In broad trees, the number of levels in the hierarchy is small and the number of alternatives in each level is large. In narrow trees, in contrast, there are relatively many levels in the hierarchy, but the number of alternatives in each level is small.

Consider a menu-based system in which users have to move across the various levels of the menu to reach a desired item by key presses. In this case, we can distinguish between two different types of activities: (1) the user chooses an item that leads to the next desired level in the hierarchy; and (2) the user presses keys to move to the chosen item. We consider the two types of key presses as resembling the two different types of user actions—choices and ballistic steps. The first is a choice among alternatives, which requires some decision-making. The other is a motor response that is cognitively simple and falls within our definition of a "ballistic step".

Two experiments were conducted to study how choices and ballistic steps affect the user's duration estimates. Both experiments employed the prospective evaluation method for duration estimation. We chose the prospective paradigm rather than the retrospective method because in the latter, a user can only perform a single duration evaluation (once a user is sensitized to duration evaluation, additional evaluations will no longer be retrospective). Thus, the results of this study apply to cases in which users anticipate system delays or are aware of the possibility of such delays, rather than cases in which delays occur unexpectedly.

3. Experiment 1

3.1. METHOD

Participants. Twenty-five third-year engineering students participated in the study as partial fulfillment of a course requirement.

Materials. A computer program, written in Visual Basic was used to display stimuli and to record the participants' actions. The program was run on 133 MHz Pentium computers, using 15″ monitors.

The task. Participants had to locate specific items (the targets) in a hierarchical menu system. The target items consisted of one, two, or three characters (letters and numerals). The number of characters in an item indicated the depth of the menu structure. For a one-character item the target was at the first menu level, for a two-character item it was present at the second level of the menu structure, and so on. When a target item consisted of more than one character (and thus was located below the first level), participants had to select a character in the menu's upper level that matched the item's first character. In the menu's second level, they had to select an item that matched the target's first two characters. Finally, if the target was comprised of three letters (the maximum possible length), participants had to find and select the matching item in the menu's third level. Figure 1 presents an example of a three-character target selection process.

Participants used the arrow keys to move among menu items and the Enter key to select the appropriate item. The use of keys instead of a mouse for navigation and selection allowed us to control for the minimal number of ballistic steps that are required to reach a target.

Overall, participants faced 40 selection trials, with 27 experimental tasks that lasted 12 s each and 13 filler trials, each lasting 16 s. The conclusion of one trial and the

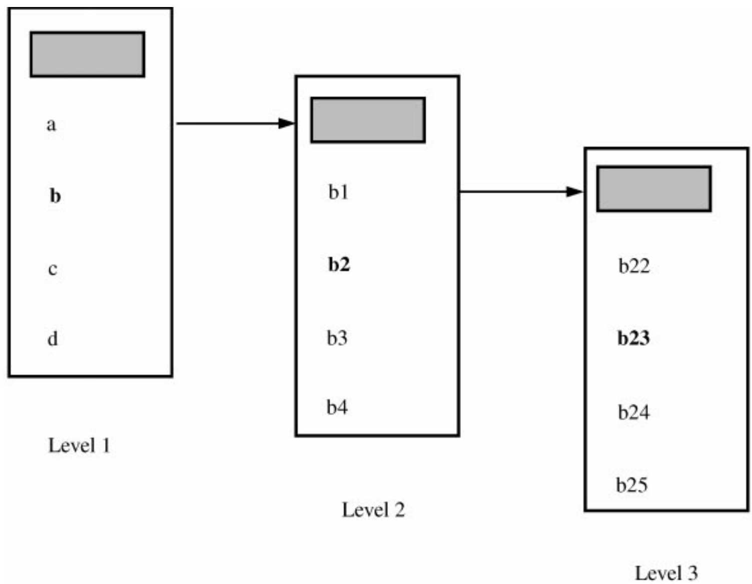

FIGURE 1. An example of finding a three-character stimulus ("b23") in the menu hierarchy. Participants had to select item "b" at the first level, then item "b2" at the second level, and then item "b23" at the lowest level. The gray rectangle at the top of each level indicates the default position of the cursor when the participant enters that level.

beginning of the next trial were separated by a 1-s delay. Since each trial had to be completed in a predetermined time interval, we had to introduce a mechanism to ensure that participants could not complete the task too early. This was achieved by introducing forced delays between navigation steps. To smooth the variance in delay times within a trial, the delay time for each step was calculated with an algorithm that took into account the remaining time for task completion and the number of remaining steps until the target item was reached. After each keystroke, the algorithm calculated whether a delay is needed, and its length, if needed. The delay was calculated according to

$$D_i = (12 - E_i)/(S_i + 1), \tag{1}$$

where D_i is the delay enforced by the system after step i was performed; E_i the time elapsed since the beginning of the trial until step i was performed and S_i the number of steps left after step i.

Time estimation method. The comparative evaluation method was used to elicit duration estimates. Before the first trial, and again after every five trials, the program displayed a criterion screen. The screen consisted of a red square that appeared for 10 s. Participants were told that this time was the equivalent of 100 units. After each trial they were asked to enter a numeric value that corresponds to the amount of time it took them to perform the experimental task, relative to the 100-unit criterion. They were told to enter larger numbers if they felt that the time was longer than the criterion, and smaller numbers if they felt that the trial was shorter than the criterion.

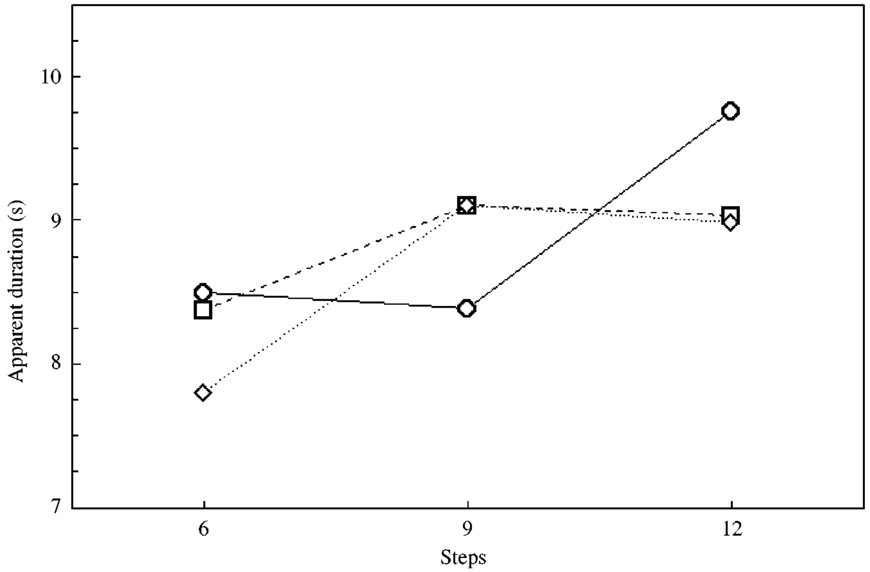

FIGURE 2. Duration evaluations (in seconds) in Experiment 1 as a function of the number of ballistic steps and the number of decisions. Decisions: —○—, 1; --□--, 2; ··◇···, 3.

Experimental design. The 27 experimental trials consisted of the combinations of 3 numbers of ballistic steps (3, 6 or 9 steps) × 3 numbers of choices (1, 2 or 3 choices) × 3 repetitions. A ballistic step was defined as a keystroke, while a choice was defined as a decision to move to a lower level of the menu hierarchy or to terminate the search (by pointing to the target item). For example, selecting the target in Figure 1 involved 9 ballistic steps and 3 decisions. Each combination of the two factors was repeated 3 times in a random order. The duration of all the trials, without regard to the number of steps and choices, was 12 s. Thirteen distracting trials were inserted randomly among the experimental trials. The distracting trials were similar in form to the experimental trials except that they were presented for 16 s.

For each trial, the actual number of keystrokes that were performed until the target was reached and the time required to complete the trial were recorded. If the completion of a trial exceeded the allotted time by more than 2 s, or if the number of keystrokes used exceeded the minimal number by more than 33%, the trial was discarded and was repeated later at some random position in the sequence of trials. Overall, participants performed 42.64 trials on average, indicating that about 6.6% of the trials had to be repeated.

3.2. RESULTS

The dependent variable in this study was the subjective estimate of the trial's duration, compared to the 100-unit baseline. While screening for outliers, we found two time estimates under the condition of 9 steps and 3 decisions, which departed from the sample's mean by more than 3 standard deviations. These two data points were not used in the ensuing analyses.

A linear transformation function was used to convert participants' time estimates from the 100-point baseline scale back to seconds. The transformed times are depicted in Figure 2. The mean duration estimate for the 12-s intervals was 8.79 s, indicating a tendency to underestimate actual time intervals. The results were analysed in a two-way analysis of variance, with the number of steps (6, 9, and 12) and the number of decisions (1, 2, and 3) as within-subject variables. The ANOVA showed a significant main effect of the number of ballistic steps on duration estimates, $F(2, 48) = 6.878$, $p < 0.001$. More ballistic steps were associated with longer duration estimates (mean estimates were 8.23, 8.87 and 9.27 s for the 6, 9 and 12 steps, respectively). There was no significant main effect of the number of choices on duration estimation, nor was there an interaction effect of Ballistic Steps × Choices. Thus, the only factor affecting duration estimation was the number of ballistic steps: more ballistic steps were associated with longer estimates of trial duration.

4. Experiment 2

Experiment 1 showed a significant main effect of the number of ballistic steps on duration estimates, which is consistent with our predictions and with previous research on duration estimation (Meyer *et al*., 1995, 1996). However, the number of choices had no significant effect on the estimates. This could be due to the very simple nature of the stimuli used in this experiment. While navigating through menu items, participants were required to decide whether a short string of characters matches that of the target item. This type of a decision can be made very quickly, and thus may not have any noticeable effect on psychological time. In order to examine this possibility, participants in Experiment 2 were required to make choices that were more complex.

Experiment 1 employed only the comparative measurement method of duration estimation. Experiment 2 was designed to test whether duration estimates are affected by the measurement method. As summarized in Table 1, there are conflicting predictions for the prospective paradigm and the comparative method (which were both used in Experiment 1) with regard to the effect of the number of choices on duration estimates. More choices should lead to longer apparent durations with comparative judgments, and to shorter apparent durations with the prospective method. In addition, the effect of the number of choices on absolute duration judgments is not clear. It is of importance to examine the absolute method for two reasons. First, we seek converging evidence from both measurement methods regarding the effects of ballistic steps and choices on duration estimation. Second, practical time estimations during human–computer interaction may be based on either method. That is, users may estimate delays or interaction times by comparing them to other time intervals (e.g. to the time it took to complete similar tasks on other systems or applications, or the connection to other web sites). Alternatively, users may estimate interaction duration directly (e.g. by evaluating a task completion time or system response time in and of itself).

4.1. METHOD

Participants. Ninety-nine third-year engineering students (who did not participate in Experiment 1) participated in the study as partial fulfillment of course requirements.

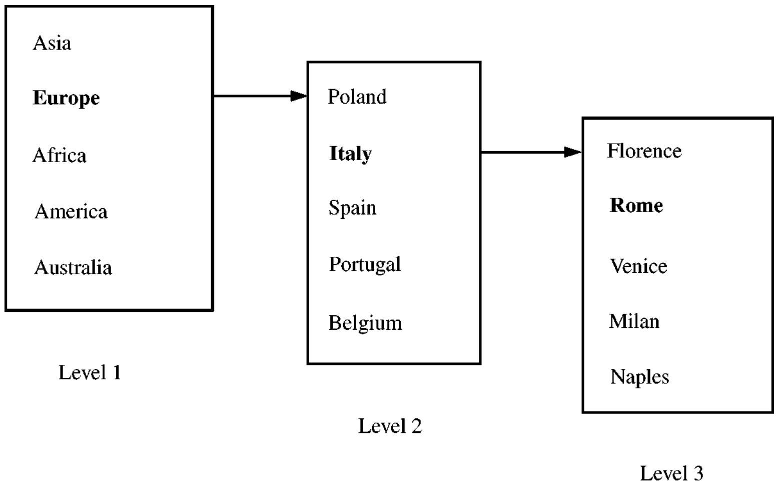

FIGURE 3. An example of finding the item "Rome" in a menu hierarchy. Participants had to select item "Europe" at the first level, then item "Italy" at the second level, and then item "Rome" at the lowest level.

Materials and procedure. Materials were the same as in Experiment 1. The task was similar to the one used in Experiment 1, except that the menu items now were more realistic and the target items more difficult to match. Figure 3 displays a menu hierarchy in which the participant is required to locate the item "Rome". While it was our intent that items would not be as simple to match as in Experiment 1, we had to balance this goal with the concern that users would still be able to select the target item within the prescribed time interval. Thus, the stimuli for this experiment had to be taken from well-known domains to ensure that participants would not require too much time to select the appropriate item in any of the menu levels.

Stimuli. Two groups of stimuli were used in this experiment. The first group was identical, in terms of the experimental design, to the stimuli of Experiment 1 (i.e. there were combinations of 3, 6 or 9 steps and 1, 2 or 3 choices), except that each of the nine combinations was only presented twice (instead of three times as in Experiment 1). Items in this group were presented in a random order and had to be completed in 12 s. A second group of items consisted of nine filler items that had to be completed in 16 s. The filler stimuli here served to test the effect of the order of different path lengths on the apparent durations of tasks. This was done by using three different types of paths, each having 12 ballistic steps and two choices. They differed in the number of ballistic steps that were performed before and after the first decision. The three types were (a) a short path (3 ballistic steps) before the first decision and a long (9 ballistic steps) path after it; (b) a long path before the decision and a short one after it, and (c) equal-length paths (6 steps each) before and after the decision.

Time estimation method. Participants were randomly assigned to one of two duration estimation groups. One group, consisting of 46 participants, reported duration estimates with the comparative method (as in Experiment 1), while the second group, which consisted of 53 participants, reported estimates with the absolute estimation method. After each trial, participants in this group had to enter the number of seconds that they believed were required to complete the trial.

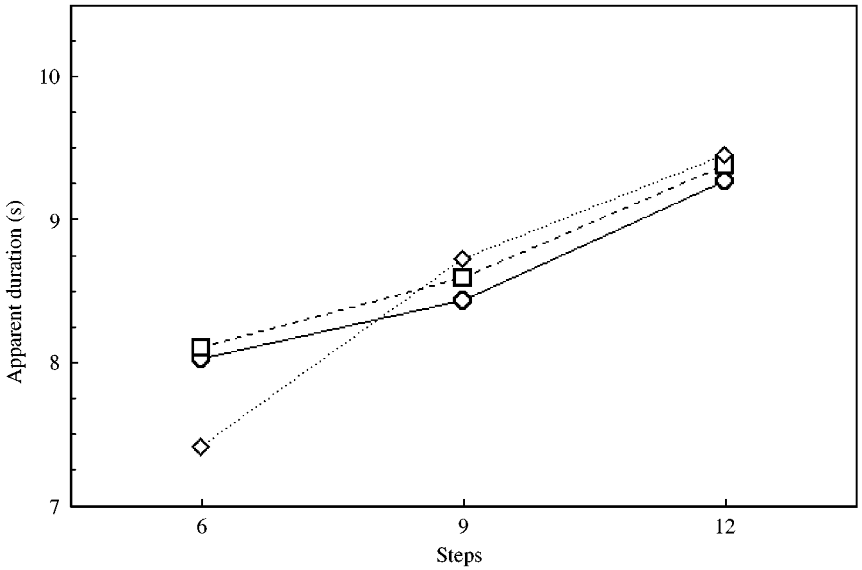

FIGURE 4. Duration evaluations (in s) in Experiment 2 as a function of the number of ballistic steps and the number of decisions. Decisions: —○—, 1; --□--, 2; ··◇···, 3.

4.2. RESULTS

Several duration estimates of two participants departed from the sample's mean by an order of magnitude. These participants (one from each estimation method group) were subsequently omitted from further analyses. For the comparative estimation group, time estimates were linearly transformed from the baseline scale (in which 100 points represented 10 s) back to seconds. Overall, the mean duration estimation for the 12-s actual intervals was 8.64 s.

The experimental items were analysed with a three-way ANOVA, with the number of ballistic steps (6, 9, and 12) and the number of choices (1, 2, and 3) as within-subject factors, and the measurement method as a between-groups factor. The ANOVA showed a significant main effect of the number of steps, $F(2, 190) = 38.81$, $p < 0.001$. In post hoc comparisons between the categories with a different number of steps, the differences between each pair of categories were significant at the 0.001 level. Consistent with Experiment 1's findings, more steps were associated with longer duration estimates (see Figure 4). The method of measurement had a significant effect on duration estimates, $F(1,95) = 4.87$, $p < 0.05$. Duration estimates with the absolute method were longer (9.14 s) than those obtained with the comparative method (8.06 s). No significant main effect of the number of decisions was found, but there was a Steps $\times$ Decisions interaction, $F(4, 380) = 2.54$, $p < 0.05$. The cause of the interaction was a significantly lower duration estimate for trials that had the minimal number of ballistic steps and the maximum number of decisions (see Figure 4). It can be seen that for the 6-step condition, the duration estimates with 3 decisions were significantly shorter than the estimates with 1 or 2 decisions. For the 9- and 12-step conditions there were no differences in duration estimates between the different numbers of decisions. A reinspection of Figure 2 shows that a similar (though not significant) pattern also existed in Experiment 1.

The results of duration estimates for the filler items (which were designed to test order effects) were analysed using a two-way ANOVA, with order as a within-subject factor and estimation method as a between-groups factor. The results indicate a significant order effect, $F(2, 190) = 3.54$, $p < 0.05$. Paths that consisted of fewer ballistic steps in the first stage (before the first choice) and more steps in the second stage seemed to last longer (mean = 11.43 s) than paths that consisted of an equal number of steps before and after the first choice (mean = 10.80 s). (Note that the actual duration of these paths was 16 s.) An LSD test between the two types was significant at the 0.01 level. The paths that included fewer steps before the first choice also appeared to be longer than the paths containing more steps before that choice (mean = 10.99 s), although the LSD test was only marginally significant ($p = 0.07$).

The results of Experiment 2 generally support the findings of Experiment 1. Participants tended to underestimate the duration of their interactions with the computer. Duration estimates were positively related to the number of ballistic steps during the interaction. Likewise, the number of choices did not affect duration estimation. However, in Experiment 2, we found an interaction effect between the number of ballistic steps and the number of choices. In addition, in Experiment 2, participants' estimates were longer when they used the absolute measurement method, relative to the comparative method. Finally, we found that, using the comparative method, participants gave shorter estimates for the same 12-s intervals in Experiment 2 (8.06 s) than in Experiment 1 (8.79 s). This finding is in accordance with the predictions of Zakay's model for the prospective-comparative elicitation method. The model states that a higher cognitive load (embodied in the more difficult stimuli of Experiment 2) will result in shorter duration estimates. These results may indicate that, during an interaction, the type of choices may have a stronger effect on duration estimation than the number of choices (given the range of one to three decisions used in our studies).

5. Discussion of the results

The results of the two experiments are consistent. There is clear evidence that the number of simple actions (ballistic steps) that are performed during a human–computer interaction affects duration estimation, even if the actual duration remains constant: the larger the number of ballistic steps, the longer the duration estimation. This finding resembles results from studies in other domains, which found that the number of events during an interval was positively correlated with the apparent duration of the interval (Zakay *et al.*, 1983; Meyer *et al.*, 1996). However, there is one important difference between those studies and ours. While previous studies showed that the pace of events affects a passive user (i.e. one who only perceives system messages), the current study demonstrates a similar effect on an active user who performs more or less frequent actions.

The number of choices had no main effect on duration estimates in either of the experiments. Thus, if indeed more decisions are associated with greater cognitive complexity, the current study does not support the notion that greater complexity leads to shorter duration estimates. The lack of an effect of the number of choices on duration estimation corresponds to the equivocal predictions by the duration estimation models (see Table 1). The combination of the comparative mode of measurement and the prospective time orientation may have triggered both time estimation mechanisms (memory trace and internal timer), with each canceling the other's effect. Another possible explanation is related to the complexity of the tasks in our study. The choices required here may have been too simple to create an effect. Further research with more complex decisions may be required. Yet, it should be noted that the paradigm of duration estimation used in our study depends on participants' completion of the task within a strict and relatively narrow time interval. The introduction of more complex tasks will inevitably increase variability in completion times and, subsequently, decrease the probability that completion times fall within the required time interval. The major advantage of our paradigm is that the actual duration of the task is independent of the number of steps and decisions. Other measurement paradigms may be needed to assess the effects of more complex choices on duration estimation.

A possible limitation of the experimental paradigm used in this study is that the algorithm that forced delays between quick keystrokes may have annoyed speedy participants. This is again a necessary trade-off that we faced when designing the experiments. The main research question we addressed was how users estimate the duration of interactive sessions. Consequently, the experimental paradigm was designed to increase the internal validity of the findings regarding this question. As with any controlled research, other questions may arise regarding the effects of the uncontrolled elements on various dependent variables. Thus, while the question of how system delays affect various user types is important, it was clearly beyond the scope of the study.

Both experiments indicate that users tend to underestimate the duration of active interaction intervals, compared to the actual time. The cognitive complexity of the interaction, as well as the elicitation method of duration estimation, may have affected duration estimates. Our experiments were not designed in advance to test these effects, so explanations for these results should be treated with some caution. Yet, from comparing the results of Experiment 1 with those of the comparative estimation group in Experiment 2, it seems that as the cognitive complexity of the stimuli increased, users' tendency to underestimate interaction intervals was accentuated. In addition, the comparison of the two measurement methods in Experiment 2 shows that absolute measurements of interaction duration alleviate those underestimations.

An interaction between the number of ballistic steps and the number of decisions was found in Experiment 2 but not in Experiment 1. However, the pattern of results in both experiments was the same and the difference in statistical significance may be due to the larger sample size in Experiment 2. In essence, the interaction was due to the fact that the only point at which results deviated from the linear pattern between duration estimates and number of ballistic steps was the combination of a minimal number of ballistic steps and a maximal number of choices. This condition has the smallest ratio of simple steps to choices. While the ratio of steps to decisions was not a significant determinant of duration estimates in general, our conjecture is that when this ratio is very small (i.e. every step is followed by a decision), it may shorten users' perception of the duration. In a sense, this condition resembles a waiting line of relatively short size (e.g. 6 steps) that was further divided into three very short sub-lines, each of which consists of two steps.

6. Implications

Although our study was carried out in the context of a search in keyboard-based hierarchical menu systems, its findings regarding duration estimates should not be limited to systems of this type. As noted earlier, searching within a menu system served to create a familiar context for the experiments and to increase their internal validity. However, we aimed to study the effect of task organization on the apparent duration of human–computer interactions in general. This issue is particularly important for the design of systems in which long delays of varying duration occur, e.g. various web-based applications. In such systems, the additional time that results from a single keystroke or even a mental operation is often negligible compared to the delay times imposed by the system. Such systems include web portals, which design their home pages in a menu-like form. Other sites use various navigation mechanisms that lead users through various arrangements of links. For example, E-commerce sites offer different solutions to help users find their way both inward (towards the store's merchandise) and outward (for completing the purchase process). Similarly, in the foreseeable future most WAP applications will have to rely on menu-based systems activated by key-presses or touch pens. These designs may reside at various points along the breadth and width dimensions. Our findings indicate that, all else being equal, leading users through a long sequence of simple actions raises the apparent duration of the interaction, relative to designs that trade off some of the simple steps with explicit choices. In a sense, this finding stands in contrast to common guidelines of HCI design, which call for the utmost simplicity. The probability of error grows with the complexity of the user's response, suggesting that simple actions (i.e. ballistic steps) may be better than more complex ones (i.e. choices). Yet, from the standpoint of improving the user experience, this study suggests that the added complexity introduced by including relatively more choices in the interaction may, in some cases, prove beneficial.

6.1. COMPARISON TO PREDICTIONS BY PERFORMANCE TIME MODELS

The results indicate that to reduce users' estimates of interaction times, the number of simple user actions should be minimized. We also found that the number of decisions required from the user during the interaction does not affect duration estimates, except for the minimum steps—maximum choices condition. It is interesting to note that minimizing the number of keystrokes (e.g. creating short cuts) has been a common guideline when the paramount criterion of the interface's design is the interaction's efficiency (e.g. Card *et al*., 1983). From this perspective, the keystroke-level model (KLM) for user performance time, for example, penalizes the design for any additional required keystroke (Card *et al*., 1983). Our study suggests that a large number of ballistic steps should be discouraged not only because it actually leads to longer interaction times, but also because it increases the psychological perception of time. This is not the case, however, with mental operations during the interaction. The KLM suggests that the mental preparation for executing an action (which is analogous to making a choice in our study) increases actual performance time much more than performing a simple action (i.e. "ballistic step"). Yet, our results suggest that given a constant interaction interval, a larger number of mental operations does not increase users' *estimates* of that interval.

In fact, a combination of relatively fewer simple actions and more mental actions reduces duration estimates. Thus, HCI designers of systems that are likely to be used under circumstances that induce long and diverse system delays (such as WWW applications) may be wise to shorten the apparent duration of interactions by designing them to be more cognitively challenging and less repetitive.

6.2. ETHICAL ASPECTS

The above conclusion may be subject to debate on ethical grounds. While our recommendation to reduce simple keystrokes or interactive actions befits general recommendations for HCI design, suggesting more cognitively demanding actions does not. After all, more cognitively demanding interaction stands in stark opposition to one of the pillars of interaction design—ease of use. Yet, it is important to bear in mind that the high level goal of interaction design is to improve the users' quality of life (e.g. Shneiderman, 1998). That quality of life is multidimensional and depends on various task and context contingencies. For example, rather than sitting idly by their computer or performing simple manual tasks, users who are subjected to long network delays may prefer to fill their waiting time by engaging in various cognitive tasks. We have shown that frequent choices can shorten the apparent duration of the interaction. We cannot be sure, however, whether this subjective feeling is worth its price, but interaction designers may want to consider this trade-off under some conditions.

7. Conclusion

Our results show that HCI design can affect users' estimation of the duration of an interaction. These estimates can, in turn, affect users' satisfaction and their interaction patterns. The results also suggest that taking psychological time into account can introduce new trade-offs for the design of computerized systems. Depending on the context of the interaction and the major goals of the designers, design recommendations that are based on duration estimates may either coincide or conflict with established HCI guidelines. We suggest that the duration estimation perspective is especially viable for the design of networked systems that are susceptible to long or variable delays. Relatively little research has been done to date on apparent duration in HCI, and further research is necessary to improve our understanding of how duration estimates are affected by various design features, as well as how they affect other dimensions of the interaction.

References

Brown, S. W. (1985). Time perception and attention: the effect of prospective and retrospective paradigms and task demands on perceived duration. *Perception and Psychophysics*, **38,** 115–124.

Brown, C. M. (1988). *Human–Computer Interface Design Guidelines.* Norwood, NJ: Ablex Publishing.

Card, S. K., Moran, T. P. & Newell, A. (1983). *The Psychology of Human–Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum.

Clausen, J. (1950). An evaluation of experimental methods of time judgment. *Journal of Experimental Psychology*, **40,** 756–761.

DIX, A., FINLAY, J., ABOWD, G. & BEALE, R. (1993). *Human–Computer Interaction*. Englewood Cliffs, NJ: Prentice-Hall.

EINHORN, H. J. & HOGARTH, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, **32,** 53–88.

GUYNES, J. L. (1988). Impact of system response time on state anxiety. *Communications of the ACM*, **31**(3)**,** 342–347.

HICKS, R. E., MILLER, G. W., GAES, G. & BIERMAN, K. (1977). Concurrent processing demands and the experience of time in passing. *American Journal of Psychology*, **90,** 413–446.

MEYER, J., BITAN, Y. & SHINAR, D. (1995). Displaying a boundary in graphic and symbolic "wait" displays: Duration estimates and users' preferences. *International Journal of Human–Computer Interaction*, **7,** 273–290.

MEYER, J., SHINAR, D., BITAN, Y. & LEISER, D. (1996). Duration estimates and users' preferences in human–computer interaction. *Ergonomics*, **39,** 46–60.

NIELSEN, J. (1999). The Top Ten *New* Mistakes of Web Design. Available online: http://www.useit.com/alertbox/990530.html.

NORMAN, K. L. (1991). *The Psychology of Menu Selection*. NJ: Ablex Publishing Co.

ORNSTEIN, R. E. (1969). *On the Experience of Time*. Middlesex, England: Penguin Books.

SHNEIDERMAN, B. (1998). *Designing the User Interface*, 3rd edn. Reading, MA: Addison-Wesley.

SLIWA, C. (1999). Web sites battle slow download times. *Computerworld*, **33,** 8.

SOLOMON, M. R. (1996). *Consumer Behavior*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

THOMAS, E. A. C. & WEAVER, W. B. (1975). Cognitive processing and time perception. *Perception & Psychophysics*, **17,** 363–367.

ZAKAY, D. (1989). Subjective time and attentional resource allocation: an integrated model of time estimation. In I. LEVIN & D. ZAKAY, Eds. *Time and Human Cognition*: *A Life Span Perspective*, pp. 365–395. Amsterdam: North-Holland.

ZAKAY, D. (1990). The evasive art of subjective time measurement: Some methodological dilemmas. In R.A. BLOCK, Ed. *Cognitive Models of Psychological Time*, pp. 59–83. Hillsdale, NJ: Erlbaum.

ZAKAY, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics*, **54,** 656–664.

ZAKAY, D. & BLOCK, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science*, **6,** 12–16.

ZAKAY, D., NITZAN, D. & GLICKSOHN, J. (1983). The influence of task difficulty and external tempo on subjective time estimation. *Perception & Psychophysics*, **34,** 451–456.

Paper accepted for publication by Associate Editor, Dr G. Sundstrom