42 Forty Years of Research on System Response Times – What Did We Learn from It?

Wolfram Boucsein

1 Setting the Stage Before the Mid 1980s

With the advent of time sharing computer systems in the late 1960s, temporal factors came into the focus of human-computer interaction (HCI). To the present author's knowledge, NICKERSON et al. (1968) and CARBONELL, et al. (1968) were the first authors to point to the psychological importance of involuntary delays in HCI. They concluded that computer-system originated waiting times should either be rather short, thus preventing the work flow from being interrupted, or be long enough to allow for the so-called job swapping which means executing another task during the waiting period.

On the background of communication theory, a similar point was made by MILLER (1968), who recommended upper limits of 2 seconds for intratask- and 15 seconds for intertask-computer elicited waiting times to prevent a breakdown during the HCI dialogue. Otherwise, the human operator should be allowed to behave as a time-sharing system, switching to other tasks during the interruption. These early insights are still valid and more recently discussed as determining optimal system response times (BOUCSEIN 2000) and under the concept of multitasking (SCHAEFER et al. 2000).

The term "system response time" (SRT) was introduced by WILLIGES and WILLIGES (1982) and SHNEIDERMAN (1984) who divided the human operator's behavior in alternating "operant" and "respondent" parts. These terms – taken from psychological learning theory – were used to classify actions during HCI depending on who initiates them: the human (operant) or the computer (respondent). The time between the termination of the user's input and the computer's "prompt" was labeled SRT.

System response times are of particular importance for HCI because of their stress inducing properties (for a summary, see BOUCSEIN 2000). Without providing empirical data, PETERS (1977) pointed to observations of an increased heart rate (HR) during waiting times in HCI. MARTIN (1973) did not only focus on the

C.M. Schlick (ed.), *Industrial Engineering and Ergonomics*,
DOI 10.1007/978-3-642-01293-8_42,

length but also on the variability of SRTs, the latter supposedly being a source of uncertainty and thus possibly stress inducing. His recommendation, though also not empirically backed up, was that the standard deviation should not exceed half of the mean SRT. In addition, Martin pointed to the importance of task complexity for the tolerability of SRTs. This had been already pointed out by SACKMAN (1970) who reported that operators became rather impatient if SRTs for simple tasks exceeded 10 seconds and/or were irregular, whereas if an operator assumed that the computer was currently busy with handling a huge amount of data, SRTs up to 10 minutes could be well tolerated. Cognitive psychologists later called this the "model which an operator has from the computer" (SHNEIDERMAN 1984).

2 Psychophysiological Stress-Inducing Properties of System Response Times

Based on psychological stress theories, BOUCSEIN et al. (1984) set the stage for a comprehensive research program on the possible stress inducing properties of mean duration and variability of SRTs, based on the following two general hypotheses:

(1) Any increase of mean SRTs beyond a well tolerated time span will induce stress by means of anticipating negative consequences such as overtime work or salary reduction (GREIF 1983). This hypothesis was probed using SRTs of different lengths.

(2) A continuous work flow which is interrupted by intervals of unpredictable length is likely to produce uncertainty, which is also a source of psychological stress (MONAT et al. 1972). This hypothesis was probed using SRTs of different variability.

Since stress induced by SRTs could be expected to show up instantaneously, measures of the autonomic nervous system would be the adequate psychophysiological responses to look for (BOUCSEIN 1988, 2006, LUCZAK 1987). Taking samples of hormones such as catecholamines and cortisol would be more suitable for determining long lasting phenomena such as the inability to unwind after work (MELIN & LUNDBERG 1997).

Between 1982 and 1997, the present author's group performed a series of systematic laboratory studies on the stress inducing properties of SRTs. Similar to the type of assessment of HCI chosen by CARD et al. (1983), a rather simple computer task was chosen to permit a close control of task variables in relation to SRTs. A space surrounded by identical letters had to be targeted within a line of otherwise randomly generated letters and spaces, which was presented in the center of a visual display. Only two studies (THUM et al. 1995, SCHAEFER & KOHLISCH 1995) used a more complicated task to individually standardize mental strain. In our studies, SRTs were varied systematically between 0.5 and 8 seconds. Additional time pressure was induced by using incentives for working as fast and as correctly as possible or by announcing overtime work in the case of incorrect solutions.

As autonomic nervous system measures, we applied skin conductance level (SCL), frequency and mean amplitude of nonspecific electrodermal responses (NS.EDRs), systolic and diastolic blood pressure (SBP and DBP), mean arterial BP, HR, heart rate variability (HRV), and respiration rate. In addition, frontalis electromyogram (EMG) was recorded as another indicator of short lasting stress responses. As performance measures, we used task completion time, numbers and relevance of keystrokes, cursor movements and errors, whereas ratings of mood and bodily symptoms were applied as subjective measures.

Table 42.1. Results from six studies on System Response Times: a: SCHAEFER et al. (1986), N= 20; b: KUHMANN et al. (1987), N= 68; c: KUHMANN (1989), N= 48; d: KUHMANN et al. (1990), N= 24; e: THUM et al. (1995), N= 40; f: KOHLISCH & KUHMANN (1997), N= 42. EDR freq. = EDR frequency; EDR amp. = EDR amplitude. (after BOUCSEIN (2000), Table 14.1)

SRT	Parameter category	Incentive (time pressure) present	Incentive (time pressure) absent
short (0.5 – 2 sec)	physio-logical	SBP [b, e] ,mean arterial BP [f] and/or DBP [e] increased HRV decreased [e] Respiration rate increased [e] EMG frontalis power in-creased [e] NS.EDR freq. increased [f]	SBP and DBP increased [a] HR increased [d]
	behavioral	Error rate increased [b, e, f] Number of cursor move-ments [b], working speed [e] and irrelevant keystrokes [f] increased	Error rate increased [a, d] Working speed increased [d]
	subjective	More headache and eye-related symptoms reported [b]	General arousal increased [d] Number of task related symptoms increased [a]
long (8 sec or longer)	physio-logical	SCL increased [b] NS.EDR freq. increased [b] NS.EDR amp. increased [e] SBP and DBP decreased [e] Respiration rate decreased [e]	Increasing amount of EDA during SRT [c] NS.EDR freq. initially increased but decreased later [d]
	behavioral	Error rate low [b, e]	Error rate low [a, d] Number of cursor move-ments increased [c] Task completion time prolonged [c]
	subjective	Negative motivation and emotion increased [e] More symptoms of head-ache reported [f]	General well being en-hanced at the beginning of work [c]

2.1 Effects of SRT Duration and the Concept of an Optimal SRT

Our results from six laboratory experiments with altogether 242 subjects are summarized in Table 42.1 (taken from BOUCSEIN 2000, Table 14.1). Contrary to our hypotheses, our general findings were that negative physiological, behavioral, and subjective consequences resulted not only from rather long but also from rather short SRTs. If in case of the latter, incentives for working fast such as time pressure were present (top of the third column of Table 42.1), our subjects developed considerable cardiorespiratory strain, i.e., increases of SBP, DBP, and respiration rate, plus a decrease in HRV. In addition, there was an increase of EMG frontalis activity and of NS.EDR frequency (NS.EDR freq.). The subjects showed an increase in error rate, cursor movements, and irrelevant keystrokes as signs of impaired performance. They also reported more subjective complaints which are typical for computer workplaces. When no time pressure was imposed on the subjects under the short SRT conditions (top of the fourth column of Table 42.1), physiological and subjective effects were less computer-task specific, and only rather slight increases in SBP, DBP, and HR were observed. However, there was an increase of general subjective arousal and the number of task-related bodily symptoms. Despite not being instructed to do so, our subjects spontaneously increased their working speed, the consequence of which was an increase of error rates. Since our subjects were not experienced computer users, our impression was that they may have felt rushed by the tasks being presented too fast for their personal tempo.

When SRTs were rather long, a different picture emerged. At first glance, it looked as if rather long SRTs simply resulted in a more relaxed and careful work style, compared to the case when incentives for working fast were present, e.g., time pressure (bottom of the third column of Table 42.1). Subjects made fewer errors, presumably because error correction was much more time consuming. However, any increased relaxation was perceived rather than real, because subjects experienced their working situation as uncomfortable, which is shown by an increasing number of negative emotions and headache symptoms. In addition, an increase in SCL, NS.EDR frequency, or mean NS.EDR amplitude suggested an augmentation of emotional strain. Results without incentives, e.g., without time pressure (bottom of the fourth column of Table 42.1) were similar with respect to a low error rate. However, subjects used their resources less effectively, increasing the number of cursor movements and the time needed for task completion. Furthermore, an enhanced subjective well-being, which had been reported at the beginning, faded during the course of work. At the same time, emotional tension developed as indicated by an increasing amount of EDA during SRTs with duration of 8 seconds (KUHMANN 1989). The decrease in EDA after an initial increase reported by KUHMANN et al. (1990) corresponds to an application of SRTs within a single task instead of SRTs between tasks. In general, effects of long SRTs on

EDA are more pronounced under time pressure, compared to conditions with no incentives for working fast.

We consider an increase of workload due to high work density a possible reason for the stress-inducing effects of short SRTs, which is in accordance with elevated cardiorespiratory activity (top of the third and fourth columns of Table 42.1). However, since increased cardiorespiratory activity can be regarded as an indicator for physical strain (BOUCSEIN 1993), it should be ruled out that our results were mediated by an increased amount of physical activity required under short SRTs, whereas long SRTs might have imposed less motor activity on our subjects. This was probed by varying motor demands and mental load as independent factors in two experiments performed by KOHLISCH & SCHAEFER (1996). The task used in the first study was to use compensatory keystrokes for keeping a cursor moving towards the left margin of the screen within a target area in the middle of the screen. They experimentally varied the required motor activity in four steps by manipulating the cursor speed, using inter-keystroke intervals of 150, 300, 600, or 1,200 milliseconds, respectively. Mental load was varied by narrowing or widening the target area, thus changing the required accuracy for keystroke synchronization. A completely balanced within-subjects design was applied, in which 24 subjects performed a training of 20 seconds and a test trial of 90 seconds for each combination of conditions. Besides mean inter-keystroke intervals, they recorded heart period (instead of HR) and NS.EDR freq. as psychophysiological measures. When the mean inter-keystroke intervals were 300 milliseconds or longer, cardiovascular and electrodermal measures remained unaffected by motor activity which is in accordance with CARRIERO'S (1975) finding that tapping below 333 milliseconds does not exert considerable influence cardiovascular and electrodermal variables. A second study with 42 subjects confirmed this result. In this experiment, an additional arithmetic task was applied for the induction of mental load. Again, no significant effects of motor activity on psychophysiological variables were found below 360 milliseconds. In HCI, the reported range for mean inter-keystroke intervals lies between 891 milliseconds and 5.01 seconds (RAUTERBERG 1992), with a mean of 453 milliseconds for data entry tasks (HENNING et al. 1989). Because the intervals had been as high as 880 milliseconds in the KUHMANN et al. (1987) study, an increased motor demand cannot account for the psychophysiological changes observed under short SRTs. Therefore, an increased mental workload under short SRTs is the most likely explanation for the observed cardiorespiratory stress effects.

To further probe the importance of mental workload for psychophysiological effects of SRTs, THUM et al. (1995) performed a study with 40 subjects. They kept metabolic demands due to mental workload constant while varying SRTs, by using an adaptive computer task developed earlier by KUHMANN (1979). This task consisted of a randomly generated 6 × 6 matrices of 36 two-digit numbers where the subjects had to decide whether one, both, or none of the two target numbers were present. An algorithm continuously varied the presentation time of the matrices, becoming shorter after a correct response and longer after two subsequent mistakes, thereby ensuring that all subjects achieved the same percentage of correct responses. After each single task, feedback was given to the subject. The first of three 7-minute trials was performed with 1.5 seconds SRTs, the two others with 0.5 seconds and 4.5 seconds in a counterbalanced order. Half of the subjects were given monetary incentive for exceeding a certain performance level. HR, SBP, DBP, respiration rate, EDA, and frontalis EMG were recorded, and after each trial, task features (e.g., difficulty in comparison with a reference trial) as well as emotional states during the task were rated. Compared to the medium SRT of 1.5 seconds, both short (0.5 seconds) and long (4.5 seconds) SRTs significantly increased HRV while HR was reduced. In addition, short SRTs increased DBP, while long SRTs increased NS.EDR amplitude and reduced SBP and respiration rate. Compared to 0.5 seconds, the 4.5 seconds SRTs increased the NS.EDR amplitude and the HRV, but reduced SBP, DBP, and respiration rate. This complex outcome of the THUM et al. (1995) study shows that SRTs being shorter and longer than a certain medium range may induce considerable psychophysiological strain. Therefore, effects of short SRTs on HR and HRV as displayed in the upper part of Table 42.1 may be due to mental workload since they no longer appear when this factor is controlled.

The adverse effects of rather short SRTs determined by our research group have been challenged by SCHLEIFER and OKOGBAA (1990). In their own experiment, 45 female professional typists performed an alphanumerical data entry task in response to a series of prompts displayed on a VDT screen on 4 consecutive days. Their subjects worked either under short and invariable (350 ms) or long and variable (3–10 seconds) SRTs. Half of their subjects received a monetary incentive for working faster and a penalty for making more errors compared to an individual baseline. Results indicated that time pressure imposed by monetary incentives led to increased cardiovascular strain (increased SBP and DBP, decreased HRV) during the last 2 days of their experiment. However, no significant effects of SRTs and no interactions between SRT and incentive conditions were found in their psychophysiological measures. One possible explanation could be the difference in subjects between our studies and the Schleifer and Okogbaa one: Their typists may have been acquainted with the combination of short SRTs and monetary incentives, which was not the case for the student subjects used by our group. Additionally, there was no chance for the adverse effects of long SRTs found by our group in electrodermal measures, since Schleifer and Okogbaa restricted themselves to cardiovascular measures.

Provided that various psychophysiological studies have shown detrimental effects of very short SRTs, the general recommendation "faster is better" (see the title of a paper by SMITH 1983) does not seem to be appropriate for HCI, except for programmers (LAMBERT 1984). Instead, our results support a concept of an optimal SRT, which is in accordance with the notion of SHNEIDERMAN (1992, p. 277) that there is a preferred SRT for a given user and task, and that both shorter and longer SRTs may generate debilitating effects on the effectiveness of HCI. Based on the research performed by our group, our recommendation in the mid

1990s was to determine an optimal SRT for each task in question. The criteria for such an optimal SRT were considered:

- no marked increases in cardiovascular activity,
- low NS.EDR freq.,
- no increased general muscle tension,
- best performance, and
- lowest reports of pain symptoms.

With the particular task and kinds of subjects used by our group, the optimal SRT turned out to be 5–6 seconds (KUHMANN 1989). Presumably, the optimum would be shorter for experienced computer users who are familiar with the task, such as the typists used by SCHLEIFER and OKOGBAA (1990), but also for other kinds of tasks (e.g., about 1.5 seconds for the task used by THUM et al. 1995). Optimal SRTs could well be much longer, as was the case in the BARBER and LUCAS (1983) study, where the optimal SRT was 12 seconds for establishing telephone circuits in interstate networks.

KOHLISCH and KUHMANN (1997) exemplified how such an optimal SRT for the task in question should be determined. Their 42 subjects performed the detection task, as applied by our group so far, for 10 minutes under three conditions in a balance repeated measurement design. SRTs were either 1 second (i.e., short), 5 seconds (i.e., medium), or 9 seconds (i.e., long). An extra amount of unpaid overtime work was announced if the subjects' performance would not meet a combined speed-accuracy criterion checked by the computer. This was introduced to provide an incentive for working as fast and correct as possible. Besides recording speed and accuracy, HR, BP and EDA as well as subjective ratings of mood, arousal and bodily symptoms were obtained. As a result, the percentage of errors, the amount of irrelevant keystrokes and mean arterial BP were significantly increased under short SRTs compared to the other two conditions. Unexpectedly from the results summarized in Table 42.1, NS.EDR freq. was significantly higher under the short, compared to the long SRT condition. The authors concluded that the 1-second SRT was stress and error inducing. Since the amount of subjectively reported headache was significantly higher under SRTs of 9 seconds compared to those of 5 seconds, KOHLISCH and KUHMANN (1997) considered 5 seconds as an optimal SRT for the specific combination of task and subjects.

Attempts to theoretically explain the adverse effects of prolonged SRTs date back to MILLER's (1968) review on response times in HCI. Based on communication theory, Miller considered a breakdown of the human computer dialogue a possible cause for stress-inducing properties of rather long SRTs. As a consequence, he recommended an upper limit of 2 seconds for intra-task and 15 seconds for inter-task SRTs. Another possible explanation is that computer users may become concerned with prolonged SRTs since the penalty for errors increases because of slowing down their work (SHNEIDERMAN 1992). In addition, prolonged SRTs have been found detrimental for the user's motivation and work satisfaction (e.g., SCHLEIFER & AMICK 1989, THUM et al. 1995, TREURNIET et al. 1985). Both may in turn contribute to performance decrements.

2.2 Effects of Variability of SRTs and Reducing Their Stress-Inducing Effects

Since temporal uncertainty may also be an important stress inducing factor in HCI (BOUCSEIN et al. 1984), it has been suggested that not only the length but also the variability of SRTs may induce strain by a mechanism of inducing uncertainty concerning the temporal course of HCI (CARBONELL et al. 1968). Unexpectedly, no dramatic effects were found when considerable SRT variability was introduced (see SHNEIDERMAN, 1992, Chap. 7.6, for a summary). For example, in two studies that introduced zero or 80% variability of SRTs as a second experimental factor besides 2 and 8 seconds SRT duration, neither significant main effects of variability nor interactions with duration were found (KUHMANN et al. 1987, SCHAEFER et al. 1986). To resolve this issue, another study was performed by SCHAEFER (1990) with 48 subjects. After 20 minutes training with 1 seconds SRTs for all subjects, four independent groups performed the detection task used earlier by our group with 2, 4, 6, or 8 seconds fixed SRTs for another 20 minutes. The following six blocks of 20 minutes each used a so-called stop-reaction time procedure. This means a sequence of events presented with constant intervals is halted from time to time. The subjects' task is to press a key as soon as they notice such an unpredictable event. Subjects were instructed that system breakdowns would occur in irregular sequence instead of the ordinary SRTs, and that a "restart key" would enable them to continue their work. To prevent subjects from speeding up their work by pressing this key, the restart needed 8 seconds. As a result, no significant effects of SRTs on HR, or SBP and DBP were found. The plot of mean stop-reaction times against SRTs fitted a positively accelerated power function with an exponent of 3.39. An alternative linear model had to be rejected, apparently due to the marked increase of stop-reaction times under the 8 seconds SRT condition. Similar results were obtained for the standard deviations of the stop-reaction times. This points to a deterioration of temporal sensitivity for SRTs with durations between 6 and 8 seconds, since means and standard deviations of stop-reaction times were in agreement with Weber's Law up to 6 seconds but no more with 8 seconds SRTs. An impaired predictability resulted in performance decrement under the 8 seconds SRT condition, where subjects needed more time to complete their tasks because more cursor movements were performed. As a consequence, the failure of our earlier studies to determine a separate effect of SRT variability was re-interpreted insofar, as temporal uncertainty could have been the critical factor for the adverse effects not only of the variable but also the constant 8 seconds SRT condition. The reason is that subjects may have not been able to subjectively distinguish between those conditions.

To address the question, how predictability of SRT duration influences the working behavior, KUHMANN and SCHAEFER (2007) conducted a study with 50 subjects who performed the error detection task used earlier by our group. The target to be identified and marked with the cursor could occur in 10 different positions. SRTs were either coupled with the target position and thus predictable, or randomly distributed, thus making SRTs unpredictable. In another condition with predictable SRTs, the part of the text line on the left side of the target remained visible on the screen during SRT, thus indicating its actual duration. In 20% of the tasks, presenting the subsequent task was delayed until the subjects pressed a "restart key" as in the stop-reaction time procedure introduced by SCHAEFER (1990). However, pressing the "restart key" before the SRT was expired resulted in an unproportional long additional waiting time, which considerably prolonged the working time. Because of the importance of correctly estimating the SRT in this setting, the stop-reaction task consisted a vehicle to access the subjects' internal representation of SRTs. Results showed that stop-reaction times under the predictable condition corresponded to the target positions of the preceding task, while they were uncorrelated with the previous target positions in the unpredictable condition and often exceeded the longest SRT. Both strategies, differential responding and responding to the longest SRT, indicate that subjects tried to optimize their total amount of working time. However, differential responding which occurred under the predictable condition was associated with additional cognitive work load compared to the unpredictable condition. The KUHMANN and SCHAEFER (2007) results strongly support the use of displaying the actual progress of the computational task in case of unpredictably variable SRTs, instead of giving an estimate for the time until task completion as used in most PC based software systems. However, in case of predictable SRTs, it is advised not to indicate elapsed or still to be expected parts of actual SRTs, since users are able to detect regularly interspersed SRTs by implicit learning and optimize their work flow according to the system's temporal behavior.

The observable detrimental effects of unpredictable SRTs during mental tasks may be partly due to the fact that SRTs are normally used for preparing a subsequent work step. Therefore, SRTs can be confusing if their anticipated duration does not correspond to their actual duration. In general, it will be more likely for preparatory-related problems to appear in inter-task SRTs than in intra-task SRTs. Therefore, an intra-task SRT paradigm was used by SCHAEFER and KOHLISCH (1995) to determine the kind of mental processes during SRTs by means of EEG recordings from Fz and Pz. Their 54 subjects performed a series of 360 target recognition tasks which served as a model for database retrieval. After presentation of an artificial target word on the screen, SRTs of 1, 2, or 4 seconds were imposed on one third of the subjects each, prior to the appearance of a recognition list. The match of SRTs with the subjects' expectation was varied as an additional within-subjects factor as follows: The standard SRT was reduced by 25% in 30 trials and increased by 25% in another 30 trials, which were randomly interspersed between standard trials. For each condition, latencies and amplitudes of N100 and

P300 following SRT termination were obtained. The N100 and P300 latencies were larger and recognition time was longer when SRTs were unexpectedly short, whereas P300 amplitudes were increased and working speed was reduced for the subsequent task when SRTs were unexpectedly long. In addition to the demonstration of distortions during information processing, this result stresses the importance of a stable and predictable SRT in HCI.

Intra-task SRTs were also probed by KUHMANN et al. (1990), using 24 subjects, who worked on a modification of the detection task used earlier by our group. Two rows of random letters with spaces as used in the other studies were combined into one task, after which the subjects had to report whether a target (i.e., a space surrounded by two identical letters) was detected or not in either of the subsequently presented lines. After 15 minutes training with constant 1 second intra-task SRT between the two lines, subjects performed four 10 minutes trials in a counterbalanced order with intra-task SRTs of 2 or 8 seconds, crossed with zero or 80% variability as a second factor. Again, SRT variability did not have any effect on physiological and performance measures, but there was significantly less subjective well being and more symptoms of general excitement in variable compared to constant SRTs. In accordance with the results from inter-task SRTs, cardiovascular activity (i.e., HR) was increased while working under short intra-task SRTs. However, an increase in NS.EDR freq. observed under long SRTs at the beginning reversed during the course of the trials. In accordance with our hypothesis that detrimental effects of SRTs would show up if they were introduced within a task, SRTs mainly affected performance in the second part of the task. Under short SRTs, subjects worked significantly faster on the second line but their error rate increased as well. Performance decrements and physiological as well as subjective strain are likely to occur not only when SRTs exceed a certain optimal range but also when SRTs are too short for the task in question. This has been found for both, intra-task SRTs and inter-task SRTs.

In accordance with our second general hypothesis, HOLLING and GEDIGA (1987) formalized a theoretical approach based on the cognitive stress theory of Lazarus (MONAT et al. 1972), using a probability model for determining the course of strain development during SRTs. According to a proposal made by CARBONELL et al. (1968), they assumed a linear non-decreasing function for psychological costs, e.g., the amount of anger associated with SRT duration. The subjective expectation of SRTs was regarded as a probability distribution. As a result, the integrated conditional expectation of the future costs allowed for a precise prediction of the time course for developing strain during the actual SRT.

To analyze the predictions of the model, Holling and Gediga performed a laboratory study with 72 subjects who worked for 6 trials of 12 minutes each on the simple detection task used by our group. Four experimental conditions made up from 2 or 8 seconds mean SRT duration crossed by constant or variable SRTs were applied in a counterbalanced order. Heart rate and BP were recorded, and ratings of subjective stress, weariness and anger were obtained after each trial. Their results supported the strain increasing property of SRT variability and stated that 8 seconds SRTs were more stress-inducing than those of only 2 seconds duration. However, the assumption of a linear cost function had little predictive power for the course of strain developing during SRTs.

A possibility for the reduction of stress induced by long and/or unpredictable SRTs may be providing feedback on the status of the task in progress. In a laboratory study with 48 subjects whose task was to correct errors detected by the computer in seven files consisting of 100 forms each, HOLLING (1993) applied two versions of the task already used by HOLLING and GEDIGA (1987) in a counterbalanced order. One version gave no information about the status of the task, while the other one continuously displayed the number of the form which was actually being checked by the computer. Standardized differences between the measure of strain, developed by HOLLING and GEDIGA (1987) under both conditions were computed for EDA (SCL) and HR. Stress detected by EDA was significantly lower when feedback was given, paralleled by a decrease in subjective strain measures of anger and arousal, whereas feedback did not exert a significant effect on HR.

Besides using indicators as feedback for the progress of an actual SRT, their stress-inducing properties may be reduced by allowing the user to behave like a time sharing system – a suggestion made as early as by MILLER (1968). In a preliminary study performed by SCHAEFER et al. (2000), 48 students worked on a mocked power plant control center setup. A panel showed 36 displays to be set to particular initial values, some of which were to be requested from a virtual remote data bank by transmission lines. During the processing times for these requests, which were 10 seconds for half of the subjects and 30 seconds for the other half, the setup procedure could be either continued or additional requests could be started. As an additional within-subjects factor, feedback on the progress of opened requests was either not provided or given as a static or a dynamic display. EDA, ECG, peripheral blood volume and respiration were continuously recorded, and keystrokes and request handling were written in log-files. Subjective ratings were taken after each feedback condition. Results showed that the multi-tasking features were more often used when the processing times were long, a condition under which subjects also reported more arousal and pain symptoms. For short processing times, both frequency and amplitudes of NS.EDRs significantly increased, indicating increased emotional strain. Cardiovascular measures did not yield significant effects. These results mean that the psychophysiological patterns seems to be reversed compared to the former results of our group with single-task systems. We concluded that under multi-tasking conditions, long SRTs seem to be more convenient since they may be used for the performance of other work steps, which is not the case with short SRTs. It is also a possibility that EDA may have indicated the inappropriateness of using multi-tasking procedures during too short processing times.

An additional resource for coping with stress-strain processes induced by SRTs may be using the multi-tasking capabilities of modern computer systems. Instead of merely waiting for the computer to respond, the user may decide to work on several processes simultaneously. However, if SRTs do not exceed a certain duration, the benefit from switching to another task instead of waiting may be more than outweighed by the additional mental load that results from scheduling (the term scheduling refers to the need for organizing the work flow of different tasks running in parallel). Maintaining an optimal scheduling is an additional mental challenge for the user and thus becomes another source of stress. Coping with this type of stress can be facilitated by providing appropriate feedback on the temporal aspects of processes running in the background while a main task is performed. Therefore, process indicators play an important role in multi-tasking systems.

The aim of the next two studies (BOUCSEIN et al. 1998) was to establish the relationship between the duration of SRTs (i.e., the duration of processes running in the background), the type of feedback given by process indicators, and the stress-reducing properties of multi-tasking. We expected the stress-reducing features of multi-tasking to become more prominent with increasing background process duration. Furthermore, we expected the stress imposed by the need for scheduling to be reduced if the temporal flow of processes running in the background is visualized by means of process indicators. Two types of process indicators were applied: a static and a dynamic-relative indicator. In order to create a realistic environment for multi-tasking, a simulated computer aided design (CAD) task was programmed permitting two different background processes. The task consisted of a series of three pages with 12 screws, each of which was to be calculated and inserted. Users were asked to get a screw symbol with the mouse from a graphic board, insert it at the predetermined position and measure the required length by using two mouse clicks. Background process 1 was started by entering the numbers for thickness of material and required strength for the screw in a separate window. The results of this process were the minimal length and diameter required for the screw. Those numbers were entered in another window for background process 2 that resulted in the norm length and diameter of the screw fitting for the particular purpose. Finally, the numbers for length and diameter were entered in a window in order to generate the selected screw, and the screw was draught to its place with the mouse.

A first study with 18 male engineering students was performed to determine the minimal background process duration for a successful multi-tasking. Three different values for the process duration (10, 20 and 40 seconds) were applied in counterbalanced order. Electrodermal activity, HR, EMG recorded from the neck, BP, respiration rate, and electrooculogram were continuously recorded, together with different working strategy parameters such as mouse movements, processes in use and activity breaks. Subjective ratings were taken after each CAD page. Covariance analyses were applied after eliminating general habituation effects over trials. Psychophysiological parameters did not yield significant effects, presumably because of the different work strategies used during the three different levels of processing times as could be shown by strategy parameters. Besides causing longer working periods, prolongation of background process duration times resulted in less arousal and more emotional strain. The highest error rate and the strongest self-rated bodily pain symptoms occurred with processing times of 20 seconds. Our conclusion was that the 20 seconds background process duration might have the most benefit from introducing process indicators.

As a consequence, the background process duration of 20 seconds was used in a second study with 42 male engineering students. Three different experimental conditions were applied in counterbalanced order: A process indicator (an hourglass showing up until the background process was done), a dynamic-relative process indicator (a horizontal bar filling from left to right proportional to percent-done of the background process), and no process indicator as a control condition. We expected the dynamic indicator to further reduce mental strain, but possibly increase emotional strain because of its property to push towards the end of the task in progress. Psychophysiological, subjective, and performance measures as well as the statistical evaluation were the same as in the first study. The presence of process indicators accelerated working speed and cardiorespiratory activity, regardless of the indicator type. The rate of utilization of the multi-tasking features increased significantly with the presence of process indicators: In only 53% of the time there was no background process active as compared to 56% under the no-process-indicator condition. This rate of utilization difference was better reflected in subprocess 1 than in subprocess 2 where the significance was only marginal. However, only in less than 2% of the time both subprocesses were simultaneously activated. The only physiological measures affected by the introduction of process indicators were the systolic blood pressure and the respiration rate, both increasing significantly. In the subjective domain, anxiety/depression decreased slightly but significantly. On the other hand, there was a marginal significant decrease in subjective well being with increasing complexity of the process indicators. The psychophysiological changes observed can be easily explained with the increase in performance.

In general, there were not many differences between physiological measures under the different experimental conditions in both studies. Instead, we observed marked differences in various measures of working strategy. Our interpretation is that our subjects had so many degrees of freedom in the CAD task that they were able to compensate for both kinds of possible stress inducing factors (mental and emotional) by changing their working strategies accordingly. In our former work on system response times, we used highly determined task sequences with almost no degrees of freedom. Therefore, the present results do not directly compare since the CAD task used allowed for multiple degrees of freedom. Those may be used as an additional resource for preventing stress caused by adverse factors in the work flow such as forced waiting periods. The introduction of process indicators increased working speed and accordingly cardiorespiratory activity. However, there is no increase in neck muscle tension and no stress relevant change in electrodermal or subjective measures. Instead, subjective anxiety and depression show a small though significant decrease.

We concluded that using multi-tasking computer systems may increase the amount of work and enhance performance but does not seem to increase psychophysiological stress. Moreover, the kind of process indicator used does not really matter, although the presence of such indicators shows superiority over the absence of process indicators. It is concluded that multi-tasking may facilitate action regulation as an additional resource for coping with psychophysiological stress during HCI.

3 System Response Times – A Persisting Problem

Despite the comprehensive research program that our group has performed and published for more than 15 years (with an extension to recently), the computer software engineering community almost completely disregarded the results from psychophysiological research on SRTs. In the field of designing interfaces, results from the so-called cognitive psychology still prevail, focusing on time estimation and perception, attention memory, and user rated satisfaction (e.g., SHNEIDERMAN & PLAISANT 2005). There are, however, a few exceptions.

HÜTTNER et al. (1995) stated that SRTs no longer contain a problem with stand-alone personal computers, however, are still of great relevance in case of access to a central unit within a network. A very common mistake in designing such systems is to minimize SRTs as much as possible. A much better strategy would be to avoid SRTs becoming too variable (the standard deviation should be less than half of the mean), and to provide a preliminary instantaneous response to inform the user about the processing being started. Especially when SRTs exceed 30 seconds, an analog indicator of the system progress is required to avoid adverse effects. However, analog indicators that do not really correspond to the time to be elapsed as seen in various Windows-based applications can not be trusted and will hence be prone to again produce adverse effects.

With the advent of the World Wide Web, SRTs became everyone's "daily bread". As a consequence, the issue of possible detrimental effects of SRTs on the course of HCI and its outcome gained some importance for the design process of interfaces (JACKO et al. 2000). Since SRTs consist an unavoidable by-product of interactions between multiple users – as they already did for multiple terminals connected to a single mainframe in the late 1960 – one does not have to be a prophet to predict that SRTs will continue to be a significant issue in designing interfaces for any kind of HCI (POLKOSKY & LEWIS 2002).

In a study with 26 subjects, grouped into skilled and unskilled Web users, TRIMMEL et al. (2003) found an increase of psychophysiological strain with prolonged SRTs. Their subjects performed three Web tasks with 2, 10 and 22 seconds SRT duration. Regardless of their level of experience, subjects who rated their subjective stress level being heightened emerged higher HRs compared to those with a lower subjective stress level, and their HRs increased with the duration of SRTs. The authors concluded that uncertainty about the outcome of the S's last action could be seen as a main source of an increase in psychophysiological strain. Hence, SRTs persist as an important source of stress in the world of Web.

Since speed has been considered one of the most important issues for Web users, interfaces that come with long and/or unpredictable SRTs may have both serious psychological and economical consequences (LIGHTNER et al. 1996). Even short SRTs of no more than 2 seconds may exert detrimental effects on the conversational nature of HCI (MILLER 1968). Delays of 8 seconds are commonly regarded as a threshold for adverse psychophysiological and performance consequences (BOUCSEIN 2007, SHNEIDERMAN & PLAISANT 2005). In addition, SRTs of 8 seconds and more may have a similar uncertainty-eliciting effect as variable SRTs (SCHAEFER 1990), thus inducing stress and feelings of anxiety, in particular with respect to the possibility of the occurrence of HCI breakdowns (BOUCSEIN et al. 1984).

Improving computer speed is not always a proper solution to prevent adverse effects of SRT; it may even have contrary effects (see Table 42.1). As a consequence of improving the page load speed from 8 to 2–5 seconds, long SRTs brought about reduced levels of trust and caused a lot of traffic as users sought alternatives (NIELSEN, 2000). This may not only lead to a decrease in user dissatisfaction but also to considerable losses in e-commerce sales, since Web users may simply give up trying to buy an item if they feel messed about by long and/or unpredictable SRTs.

Since making computer systems and networks faster will provide no general solution for the nagging problem of SRTs, software engineers should focus on how to mitigate their adverse effects. Based on our research as summarized in Sect. 2, the following measures could be considered:

- Fostering multi-tasking during which the user may be encouraged to behave like a time-sharing system (SCHAEFER et al. 2000), combined with
- providing indicators for elapsed time instead of imprecise estimated time until task completion (KUHMANN & SCHAEFER 2007).

Such a venture can not be successfully performed without empirical studies which include psychophysiological measures (BOUCSEIN et al. 1998). Otherwise, stress-inducing properties of interfaces for HCI may be left undetected, making the whole designing process unfeasible.

4 References

Barber RE, Lucas HC (1983) System response time, operator productivity, and job satisfaction. Communications of the ACM, 26: 972–986.

Boucsein W (1988) Wartezeiten am Rechner – Erholung oder Streß? Zeitschrift für Arbeitswissenschaft, 42: 222–225.

Boucsein W (1993) Psychophysiology in the work place – goals and methods. In: Schriftenreihe der Bundesanstalt für Arbeitsmedizin, Sonderschrift 2: Psychophysiology of mental workload. Haefner, Berlin, 35–41.

Boucsein W (2000) The use of psychophysiology for evaluating stress-strain processes in human-computer interaction. In: Backs RW, Boucsein W (Eds.) Engineering Psychophysiology. Issues and Applications. Lawrence Erlbaum Associates, Mahwah, NJ, 289–309.

Boucsein W (2006) Psychophysiologische Methoden in der Ingenieurpsychologie. In: Zimolong B, Konradt U (Eds.) Enzyklopädie der Psychologie. Themenbereich D Praxisgebiete. Serie III Wirtschafts-, Organisations- und Arbeitspsychologie. Band 2 Ingenieurpsychologie. Hogrefe, Göttingen: 317–358.

Boucsein W (2007) Systemresponsezeiten. In: Landau K (Ed.) Lexikon Arbeitsgestaltung. Best Practice im Arbeitsprozess. Gentner Verlag, Stuttgart: 1215–1217.

Boucsein W, Figge BR, Goebel M, Luczak H, Schaefer F (1998) Beanspruchungskompensation beim Multi-Tasking während der Bearbeitung einer CAD-Simulation. Zeitschrift für Arbeitswissenschaft, 52: 221–230.

Boucsein W, Greif S, Wittekamp J (1984) Systemresponsezeiten als Belastungsfaktor bei Bildschirm-Dialogtätigkeiten. Zeitschrift für Arbeitswissenschaft, 38: 113–122.

Carbonell JR, Elkind JI, Nickerson RS (1968) On the psychological importance of time in a time sharing system. Human Factors, 10: 135–142.

Card SK, Moran TP, Newell A (1983) The psychology of human-computer interaction. Lawrence Erlbaum Associates, Hillsdale, NJ.

Carriero NJ (1975) The effects of paced tapping on heart rate, skin conductance, and muscle potential. Psychophysiology, 12: 130–135.

Greif S (1983) Streß und Gesundheit: Ein Bericht über Forschungen zur Belastung am Arbeitsplatz. Zeitschrift für Sozialisationsforschung und Erziehungssoziologie, 3: 41–58.

Henning RA, Sauter SL, Salvendy G, Krieg EF (1989) Microbreak length, performance, and stress in a data entry task. Ergonomics, 32: 855–864.

Holling H (1993) Reduction of stress by the display of system operations in human computer interaction. In: Luczak H, Çakir A, Çakir G (Eds.) Work with display units. North-Holland, Amsterdam, 129–134.

Holling H, Gediga G (1987) Some theorems on stress in waiting times – predictions from a mathematical model. In: Roskam EE, Suck R (Eds.) Progress in mathematical psychology. North-Holland, Amsterdam: 233–249.

Hüttner J, Wandke H, Rätz A (1995). Benutzerfreundliche Sofware. Psychologisches Wissen für die ergonomische Schnittstellengestaltung. Berlin, Paschke.

Jacko JA, Sears A, Borella MS (2000) The effect of network delay and media on user perception of web resources. Behavior and Information Technology, 19: 427–439.

Kohlisch O, Kuhmann W (1997) System response time and readiness for task execution – the optimum duration of inter-task delays. Ergonomics, 40: 265–280.

Kohlisch O, Schaefer F (1996) Physiological changes during computer tasks: responses to mental load or to motor demands? Ergonomics, 39: 213–224.

Kuhmann W (1979) Realisation eines Verfahrens zur experimentellen Anforderungssteuerung bei einer einfachen Konzentrationsaufgabe. In: Vaitl D (Ed.) Psychophysiologie und Verhaltensmodifikation, Bericht über das IV. Kolloquium "Psychophysiologische Methodik", 13. – 16.12. 1979, Spitzigsee, Obb. Justus-Liebig-Universität, Gießen, 223–228.

Kuhmann W (1989) Experimental investigation of stress-inducing properties of system response times. Ergonomics, 32: 271–280.

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: 933–943.

Kuhmann W, Schaefer F (2007) Benutzerverhalten und Benutzerzustand in Abhängigkeit von der Vorhersehbarkeit systembedingter Wartezeiten. Zeitschrift für Arbeitswissenschaft, 61: 183–194.

Kuhmann W, Schaefer F, Boucsein W (1990) Effekte von Wartezeiten innerhalb einfacher Aufgaben: Eine Analogie zu Wartezeiten in der Mensch-Computer-Interaktion. Zeitschrift für Experimentelle und Angewandte Psychologie, 37: 242–265.

Lambert GN (1984) A comparative study of system response time on program developer productivity. IBM Systems Journal: 23, 36–43.

Lightner NJ, Bose I, Salvendy G (1996) What is wrong with the World-Wide-Web? A diagnosis of some problems and prescription of some remedies. Ergonomics, 39: 995–1004.

Luczak H (1987) Psychophysiologische Methoden zur Erfassung psychophysischer Beanspruchungszustände. In: Kleinbeck U, Rutenfranz J (Eds.) Enzyklopädie der Psychologie, Serie III, Band 1. Hogrefe, Göttingen, 185–259.

Martin J (1973) Design of man-computer dialogue. Prentice Hall, Englewood Cliffs, NY.

Melin B, Lundberg U (1997) A biopsychosocial approach to work-stress and musculoskeletal disorders. Journal of Psychophysiology, 11: 238–247.

Miller RB (1968) Response time in man-computer conversational transactions. Proceedings of the Fall Joint Computer Conference, Vol. 33. AFIRS Press, Reston, VA, 267–277.

Monat A, Averill JR, Lazarus RS (1972) Anticipatory stress and coping reactions under various conditions of uncertainty. Journal of Personality and Social Psychology, 24: 237–253.

Nickerson RS, Elkind JI, Carbonell JR (1968) Human factors and the design of time sharing computer systems. Human Factors, 10: 127–134.

Nielsen J (2000) Designing web usability: the practice of simplicity. New Riders, Indianapolis.

Peters T (1977) Arbeitsmedizinische Forderungen an die Bildschirmtextverarbeitung. Report, 5th European Congress for Text Processing, Intertext 1977.

Polkosky MD, Lewis JR (2002) Effect of auditory waiting cues on time estimation in speech recognition telephony applications. International Journal of Human-Computer Interaction, 14: 423–446.

Rauterberg M (1992) An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. Behaviour and Information Technology, 11: 227–236.

Sackman H (1970) Experimental analysis of man-computer problem-solving. Human Factors, 12: 187–201.

Schaefer F (1990) The effect of system response times on temporal predictability of work flow in human-computer interaction. Human Performance, 3: 173–186.

Schaefer F, Kohlisch O (1995) The effect of anticipatory mismatch in work flow on task performance and event related brain potentials. In: Grieco A, Molteni G, Occhipinti E, Piccoli B (Eds.) Work With Display Units 94 North-Holland, Amsterdam, 241–245.

Schaefer F, Kuhmann W, Boucsein W, Alexander J (1986) Beanspruchung durch Bildschirmtätigkeit bei experimentell variierten Systemresponsezeiten. Zeitschrift für Arbeitswissenschaft, 40: 31–38.

Schaefer F, Schäfer R, Boucsein W (2000) Auswirkungen von Prozeßlaufzeit und Prozessindikatoren beim Multi-Tasking auf Arbeitsstrategie und Beanspruchung des Benutzers. Zeitschrift für Arbeitswissenschaft, 54: 267–275.

Schleifer LM, Amick BC (1989) System response time and method of pay: stress effects in computer-based tasks. International Journal of Human Computer Interaction, 1: 23–39.

Schleifer LM, Okogbaa OG (1990) System response time and method of pay: cardiovascular stress effects in computer-based tasks. Ergonomics, 33: 1495–1509.

Shneiderman B (1984) Response time and display rate in human performance with computers. Computing Surveys, 16: 265–285.

Shneiderman B (1984) The psychology of human-computer interaction. Datamation, 30: 236–240.

Shneiderman B (1992) Response time and display rate. In Shneiderman B (Ed.) Designing the user interface. Addison-Wesley, Reading, Mass, 277–301.
Shneiderman B, Plaisant C (2005) Designing the user interface: strategies for effective human-computer interaction. Addison-Wesley, Boston.
Smith D (1983) Faster is better: a business case for subsecond response time. In Depth, Computerworld (April 18), 2–22.
Thum M, Boucsein W, Kuhmann W, Ray WJ (1995) Standardized task strain and system response times in human-computer interaction. Ergonomics, 38: 1342–1351.
Treurniet WC, Hearty PJ, Planas MA (1985) Viewers' responses to delays in simulated teletext reception. Behaviour and Information Technology, 4: 177–188.
Trimmel M, Meixner-Pendleton M, Haring S (2003) Stress response caused by system response time when searching for Information on the Internet. Human Factors, 45: 615–621.
Williges RC, Williges BH (1982) Modeling the human operator in computer-based data entry. Human Factors, 24: 285–299.