Behavioral and Emotional Consequences of Brief Delays in Human-Computer Interaction

André J. Szameitat[a,b], Jan Rummel[c], Diana P. Szameitat[d], Annette Sterr[b]
[a]Ludwig Maximilians University, Munich, Germany, [b]University of Surrey, Guildford, UK, [c]University of Marburg, Marburg, Germany, [d]University of Tübingen, Tübingen, Germany,

Present address & address for correspondence

Dr André J. Szameitat	email: szameitat@psy.lmu.de
Department of Psychology	phone: 0049-(0)89-2180 6778
Ludwig Maximilians University	fax: 0049-(0)89-2180 4866
Leopoldstrasse 13	
80802 Munich	
Germany	

Author's email addresses

André J Szameitat	szameitat@psy.lmu.de
Jan Rummel	rummel@uni-mannheim.de
Diana P. Szameitat	d.szameitat@gmx.de
Annette Sterr	a.sterr@surrey.ac.uk

Abstract

It was thought that computer-performance related problems in human-computer interaction (HCI) would become negligible with the constant increase in computing power. However, despite major advances in computer technology, contemporary HCI is still characterized by brief delays in computer responsiveness caused, for example, by background processes or network delays. Research on long System Response Times (SRTs) indicates that delays may have negative behavioral and emotional consequences. However, there are fundamental differences between previously researched long SRTs and delays as they occur in contemporary HCI, such as different timings and occurrence probabilities. Therefore the previous research is not necessarily applicable to modern HCIs. We developed a paradigm aimed at mimicking important aspects of contemporary HCI to empirically test the effects of sporadic brief delays with an average duration of 1.6s. Results showed performance decrements in the response directly following a delay in terms of increased reaction times and error rates. Furthermore, blocks in which delays occurred were less liked than blocks without delays, suggesting that delays may affect the emotional state. The data provide evidence that delays cause a significant deterioration of performance, and indicates that delays in contemporary HCI may negatively affect work productivity, work satisfaction, and health-and-safety. Suggestions for HCI design and the relationship to user interruption are discussed.

Full reference

Szameitat AJ, Rummel J, Szameitat, DP, Sterr, A. (2009). Behavioral and emotional consequences of brief delays in human-computer interaction. *International Journal of Human-Computer Studies, 67(7),* 561-570.

1 Introduction

Twenty-five years ago few people would have anticipated the tremendous processing speed of contemporary computer systems. Even though major improvements have been made in many areas regarding HCI, important issues still remain. One of those issues relates to the fact that computers still show delayed responses to user input on many occasions. Such delays may arise from network congestion or concurrently running operations such as automatic background saving, system updates, or virus scanners. Delays can severely interrupt workflow and may result in detrimental effects such as dissatisfaction, stress, or even a decrease in performance (Barber & Lucas, 1983; Guynes, 1988; Schleifer & Amick, 1989). Thus, it appears that delays in HCI can negatively affect work productivity, work satisfaction, and health–and–safety. Understanding the behavioral and emotional effects of these contemporary delays is therefore critical.

Much of the research investigating delays has focused on the so-called System Response Times (SRTs) [Footnote 1: It is important to distinguish between delays/long SRTs and interruptions. In interruption research the execution of the primary task is delayed, as in delays / long SRTs, but a secondary task also has to be performed during the interruption as well. Thus, detrimental effects of interruptions may be caused not only by the delay, but by a number of other processes such as the demand to keep two tasks in memory, to switch tasks, or to re-orient attention. Therefore we suggest that findings from interruption research are not directly applicable to research on delays / long SRTs.]. SRTs are defined as the time between an input from the user and when the computer has processed this input and is ready to receive the next input (e.g. Shneiderman, 1984). This research has shown that long SRTs, or delays, have detrimental effects on behavioral performance, can cause stress, and are perceived as aversive (e.g., Barber & Lucas, 1983; Guynes, 1988; Kohlisch et al., 1991; Kuhmann et al., 1987; Schleifer & Amick, 1989). For instance, Barber and Lucas (1983) observed in a field study that increasing SRTs from 6 to 14s resulted in less successful transactions, as well as poorer satisfaction with the SRTs, the job environment, and the online system use. Testing a cohort of 86 college students, Guynes (1988) reported that an increase in SRTs from 1.25 to 8.25s resulted in more careful and less efficient text editing. In addition, students showed higher levels of state anxiety in the long SRT condition. This is further supported by Polkovsky and Lewis (2002) who investigated different auditory waiting cues in interactive telephone systems. They found that even the shortest waiting time of 3s resulted in initial signs of perceived anxiety, stress, and impatience, with the greatest effects at the longer waiting times of up to 18s. Schleifer & Amick (1989) studied professional typists over a period of four consecutive days and found that prolonged SRTs of 3-10s duration resulted in mood disturbances such as increased frustration, impatience, and irritation. Finally, Schaefer (1990) and Butler (1982) both reported decreased behavioral performance with increases in SRT in terms of prolonged task completion time and user response time, respectively.

However, much of the research summarized above employed computers with longer SRTs than contemporary computers and therefore does not reflect the delay characteristics of modern HCI. The computer infrastructure of large companies in the 1970s and 1980s was typically based on a server-client architecture, with clients serving as text-editing terminals and mainframes serving as central storage and processing facilities. Computing, storage, and network capabilities were often not sufficient to serve all clients immediately so that client-server transactions often resulted in long SRTs, or delays. While text input or editing usually occurred promptly (because it was performed on the client), submitting the text to the server (e.g. by pressing the ENTER key after a line of input has been typed) produced a delay, or long SRT. Accordingly, long SRTs in the 1970s and 1980s usually occurred systematically and predictably.

In present HCI, however, delays have different causes and are primarily the result of multi-tasking systems which allow the parallel execution of several applications. Thus, constantly running background tasks or threads such as virus scanners, update checks, and indexing services may occasionally demand system resources, which slow down the computer and the user perceives a brief delay in the interaction. Significantly, delays in this context occur only sporadically and at random. In addition, in the 1970s and 1980s, most likely caused by the lower computing power available, SRTs were quite long, ranging from 2s up to 32s (e.g. Butler, 1982; e.g. Schaefer, 1990; Weiss et al., 1982), while present delays generally range from a couple of hundred ms up to 2 or 2.5s. Due to these differences, previous research may not be informative with respect to the effects of delays in present HCI. Therefore, the aim of the present study was to investigate the effects of delays characteristic of contemporary HCIs on behavioral performance using reaction times and error rates and on the emotional state of the user by means of questionnaires. Based on previous findings from related research on long SRTs (Barber & Lucas, 1983; Butler, 1982; Guynes, 1988; Schaefer, 1990), we expected delays of contemporary HCIs to result in behavioral performance decrements and adverse emotional effects.

1.1 Experimental Rationale

In order to investigate this type of delay we developed a paradigm which aimed to mimic the characteristics of present HCIs. Activities in modern HCI usually rely on navigating a user-interface (UI) using a mouse or keyboard. For example, when working with desktop publishing, graphics, or office software users frequently have to use menus of the UI, enter dialogues and sub-dialogues, specify values in these dialogues and confirm actions. Often, the same procedures are performed repeatedly and the performance of the user, over time, becomes highly proficient. Such performance is characterized by a smooth and fast action flow and the anticipation of action consequences. For instance, when deleting files on a computer, typically a message appears which asks to confirm the deletion. Proficient users anticipate such extra messages and may answer them with very quickly without reading them.

To simulate this situation, the paradigm developed consisted of a computer game in which participants had to navigate a character stimulus (a donkey) towards an intermediate goal (a fence) by a series of six speeded ideomotoric two-choice response tasks. Critically, such tasks are fast, easy, and fluent to perform, and should therefore reflect the proficient UI use of the expert user. Another important characteristic of contemporary HCIs is that typically a number of interdependent steps have to be fulfilled to reach a subgoal. For instance, changing the font in a text-processor involves navigating to the menu, selecting *Format*, then *Font*, then choosing the font, and finally confirming the action. Since these steps have to be performed in a fixed sequence, each step predicts (at least partially) what to do next. We simulated this interdependency of successive steps by making the next response partially predictable on the basis of the current response. Finally, we chose a computer game instead of a real-world UI navigation task because we expected our participants (mainly University students) to be more highly motivated in this type of task. This aimed to replicate the motivational state of normal everyday computer users who desire to achieve their goal.

To test whether delays, in the form of long SRTs, lead to negative behavioral and emotional effects, participants performed the game under two different conditions. In one condition there were blocks with fluent game performance (No-Delay condition). In the other condition there were blocks with occasional delays (Delay condition). The delays were implemented by briefly "freezing" the game after a response has been made. To test for behavioral performance decrements we compared reaction times and error rates firstly for specific steps in a block, i.e. steps preceded by a delay with steps not preceded by a delay, and secondly across blocks, i.e. Delay blocks with No-Delay blocks. To test for emotional effects of delays, we presented a questionnaire within each block which was based on a multidimensional model of emotion, and compared the ratings of Delay and No-Delay blocks. Taken together the present study aimed to test for behavioral and emotional effects of delays as they occur in contemporary HCIs.

2 Methods

2.1 Participants

16 participants (4 male/12 female) took part in the experiment. The age ranged between 19 and 58 (average 28, s.d. 10) years, and all participants were right handed (except for one describing himself as ambidextrous) as assessed by self-report. Prior to the experiment, all participants gave written informed consent in accordance with the Ethical Review Board of the University of Surrey. Participants received between £9 and £11 for participation (see Results section for details).

2.2 Task and Instructions

Participants had to collect a given number of target stimuli ("carrots") in a given time in each block to earn a monetary bonus. Prior to the experiment the participants were told the cover story of a computer game: a donkey (character stimulus) had to chase a carrot, whereby the carrot would always “escape”. The goal was to chase the carrot along the way until it could finally be caught at a fence (goal stimulus).

Participants received instructions verbally and on the screen, were walked through a demonstration of the game and questionnaire, and practiced the game for 40 trials in two blocks. A whole session including instruction lasted approx. 1h. Participants were informed that delays would occur in some blocks and that the delays were included on purpose and formed part of the experiment.

The experiment was programmed in PCL using the stimulus presentation software Presentation (Neurobehavioral Systems Inc, Albany, CA).

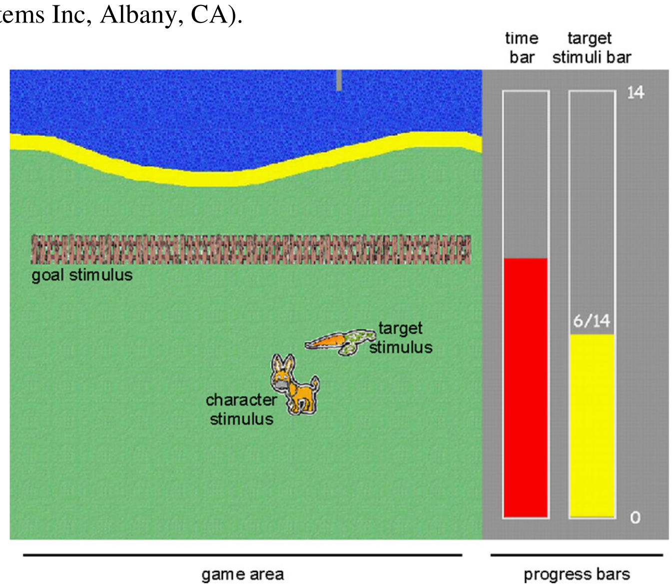

Figure 1. The display during the game play. In this example, the participant had to collect 14 carrots during the block. In this example the time bar is red, showing that the participant was lagging behind in time.

2.3 Stimulus display

For the stimulus display a 17 inch computer screen was used (visible area 30.4 (width) x 22.8 (height) cm, diagonal 38 cm). The stimulus display consisted of two parts, a game area (presented in the larger part on the left side of the screen) and an area depicting two progress bars (presented in the smaller part on the right side of the screen) (Figure 1).

The game area showed a stylized landscape with a fence as the goal stimulus and two moving stimuli, the character stimulus (donkey) and the target stimulus (carrot). The goal stimulus (fence, 20.2 x 1.5 cm) was located 13.5 cm from the bottom of the screen. The character stimulus measured 2.3 x 3 cm and the target stimulus measured 3.4 x 1.4 cm; both stimuli had black and white contours. In each trial, the character stimulus started in the center at the bottom of the game area. The target stimulus appeared diagonally either to the top-left or top-right of the character stimulus at a distance of 2 cm.

The progress bars provided feedback information, whereby they increased from the bottom to the top of the screen as the game progressed. The left progress bar showed the time passed (time bar) and the right progress bar (target stimuli bar, TS bar) showed the number of target stimuli collected in this block. The time bar increased continuously, while the TS bar increased in steps every time a target stimulus had been collected. The increment of the TS bar depended on the number of target stimuli to collect in that block and was adjusted in a way so that collecting all target stimuli would result in the bar reaching the top. If the TS bar reached the top before the time ran out participants won, whereas if the time bar reached the top first, they lost. The TS bar was always yellow, while the color of the time bar was performance dependent. If the participants were on pace to finish the game before time ran out, the bar was green, otherwise it was red. Due to the size of the bars it was easy to identify their color, and therefore feedback processing was possible with peripheral vision (Thibos & Bradley, 1991).

2.4 Procedure

The experiment can be described on three levels of granularity: steps, trials, and blocks. There were six steps in a trial, a variable number of trials in a block, and 14 blocks in the experiment. Steps consisted of a two-choice response task. Reaction times and error rates for these responses were taken as measures of behavioral performance. Six steps combined to form a trial; at the end of a trial, the participant reached the goal stimulus and collected the carrot. A block consisted of a set of a variable number of trials, dependent on individual performance, and was conducted under one of two conditions (delay or no-delay).

A step consisted of presenting the target stimulus, the response of the participant, and the actual onscreen movement of the character stimulus towards the target stimulus. If the target stimulus was presented to the top-left of the character stimulus, participants had to press the left button with the index finger of their right hand, and if it was presented to the top-right, they had to press the right button with the ring finger of their right hand on a standard 2-button mouse. If participants pressed the correct button, the character stimulus moved smoothly in a 150 ms lasting animation to the position where the target stimulus was presented. Once it reached the target stimulus, the target stimulus was immediately displayed at a new position, again either to the top-left or top-right of the current position of the character stimulus (Figure 1). This gave the impression that the carrot "jumped" away from the donkey. If participants pressed the wrong button, a short feedback sound was played and the donkey turned red for 250 ms. While the donkey was red, the mouse buttons were disabled and responses could not be corrected. Giving a wrong response was therefore disadvantageous.

A trial denoted the period from starting at the bottom of the screen until the character stimulus reached the goal stimulus, i.e. the fence. Each trial consisted of six steps. Consecutive target stimuli appeared with a higher probability (70%) at the opposite position as on the previous step, thus favoring a "zigzag" course. This latter manipulation induced the desired interdependency between the response of the current step and the response of the previous step. In other words, the response of the next step was partially predictable from the response of the current step. When the target stimulus (carrot) was collected after the last step (i.e., at the fence), a short "achievement" feedback sound was presented and the next trial started. For this, the character stimulus was presented again at the bottom of the screen and 350 ms later the next target stimulus was displayed.

A No-Delay block always lasted 90s, while Delay blocks, due to the inclusion of the delays, were longer. In Delay as well as No-Delay blocks, participants performed between 14 and 28 trials. At the beginning of each block time and TS bars were set to zero (i.e. they started at the bottom). If participants performed the required number of trials, the block ended by presenting a "congratulating" feedback sound, otherwise it ended by presenting a "failed" feedback sound. After each block, participants were informed about the monetary bonus they had won in this block and during the game so far. Before continuing with the next block, they had the opportunity for a self-paced break.

2.5 Assessment of emotional state

The effects of delays on the emotional state were tested by two different questionnaires: The In-Game questionnaire administered repeatedly, in the middle of each block, and the Post-Game questionnaire, presented only once after the game had finished. The In-Game questionnaire incorporated four questions derived from a multidimensional model of emotion, which proposes that each emotion can be described according to three underlying emotional dimensions, i.e. valence, arousal, and dominance/coping power (Osgood et al., 1975; Russell & Mehrabian, 1977; Wundt, 1905). The dimension of valence was operationalized by the question "How do you like this level?" (ranging from like to dislike; note that for the participants blocks were called levels) and the dimension of arousal by the question "How activated do you feel?" (ranging from calm to activated). During instruction, participants were explained that *activated* refers to physical activation or arousal. Finally, the dimension of coping power was assessed by the question "Do you think you will win this level?" (ranging from win to lose). Based on pilot work we further included the question "How do you feel about this level?" (ranging from annoyed to pleased).

The In-Game questionnaire was presented in the middle of each block. For this, the whole screen was cleared and the four questions were presented successively. Subsequently, the game screen was restored and the next trial started after 1s. The In-Game questionnaire was presented only between two trials, never within a trial.

The four items of the In-Game questionnaire were presented at the top of the screen one at a time. Response format was conceptualized as a computerized version of a paper-and-pencil continuous visual-analogue scale. For this, a continuous horizontal response line (length 22.2 cm) was presented in the centre of the screen, with the labels of the end points written below the left and right end points of the scale. To answer the questions, participants had to move the mouse either to the left or right, which resulted in the display of a continuous answering bar (height 0.55 cm) overlaid on the line indicating the judgment. After participants had chosen a point on the response line, they made their response by pressing a mouse button and the next question appeared immediately. Values of the rating scale were internally transformed into values ranging from -100 to +100.

The In-Game Questionnaire was presented in the middle of each block to avoid judgment of the block being influenced by the outcome of the block. Prior testing revealed that if the questionnaire is presented at the end of each block, the appraisal of the block is based mainly on whether participants won or lost.

After participants had finished the game they filled out the paper-and-pencil Post-Game questionnaire. In this questionnaire, delays were called "freezing" of the computer.

To assess the appraisal of the delays, participants were asked to give their agreement to four statements using a four-point Likert scale (strongly disagree/--; disagree/-; agree/+; strongly agree/++). The statements were "The freezing annoyed me", "The freezing frustrated me", "I liked the freezing", "The freezing made me angry", and "The game would be more enjoyable without the freezing". To assess whether participants thought the delays would influence their performance, they were asked to agree or disagree with one of the three statements "The freezing resulted in worse performance", "The freezing did not affect anything", or "The freezing resulted in better performance". The Post-Game questionnaire was administered only once per experiment and referred to the game as a whole and did not differentiate between different blocks.

2.6 Monetary bonus system

Participants earned monetary credit for won blocks (i.e., finished in time). The bonus system started at £0 and rewarded £1 per block finished in time resulting in a maximum gain of £14. Regardless of task performance participants would receive a minimum payment of £5.

Furthermore, to ensure that participants would try to be as fast as possible right from the beginning, an additional bonus was added relative to the amount of time remaining when the block was completed. For every 5s finished early, participants received £0.05. For instance, finishing 43s early would result in an additional bonus of £0.40 in that block.

The algorithm of the bonus system was implemented as follows. In the first block, participants had to collect 14 target stimuli. As this was easy to achieve, all participants won the first block of the experiment. From the second block on, an algorithm determined the number of target stimuli to be collected in the next block. In detail, based on the average time it took in each block to collect a target stimulus it was calculated how many target stimuli the participant could have collected to finish the block just in time. The number of target stimuli to collect in the next block was adjusted based on this number. As an additional condition of the algorithm the number of target stimuli was always increased at least by one if the block was finished in time and the number of target stimuli was always decreased at least by one if the block was not finished in time.

2.7 Implementation of delays

The main manipulation in this paradigm was that in half of the blocks game-flow was disturbed by occasional delays. During such a delay the computer did not process the response for some time so that the character stimulus did not move. After the delay, the procedure continued as usual; the character stimulus moved to its target position in the 150 ms lasting animation or error feedback was provided.

Duration of the delays was 1650 ms on average and randomly varied between 500 and 2800 ms to avoid anticipation of the end of the delay. In Delay blocks 12.5% of all steps were delayed, so that on average a delay occurred every 8th response. There was never more than one delay per trial.

To increase the feeling of time pressure, the time bar continued to increase during the duration of the delay, thus indicating that time was running out while participants had to wait for the end of the delay.[Footnote 2: Please note that preliminary data of follow-up studies suggest that this manipulation is not critical, since a comparable pattern of results was observed when the time bar stopped during the delay.] However, this would have resulted in a potential confound, namely that due to the delays participants would not have been able to collect as many target stimuli as in No-Delay blocks, which in turn would result in losing more Delay blocks than No-Delay blocks. In the same way, potential performance decrements caused by the delay such as prolonged reaction times may result in participants losing more Delay blocks. To circumvent both potential confounds, firstly the duration of the delays was added to the available time for finishing the block and, secondly, to account for performance decrements a further 250 ms were added for each delay. Thus, a block with 28 trials, in which 21 steps were delayed by an average duration of 1650 ms lasted 129s (90s + 21 * 1650ms + 21 * 250ms). Participants were not informed about this additional available time, and the rate by which the time bar increased was not fundamentally different to No-Delay blocks.

2.8 Design

The study was based on a 1-factor repeated measures design with two levels (Delay and No-Delay). There were 7 blocks of each condition, and the presentation order was pseudo-randomized for each participant. The randomization balanced transition probabilities between conditions and ensured a roughly homogeneous distribution of conditions along the course of the experiment. Dependent variables were reaction times (RTs), error rates, and responses from the In-Game- and Post-Game questionnaire.

3 Results

Debriefing of participants revealed that the game was well perceived and that most participants were keen on performing well. Despite participants' reported feeling of time pressure they found the game easy to perform. This fits with the fact that all participants won money in excess of the guaranteed bonus of £5, with a total of £9.98 on average (s.d. 0.685, range £9-11).

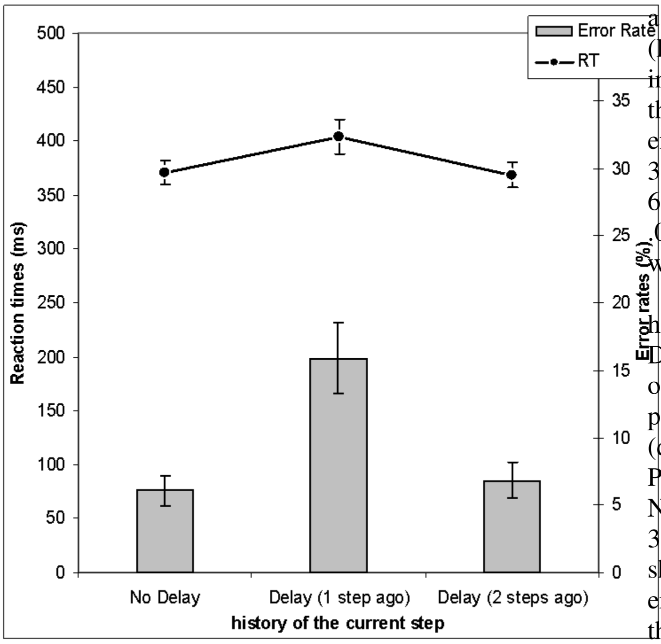

Figure 2. Local costs of delays in terms of error rates (bars and right axis) and reaction times (lines and left axis). "No Delay" refers to steps not preceded by a delay, "Delay (1 step ago)" refers to steps directly preceded by a delay, and "Delay (2 steps ago)" refers to steps preceded by a delay two steps ago. Analysis is based only on Delay blocks. Error bars denote standard error of mean (s.e.m.).

3.1 Reaction Times & Error Rates

Generally, delays may affect behavioral performance in two ways. Firstly, delays could result in local costs, i.e. performance decrements occurring directly after a delay. Secondly, delays could result in global costs, for example causing participants to be generally slower because they are more cautious.

To test whether delays result in local costs we analyzed only the steps of Delay blocks. For this, we first conducted a 1-factorial repeated measures Analysis of Variance (ANOVA) including three steps as factor levels: steps not preceded by a delay, steps directly following a delay [Footnote 3: Note that the step incorporating the delay itself cannot be affected, since the delay was induced *after* the participant's response], and steps occurring two steps after a delay. This analysis showed a significant main effect of *step* on reaction times ($F(2,30) = 11.105$; $p < .001$; partial $\eta^2 = .425$) and error rates ($F(2,30) = 19.975$; $p < .001$; partial $\eta^2 = .571$). Thus, the delay significantly affected behavioral performance.

To scrutinize these effects in more detail, we then calculated contrasts comparing the factor levels with each other. Firstly, we compared the steps directly following a delay with the steps not preceded by a delay. This analysis revealed performance decrements in the steps that directly followed a delay (Figure 2), with RTs prolonged by 33.2 ms (Delay 403.9 ms, No-Delay 370.7 ms; $F(1,15) = 11.361$; $p < .005$; partial $\eta^2 = .431$) and error rates increased by 9.8 %, (Delay 15.9%, No-Delay 6.1%; $F(1,15) = 22.672$; $p < .001$; partial $\eta^2 = .602$).

To assess the development of the disruptive effect of the delay over time we compared steps occurring two steps after a delay with steps not preceded by a delay. This analysis (Figure 2) revealed that performance decreased considerably in the first step after a delay, but returned to a normal level in the second step after a delay, indicated by non-significant effects for RTs (Delay(2 steps ago) 368.7 ms, No-Delay 370.7 ms; $F(1,15) < 1$) and error rates (Delay(2 steps ago) 6.8%, No-Delay 6.1%; $F(1,15) = 1.042$; $p > .05$; partial $\eta^2 = .065$). Therefore, the detrimental local effects of the delay were rather short-lived.

To test whether the occurrence of delays within a block has a global effect on performance we compared steps of Delay blocks which were *not* preceded by a delay with steps of No-Delay blocks. In both cases, the responses are not preceded by a delay, but the general task context is different (delays potentially occur vs. delays do not occur). Participants were generally slower in Delay blocks than in No-Delay blocks by 5 ms (Delay (No-Delay) blocks RT = 370.7 (365.7) ms; $t(15) = 2.638$, $p < .05$), while error rates showed no significant differences (Delay (No-Delay) blocks error rate= 6.1 (6.3) %, $t(15) = 1.058$, $p > .05$). Thus, already the potential of delays occurring affected behavioral performance negatively.

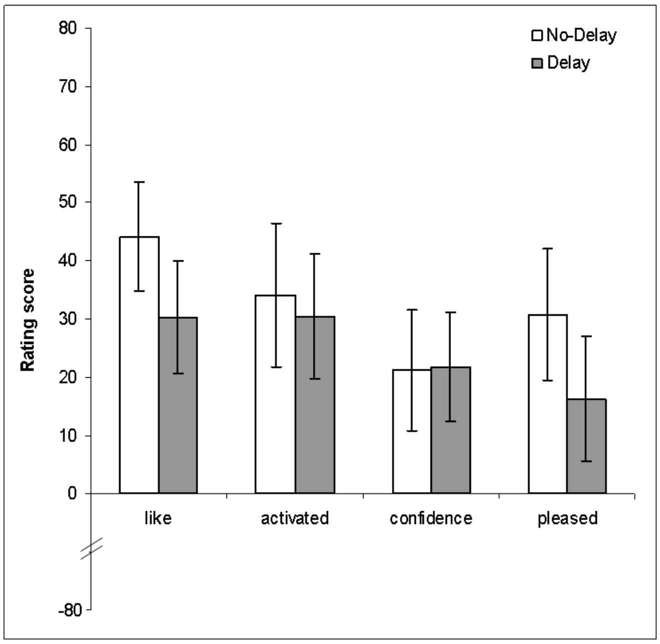

Figure 3. Rating scores of the In-Game questionnaire for No-Delay blocks (white bars) and Delay blocks (grey bars). Error bars denote s.e.m.

3.2 In-Game questionnaire

To test for the relationship between delays and the emotional state of the participants, we compared the In-Game questionnaire ratings of Delay blocks with the ratings of No-Delay blocks using paired sample t-tests. Results showed aversive tendencies towards Delay blocks (Figure 3). Delay blocks were significantly less liked than No-Delay blocks (Delay (No-Delay) 30.3 (44.2); $t(15) = 3.137$, $p < .05$ Tukey corrected for multiple comparisons) and showed a tendency to be less pleased with the Delay blocks (Delay (No-Delay) 16.2 (30.8); $t(15) = 2.470$, *ns* Tukey corrected, $p < .05$ uncorrected). However, there was no significant difference in arousal (Delay (No-Delay) 30.5 (34.1); $t(15) = .984$, *ns*) and participants were equally confident about winning the block in Delay as well as No-Delay blocks (Delay (No-Delay) 21.8 (21.2); $t(15) = .170$, *ns*), suggesting that at the time the questionnaire was presented (in the middle of the block) participants were not able to predict the outcome of the block.

3.3 Post-Game questionnaire

The Post-Game questionnaire assessed the appraisal of the delays more explicitly. Here we assumed an ordinal scale and tested for significant deviations from the middle tendency by calculating non-parametric one-sample Wilcoxon signed tests for each question. Results confirmed the findings from the In-Game questionnaire in that participants had an aversive attitude towards the delays. Participants disliked the delays (disagreement with "*I liked the freezing*"; $Z = 2.372$, $p < .05$) and were annoyed by them ($Z = 2.288$, $p < .05$), but the delays did not make them angry (disagreement with "*The freezing made me angry*"; $Z = 3.051$, $p < .01$). By trend participants agreed to the statements "*The game would be more enjoyable without the freezing*" and "*The freezing frustrated me*", but this tendency did not reach statistical significance (*enjoyable*: $Z = 1.044$, *frustrate*: $Z = 1.498$, both *ns*). Finally, participants did not assume delays would affect their performance in a positive or negative way ($Z = 1.807$, *ns*).

3.4 Blocks won

To counteract the anticipated performance decrements in Delay blocks and to ensure that Delay blocks were not lost more frequently, we prolonged the available time to finish a block (see Methods). This resulted in the fact that Delay blocks were won considerably more often than No-Delay blocks ($t(15) = 6.985$, $p < .001$). 85.7% of the Delay blocks were won whereas only 51.8% of the No-Delay blocks were won. Accordingly, on average in Delay-blocks £6 were won, whereas in No-delay blocks only £3.63 were won. Thus, participants were averse towards Delay blocks, despite the considerably higher likelihood to win money (£1 per block).

4 Discussion

This study investigated the behavioral and emotional consequences of occasional brief delays in HCIs by using a paradigm which mimicked the fluent performance of a proficient computer user. We found that brief delays of 1.6s average duration (range 500-2800 ms) resulted in detrimental behavioral and emotional effects. In particular, error rates more than doubled and reaction times were prolonged by 33 ms for responses directly following the delay. Further analyses showed that these local performance decrements were short-lived, as performance returned to pre-delay levels by the second response after the delay. Besides these local effects of the delay, steps in Delay blocks not preceded by delays showed prolonged reaction times (but not error rates) compared to steps of No-Delay blocks, but the magnitude of the global effect was smaller (5 ms). Regarding the emotional ratings, participants clearly liked Delay blocks less. When asked about the delays after the experiment, participants found the delays to be annoying and disliked them. It should be noted that the effect of disliking the delays was evident even though participants did not think that delays would affect their performance negatively (as assessed by the Post-Game questionnaire), and despite the fact that participants won considerably more money in Delay blocks as compared to No-Delay blocks. Thus, while previous findings based on much longer delays were rather inconsistent and sometimes even indicated performance increases with increasing delays, the present results demonstrate that already very brief delays can have considerable detrimental effects on behavioral performance and emotional wellbeing.

4.1 Previous research related to delays in HCI

Delays as they occur in contemporary HCI have, to our knowledge, not been investigated before (but see Monk et al., 2004, for a related study). However, research on long SRTs, mainly investigated in the 1970s until mid-1990s, is closely related to the present findings, so it will be discussed in relation to the current data.

Firstly, delays were perceived as aversive, which is in line with the majority of previous research on long SRTs (Barber & Lucas, 1983; Guynes, 1988; Polkovsky & Lewis, 2002; Schleifer & Amick, 1989). Moreover, the present results confirm and extend these findings by showing that much shorter and less frequent delays, as they occur in contemporary HCI, are perceived as aversive. Therefore, although today's computers are much faster and delays in HCI have been profoundly reduced, they continue to be a problem.

Secondly, the presently observed behavioral costs caused by the delay are in line with studies showing decreased performance with increasing SRTs (Barber & Lucas, 1983; Butler, 1982; Guynes, 1988; Schaefer, 1990). Some previous studies, however, have reported the counterintuitive finding of performance increases with longer SRTs (Bergman et al., 1981; Dannenbring, 1984; Kohlisch & Kuhmann, 1997; Kuhmann et al., 1987; Kuhmann et al., 1990; Weiss et al., 1982). We suggest that these contradictory findings are due to fundamental differences in the study paradigms. For instance, previous paradigms differ from the present regarding how the delay may disrupt the user’s workflow. In general, performing a coherent task usually results in the perception of workflow. With workflow we refer to a situation in which the user is focused and absorbed, and smoothly performs one step of the task after another, and may experience some feeling of "flow" (cf. the flow concept of Csikszentmihalyi, e.g. Csikszentmihalyi, 1998). When the user has finished one task and changes to the next, a natural break in the (work)flow may occur and delays occurring at such task/sub-task boundaries are presumably less disruptive and disturbing than delays occurring within a task or at less fundamental sub-task boundaries (cf. Adamczk & Bailey, 2004; Bailey & Konstan, 2006; cf. Kohlisch & Kuhmann, 1997; Monk et al., 2002).

In previous studies the long SRTs usually occurred at fundamental (sub-)task boundaries, such as the submission of a completed transaction (Barber & Lucas, 1983). In combination with the predictability of the long SRT, users might have used the long SRT to prepare the next task, either by preparing paperwork (Barber & Lucas, 1983) or by thinking about how to solve the next task (e.g. Bergman et al., 1981). Thus, such an enforced waiting time may well explain the increased task performance found in previous studies. However, in the present paradigm delays were unpredictable and could occur at virtually any stage of the task. Therefore, the disturbing effect of the delay in the present study may be due to the fact that delays occurred only at the most fine-grained sub-task level (after an individual step) and not, as in previous research, only at more fundamental task/subtask boundaries. It is important to note that in contemporary HCIs the situation might even be more severe, since delays may occur in the middle of a subtask where no boundary is present at all (e.g. during a mouse movement towards a target or a delay during typing a word, caused by an automatic background auto-save) [Footnote 4: Please note that further research is required to test which tasks the present results generalize to. The present task is a two-choice response tasks in which the information about the correct response is given by the spatial position of a stimulus. Such tasks involve some visuo-spatial processing and an explicit decision (the choice response) which mouse button has to be pressed. It is presently unclear whether the current findings generalize to other tasks, for example a task in which the individual responses are performed in a less explicit way (e.g. typing by a typist).]. Tentatively extrapolating our results to situations in which a delay occurs in the middle of a subtask where no boundary is present suggests that the negative effects of delays may be even more severe in real-life contemporary HCI.

A further difference of this study compared to previous studies is that in previous studies, the occurrence of long SRTs was predictable since it occurred always after completing a specific task, while it occurred only occasionally and therefore unpredictably in our paradigm. In addition, in virtually all previous studies even the shortest SRTs (which were taken as baseline to test for the effects of prolonged SRTs) had durations of 1-2 s, which would count as a long delay in contemporary HCI. This means the general expectations of a user regarding the responsiveness of the computer might have been different.

Taken together, the present paradigm (and modern HCI) may be characterized by a more fluent and flow-like performance in which the different sub-tasks neatly and smoothly merge into each other without any delay, while tasks employed in long SRT studies were more strongly subdivided into segregated subtasks with delays occurring predictably at each subtask boundary. We suggest that these differences have caused the seemingly contrary result that in long SRT studies performance often increased with increasing SRT while it decreased with the occurrence of delays in our study.

4.2 Delays and interruptions

While delays are merely a brief freezing, halting, or stopping of the computer's responsiveness, interruptions usually embody an additional need to change the task (Adamczk & Bailey, 2004; Bailey & Konstan, 2006; McFarlane, 1999, , 2002; McFarlane & Latorella, 2002; Zijlstra et al., 1999). Prototypical examples of interruptions in a standard office workplace may be alerts and notifications of incoming emails or scheduled appointments. Sometimes such notifications do not require any action by the user, but mostly the user has to acknowledge the notification, and may sometimes be forced to focus on the interrupting task (McFarlane & Latorella, 2002). Thus, while interruptions incorporate an even more complex set of cognitive problems, namely task switching, they share with delays the basic problem of a breakdown in workflow. Because of this similarity, interruptions requiring directing attention or action towards a different task may suffer at least (if not more so) from the same performance decrements and adverse emotional responses as shown for delays in the current study. Accordingly, the following discussion on design implications does not only hold for delays but may be generalized to interruptions as well.[Footnote 5: Note, however, that since interruptions raise additional cognitive problems, such as task-switching, research on interruptions is not necessarily informative with respect to the effects of delays.]

4.3 Implications for HCI design

Our data suggest that tools used for HCI, such as user interfaces, should be designed to minimize the occurrence of delays and interruptions. Recently, a trend has emerged where more and more background processes are running on standard computer workstations. While some of these processes are essential (e.g. virus scanners) there are usually a number of other processes which do not necessarily need to run constantly (e.g. indexing, checking for updates). According to our findings, it may be advisable to shift such processes to defined points in time where user disruption is minimal. For instance, a check for new updates may be performed only once a day in the morning, or indexing services only run before shutdown in the evening.

Thus, a careful review of which background processes are essentially required to be active may reduce workflow disruptions. This might comprise transferring responsibilities back from the computer to the user. For instance, the automatic background saving operations performed every 10 min by most office software packages may stay unnoticed with smaller documents on fast computers. However, large or complex documents sometimes take several seconds to save (especially when they are stored on a fileserver) and, therefore, the auto-save induces a delay every 10 minutes. Although the time to save the document is the same for auto-save and manual save, these modes may affect workflow in different ways. This is, because a manual saving operation is usually only performed at a convenient stage in task processing (i.e., probably at more fundamental (sub-)task boundaries) whereas the auto-save occurs at random stages in task processing. Manual saving should therefore be less disruptive than auto-save, even if performed at the same frequency.

Whether to choose automatic or manual modes, however, strongly depends on the nature of the computer work, particularly on the associated costs of potential errors: In some cases the loss of an unsaved document may be more serious than a slower, more error prone, and user-frustrating performance, while in other cases the opposite may be true. Therefore, it may be preferable to use minimalist systems which are designed for fast and efficient performance as standard, and to add only required further processes and features in a way that minimizes the disruption of workflow.

4.4. Conclusions

The present findings demonstrate that even small delays in the performance of computers have negative effects. From the perspective of the user our data suggests that eliminating even occasional small delays has the potential to improve work-satisfaction and well-being, which may in turn result in increased productivity and decreased stress. From the employer's perspective the increase in error rates points to a high risk of action slips which may sometimes, due to the fast and fluent performance, stay unnoticed. Thus, the present findings show that the problem of delays in HCI has not resolved itself by the mere rise of computer power. Instead, delays as they occur in contemporary HCI may even pose more serious problems than had been assumed for long SRTs years ago.

Acknowledgements

The study was funded by The British Academy, UK (SG-39619 and SG-45132; A.J. Szameitat). We thank Andrea Lynch and Phil Dean for editorial assistance.

Figure captions

Figure 1. The display during the game play. In this example, the participant had to collect 14 carrots during the block. In this example the time bar is red, showing that the participant was lagging behind in time.

Figure 2. Local costs of delays in terms of error rates (bars and right axis) and reaction times (lines and left axis). "No Delay" refers to steps not preceded by a delay, "Delay (1 step ago)" refers to steps directly preceded by a delay, and "Delay (2 steps ago)" refers to steps preceded by a delay two steps ago. Analysis is based only on Delay blocks. Error bars denote standard error of mean (s.e.m.).

Figure 3. Rating scores of the In-Game questionnaire for No-Delay blocks (white bars) and Delay blocks (grey bars). Error bars denote s.e.m.

References

Adamczk, P. D., & Bailey, B. P. (2004). *If not now, when?: The effects of interruption at different moments within task execution.* Paper presented at the Human Factors in computer Systems: Proceedings of the CHI'04, Vienna.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior, 22*(4), 685-708.

Barber, E. R., & Lucas, C. H. (1983). System Response Time - Operator Productivity, and Job Satisfaction. *Communication of the ACM, 26*(11).

Bergman, H., Brinkman, A., & Koelega, H. (1981). System response time and problem solving behavior. *Proceedings of the Human Factors Society*.

Butler, T. W. (1982). Computer Response Time and user Performance During Data Entry. *AT&T Bell Laboratories Technical Journal, 63*(6).

Csikszentmihalyi, M. (1998). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal Experience: Psychological studies of flow in consciousness* (pp. 15-36). Cambridge: Cambridge University Press.

Dannenbring, G. L. (1984). System Response Time and user Performance. *Transactions on Systems, man, and cybernetics, 14*(3).

Guynes, J. L. (1988). Impact of System Response Time on State Anxiety. *Communication of the ACM, 31*(3).

Kohlisch, O., & Kuhmann, W. (1997). System response time and readiness for task execution - The optimum duration of inter-task delays. *Ergonomics, 40*(3), 265-280.

Kohlisch, O., Kuhmann, W., & Boucsein, W. (1991). Auswirkungen systembedingter Arbeitsunterbrechungen bei computerunterstutzter Textverarbeitung: eine Feldstudie. *Zeitschrift fur Experimentelle und Angewandte Psychologie, 4*, 585-604.

Kuhmann, W., Boucsein, W., Schaefer, F., & Alexander, J. (1987). Experimental investigation of psychophysiological stress-reactions induced by different system response times in human-computer interaction. *Ergonomics, 30*(6), 933-943.

Kuhmann, W., Schaefer, F., & Boucsein, W. (1990). Effects of waiting time in simple task performance-Analogy to computer-response times in man-machine interaction. *Zeitschrift für Experimentelle und Angewandte Psychologie, 37*(2), 242-265.

McFarlane, D. C. (1999). Coordinating the Interruption of People in Human-Computer Interaction. *Human-Computer Interaction, 13*, 295-303.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction, 17*(1), 63-139.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction, 17*(1), 1-61.

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2002). *The attentional costs of interrupting task performance at various stages.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Santa Monica.

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Very brief interruptions result in resumption cost. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*.

Osgood, C. H., May, W. H., & Miron, M. S. (1975). *Cross-Cultural Universals of Affective Meaning*. Urbana: University of Illinois Press.

Polkovsky, M. D., & Lewis, J. R. (2002). Effect of Auditory Waiting Cues on Time Estimation in Speech Recognition Telephony Applications. *International Journal of Human-Computer Studies, 14*(3 & 4).

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11*, 273-294.

Schaefer, F. (1990). The effect of system response time on temporal predictability of work flow in human-computer interaction. *Human Performance, 3*(3), 174-186.

Schleifer, L. M., & Amick, B. C., III. (1989). System response time and method of pay: stress effects in computer-based tasks. *International Journal of Human-Computer Interaction, 1*(1), 23-39.

Selvidge, P. R., Chaparro, B. S., & Bender, G. T. (2002). The world wide wait: effects of delays on user performance. *International Journal of Industrial Ergonomics, 29*, 15-20.

Shneiderman, B. (1984). Response Time and Display Rate in human Performance with Computers. *Computing Surveys, 16*(3).

Thibos, L. N., & Bradley, A. (1991). The limits of performance in central and peripheral vision. *SID '91 Digest of Technical Papers (Society for Information Display, Playa del Rey, Calif., 1991), 22*, 301-303.

Weiss, S. M., Boggs, G. J., & Lehto, M. (1982). Computer System Response Time and Psychophysiological Stress. *Proceedings of the Human Factors Society, 26th Annual Meeting*.

Wundt, W. (1905). *Grundzüge der physiologischen Psychologie*. Leipzig: Engelmann.

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology, 72*, 163-185.