

# Perceived Time as a Measure of Mental Workload: Effects of Time Constraints and Task Success

Morten Hertzum and Kristin Due Holmegaard

Computer Science, Roskilde University, Roskilde, Denmark

---

The mental workload imposed by systems is important to their operation and usability. Consequently, researchers and practitioners need reliable, valid, and easy-to-administer methods for measuring mental workload. The ratio of perceived time to clock time appears to be such a method, yet mental workload has multiple dimensions of which the perceived time ratio has mainly been linked to the task-related dimension. This study investigates how the perceived time ratio is affected by time constraints, which make time an explicit concern in the execution of tasks, and task success, which is a performance-related rather than task-related dimension of mental workload. A higher perceived time ratio is found for timed than untimed tasks. According to subjective workload ratings and pupil-diameter measurements, the timed tasks impose higher mental workload. This finding contradicts the prospective paradigm, which asserts that perceived time decreases with increasing mental workload. A higher perceived time ratio was also found for solved than unsolved tasks, whereas subjective workload ratings indicate lower mental workload for the solved tasks. This finding shows that the relationship between the perceived time ratio and mental workload is reversed for task success compared to time constraints. Implications for the use of perceived time as a measure of mental workload are discussed.

---

## 1. INTRODUCTION

The safe and efficient operation of complex systems requires that the mental workload imposed on users does not exceed their capacity (Eggemeier, Wilson, Kramer, & Damos, 1991; Jex, 1988; Tsang & Vidulich, 2006). Even if capacity is not exceeded, a system design that imposes a high mental workload leaves less capacity for the task as such and is more taxing to use than a system design that imposes low mental

---

The first author provided the initial idea for the study, analyzed the data, and wrote most of the article. The second author planned, set up, and ran the experimental sessions. The authors took equal part in refining the initial idea for the study, and both authors critically read and revised draft versions of the article. We are grateful to Kasper Hornbæk, who made us aware of the possible use of perceived time as a measure relevant to system usability, and to Signe Arnklit, who recruited the participants for the experiment. Special thanks are due to the participants.

Address correspondence to Morten Hertzum, Computer Science, Roskilde University, Universitetsvej 1, Bldg 43.2, DK-4000 Roskilde, Denmark. E-mail: mhz@ruc.dk

workload. This may have important consequences, because users change their behavior when they experience high mental workload by, for example, responding more quickly to catch up, lowering their performance criteria, postponing minor tasks to preserve resources for major tasks, or experiencing distress (Eggemeier & Wilson, 1991). For these reasons reliable, valid, and easy-to-administer methods for measuring mental workload are important to the evaluation and iterative design of systems (Tsang & Vidulich, 2006; Xie & Salvendy, 2000). Perceived time has been proposed as one such method (e.g., Block, Hancock, & Zakay, 2010; Hart, 1975; Liu & Wickens, 1994; Zakay & Shub, 1998).

The ratio of perceived time to clock time has repeatedly been found to explain variations in mental workload (Block et al., 2010). Typically, perceived time ratios are lower (i.e., time flies) for high-workload than low-workload tasks, provided that users know in advance that they will be estimating time (Brown, 1997, 2008). The perceived time ratio has, however, mainly been linked to differences in task demands. Beyond workload differences imposed by task demands there is uncertainty about the relationship between the perceived time ratio and mental workload. This makes it difficult to use perceived time ratios as a measure of mental workload in real-work environments, in which multiple workload dimensions, such as task demands, time, and performance, tend to be simultaneously present. In this study, we aim to investigate the relationship between the perceived time ratio and two aspects of mental workload:

- *Time constraints*, which is a task-related aspect of mental workload. We consider it important to investigate time constraints in relation to perceived time because such constraints make time an explicit concern in the execution of the task and this may affect users' time perception and its relation to their mental workload. In addition, time constraints are frequent in air-traffic control, emergency response, process control, and other domains in which the pace is set by external events as well as in computer games, which use time constraints to create challenge. Finally, it is well known that time constraints affect mental workload (Czaja &

Sharit, 1993; Hart & Staveland, 1988) and the perceived usability of a system (Kokini, Lee, Koubek, & Moon, 2012).

- *Task success*, which is a performance-related aspect of mental workload. Performance is, for example, included as one of six subscales in the NASA task load index (TLX; Hart & Staveland, 1988). Task success is important to investigate in relation to measures of mental workload because users dynamically adjust their mental efforts to perform successfully and because task success—unlike, for example, time constraints—is associated with the end product of the task and therefore does not become apparent until at the end of the task. In addition, some studies suggest that time perception may be affected by task success (Czerwinski, Horwitz, & Cutrell, 2001; Schiffman & Greist-Bousquet, 1992). Finally, task success is important to all goal-directed behavior.

We investigate the effects of time constraints and task success on perceived time in an experiment where participants work on a mentally demanding code-breaking task similar to the game of mastermind. Participants attempt the code-breaking task in a timed version with a visually indicated time constraint on the completion of the tasks and in an untimed version with no time constraint on task completion. Task success is measured by whether participants succeed or fail at breaking the code. To manipulate mental workload further, the experiment also involves a periodic interruption of a kind that has previously been found to increase mental workload (Renaud & Blondin, 1997). Participants are notified of the interruptions auditorily, visually, or audiovisually because we are interested in knowing whether these frequent interruption types differentially affect participants' time perception. For reasons of comparison, independent measures of mental workload are obtained by means of TLX ratings (Hart & Staveland, 1988) and pupil-diameter measurements (Beatty, 1982).

In the following sections, we review related work on perceived time as a measure of mental workload, describe our experimental method, present our results, and discuss their implications. If the perceived time ratio can be developed into a robust and well-understood workload measure, it is much easier to administer than, for example, pupil-diameter measurements and may—because the relationship between response and measure is not readily apparent—avoid the response bias to which subjective workload measures such as TLX are susceptible. This could make the perceived time ratio an attractive workload measure in both research and applied settings. More generally, knowing how users perceive time may also inform the design of user interfaces by providing the understanding necessary to engineer perceived time.

## 2. RELATED WORK

Mental workload is a multidimensional concept with no single accepted definition. A working agreement exists, however,

about the meaning of the concept. For example, Kramer (1991) summarized several previous definitions when he wrote that “mental workload can be conceptualized as the interaction between the structure of systems and tasks on the one hand, and the capabilities, motivation, and state of the human operator on the other” (p. 279). Tsang and Vidulich (2006) proposed the even simpler consensus that “mental workload is very much a function of the *supply and demand* of attentional or processing resources” (p. 246).

### 2.1. Measures of Mental Workload

Measures of mental workload can be divided into three classes: performance-based measures, physiological measures, and subjective measures (Eggemeier et al., 1991; Tsang & Vidulich, 2006). Although singular measures of mental workload have been proposed (e.g., Jung & Jung, 2001), most researchers agree that different measures tap different dimensions of mental workload and that the simultaneous application of multiple measures, preferably from more than one of the three classes, is therefore advisable because it provides better diagnosticity (Eggemeier & Wilson, 1991; Tsang & Vidulich, 2006).

*Performance-based measures.* Users have a finite capacity for dealing with task demands. Hence, mental workload can be assessed by monitoring a user's performance and noting how it varies with changes in task demands. Such measures include, for example, strategy shifts (Eggemeier & Wilson, 1991) and deviations from optimal performance (Baldauf, Burgard, & Wittmann, 2009). A limitation of these measures is, however, that they tend not to reflect variation in the user's investment of resources to maintain performance when task difficulty increases (Tsang & Vidulich, 2006). Rather than direct measurement of task performance, mental workload may also be quantified as the user's residual capacity when performing a task. This residual capacity can be assessed by introducing a secondary task and monitoring the user's performance on this secondary task and how it varies with changes in the demands of the primary task (Fisk, Derrick, & Schneider, 1986). The choice of secondary task is important because it should not degrade primary-task performance yet it should consume sufficient residual capacity to be sensitive to changes in primary-task demands. Time estimation has been found to fulfill this requirement and may therefore be a suitable secondary task (Block et al., 2010; Brown, 1997).

*Physiological measures.* Mental workload has several physiological correlates, and multiple measures of brain, cardiac, skin, and eye activity have therefore been used to assess mental workload. These measures include event-related brain potentials, heart rate, skin conductance, and the diameter of the pupil (Just, Carpenter, & Miyake, 2003; Tsang & Vidulich, 2006; Wilson & Eggemeier, 1991). Physiological measures do not require the user to generate overt responses, they provide for continuous data recording, and they have high temporal sensitivity and can thus detect short periods of elevated workload (Wilson & Eggemeier, 1991). In terms of limitations,

these measures impose larger instrumentation requirements than other mental-workload measures, and they may be confounded by a number of variables not reflecting mental workload; for example, the pupil diameter is also affected by glare and lighting conditions (Kramer, 1991). Among the physiological measures of mental workload, changes in the pupil diameter are recognized as robust and thoroughly validated (Beatty, 1982; Just et al., 2003; Kramer, 1991). For example, Beatty (1982) found an increase in peak pupil dilation as demands increased in tasks concerning memory, language, reasoning as well as perception.

*Subjective measures.* Finally, mental workload can be measured by asking users about their experience of the demands imposed by a current or recently completed task. Proponents of subjective measures argue that mental workload is in the eye of the beholder and that subjective measures may therefore come closest to tapping the essence of mental workload (Hart & Staveland, 1988). In addition, subjective measures are generally easy to administer. The limitations of subjective measures include that users must temporarily suspend their work to rate their mental workload and that the overt rating process is susceptible to distortions such as response bias. Yeh and Wickens (1988) found evidence of dissociation between subjective and performance-based measures, but subjective measures “have generally demonstrated good concurrent validity with performance” (Tsang & Vidulich, 2006, p. 254). Well-established subjective measures of mental workload include the modified Cooper–Harper scale (Wierwille & Casali, 1983), the subjective workload assessment technique (Reid & Nygren, 1988), and TLX (Hart & Staveland, 1988). Of these, TLX consists of the six subscales Mental Demand, Physical Demand, Temporal Demand, Effort, Performance, and Frustration. TLX has been found to measure the average mental workload of a task, as opposed to, for example, accumulated or peak workload (Xie & Salvendy, 2000; Yeh, Wickens, & Hart, 1985).

## 2.2. Perceived Time

Perceived time and its relation to clock time have been researched for more than a century. For example, James (1890, Vol. I) wrote, “In general, a time filled with varied and interesting experiences seems short in passing, but long as we look back. On the other hand, a tract of time empty of experiences seems long in passing, but in retrospect short” (p. 624). This quote illustrates two points. First, the interest in perceived time as a measure of mental workload is but one possible application of time estimation. Others have, for example, related perceived time to engagement (Larsen & von Eye, 2006), flow (Csikszentmihalyi, 1990), and motivation (Conti, 2001). Second, the relationship of perceived to clock time is reversed for concurrent as opposed to retrospective time estimates. Studies of the use of perceived time as a measure of mental workload display a similar reversion (Block et al., 2010): When people know beforehand that they will be requested to

estimate the duration of a task after completing it, the ratio of perceived to clock time decreases as the mental workload of the task increases. Conversely, when people are not made aware that they will be requested to estimate task duration until after they have completed the task, the ratio of perceived to clock time increases with higher mental workload. The two situations involve different mental processes.

When people are told before they perform a task that they will afterward be asked to estimate its duration (known as the prospective paradigm), a secondary time-keeping task is introduced. The details of how people accomplish this secondary task are debated, but many models assume an internal clock or pulse that, when attended to, gives the experience of the passing of time (Brown, 1997, 2008; Grondin, 2010). Roughly speaking, perceived time is proportional to the number of pulses to which the user attends. In low-workload situations, the user’s residual mental capacity allows for attending to most of the pulses. However, when the workload imposed by the primary task increases then the residual capacity left for attending to the pulses decreases and pulses increasingly go unnoticed (Block et al., 2010; Hancock & Weaver, 2005). This model predicts that time perception will be increasingly distorted as workload increases and that the distortion will be in the direction of lower ratios of perceived to clock time at high workloads. The model is consistent with multiple empirical studies. Brown (2008) reported that 48 (67%) of 72 experiments that compare time estimates at different levels of mental workload find that increased workload leads to greater misestimation.

When people are unaware that they will be asked to estimate the duration of a task until after they have completed it (known as the retrospective paradigm), the estimate must be constructed from incidental memory. The details of how this is done are debated. The contextual-change model (Block, 2003; Block & Reed, 1978) proposes that time estimates are based on the availability of emotional, environmental, and other contextual changes encoded in memory during the time period. According to this model, time estimates are proportional to the number of available contextual changes. Contextual changes are, for example, encoded in memory when a task requires users to change the way they process information, such as during task switching (Block & Reed, 1978). Thus, this important source of increased mental workload yields increased retrospective time estimates. However, information-processing difficulty, another important source of mental workload, seems to have little influence on the number of encoded contextual changes (Block & Zakay, 1996) and, thus, cannot be measured with retrospective time estimates. This means that retrospective time estimates are sensitive to only some sources of mental workload (Block et al., 2010). In this study we, therefore, focus on prospective time estimation. A further reason for focusing on prospective time estimation is that it allows for asking users to give time estimates for a series of tasks, whereas retrospective timing is restricted to one task after which users know they are asked for time estimates.



A few studies relate perceived time to time constraints and task success. With respect to *time constraints*, Brown (2008) made the general observation that perceived time is strongly affected if attention is directed to time during the execution of tasks. If users are, for example, asked to estimate the duration of tasks during which they focus either on the passage of time or on the task itself, then a focus on the passage of time lengthens perceived time. Similarly, Fraisse (1984) concluded that “the more one pays attention to time, the longer it seems” (p. 31). Specifically, time constraints may increase mental workload as well as direct attention to time. In this case, time estimates will be expected to shorten if prospective models are extended to workload changes imposed by time constraints, and to lengthen because attention is directed to time.

Liu and Wickens (1994) found no effect of time constraints on perceived time. Although Liu and Wickens assumed that tighter time constraints increase mental workload, their participants’ TLX ratings of subjective mental workload did not differ for the loose and tight time-constraint conditions, though they approached a significant difference for the subscale on temporal demand ( $p < .1$ ). This weakens their results because it suggests that the two time-constraint conditions were insufficiently different to be experienced by participants as constituting two distinct levels of time constraint. Thus, it is difficult to interpret the absence of a difference in perceived time for the loose and tight time-constraint conditions in Liu and Wickens’s study.

The relationship between perceived time and *task success* cannot be fully explained by internal-clock models of prospective time estimation because these models do not account for workload changes that are induced at the end of a task, which is when task success or failure becomes known. Czerwinski et al. (2001) proposed an alternative model of the relationship between perceived time and task success derived from the Zeigarnik effect, which concerns the status in memory of a task that is interrupted before completion versus one that is completed without interruption. Zeigarnik (1927) showed that people who are interrupted can recall more task information than people who are allowed to complete the task. This effect is explained in terms of closure. An interrupted task does not offer closure and, as a consequence, involves some continued mental activity accompanied by the retention of task information in memory. Conversely, a completed task has reached closure and mental activity on the task has terminated, releasing task information from memory. The Zeigarnik effect extends to time perception in that it has been found that time estimates are longer for interrupted than completed tasks (Schiffman & Greist-Bousquet, 1992; Weybrew, 1984). This finding appears consistent with retrospective timing models, which suggest that more information retained in memory after interrupted tasks will lengthen time estimates. Czerwinski et al. extended this work further by proposing that task success resembles completed tasks and gives closure, whereas failure to solve a task resembles interrupted tasks and does not offer closure. Their empirical data, collected from a usability study, support this

proposition and show that task success shortens perceived time. However, in their usability study the experimenter intervened by offering hints when users were at risk of failing to complete a task. Czerwinski et al. noted that it is unknown whether the cause of the difference in perceived time is task success/failure or the absence/presence of experimenter interventions.

### 3. METHOD

To empirically investigate perceived time as a measure of mental workload we conducted a within-subjects experiment about how perceived time is affected by time constraints and task success.

#### 3.1. Participants

Sixteen participants (6 female, 10 male) took part in the experiment. Participants’ age ranged from 21 to 33 years with an average of 25.4 years. All participants were experienced computer users who used computers daily or near daily. In terms of background, 13 participants were students at a technical university, two were professionals, and one did not report his background. Nine participants indicated that they played computer games. Finally, all participants had normal or corrected-to-normal vision, a requirement introduced by the eye-tracking equipment used for the pupil-diameter measurements.

#### 3.2. Tasks

The task, similar to the game of mastermind, consisted of breaking a four-digit code by making repeated guesses and receiving feedback for each guess. The code was restricted to the digits 1 through 6 (e.g., 2265), and participants were provided up to eight guesses to break the code. These design choices were made on the basis of pilot tests aimed at finding a level of task difficulty where some codes were broken and others not and the task remained challenging throughout the session.

The screen area for solving the code-breaking task occupied the right-hand side of the full-screen application used for running the experiment (see Figure 1). When participants made a guess they received feedback in terms of (a) the number of correct digits in their correct position, (b) the number of correct digits not in their correct position, and (c) the number of incorrect digits. It is important to note that the feedback gave only the number of digits in each of the three categories and was devoid of information about which digits belonged in which category. Once a guess had been made it could not be changed, but the guess and the associated feedback remained visible on the screen. To solve the task, participants had to merge the feedback from their previous guesses into an understanding that gradually narrowed down the possible digit combinations for the code.

We chose this task because it is a cognitive task and sufficiently demanding to impose considerable mental workload, because its brevity allows for multiple iterations within a single session, and because we hoped its game qualities would

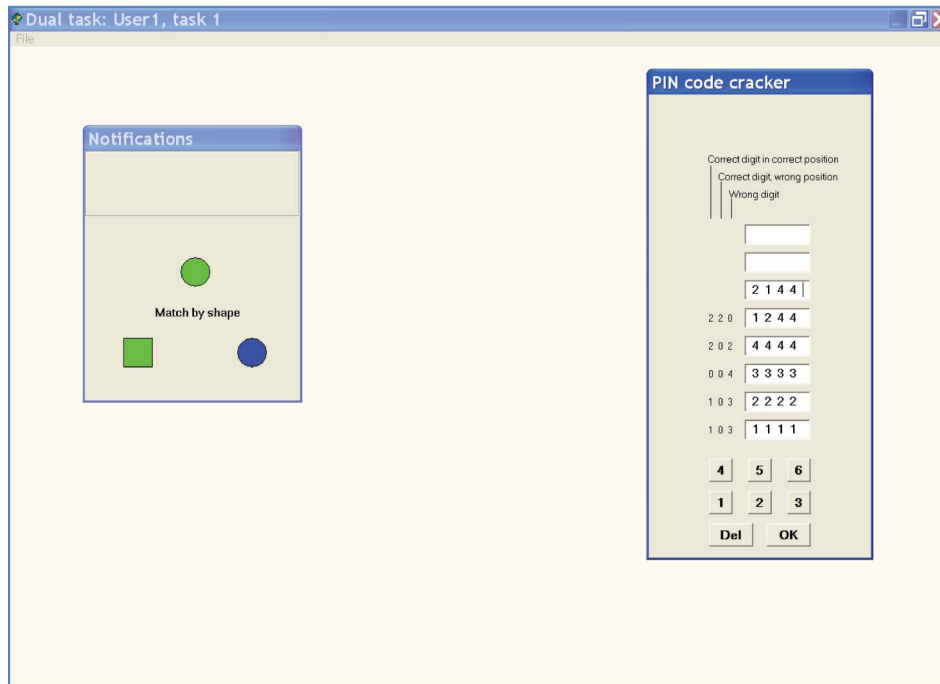


FIG. 1. Test application. *Note.* The left-hand part of the screen shows the expanded interruption bar with a green circle (the reference figure) above a green square and a blue circle (the target figures). The right-hand part of the screen shows the code-breaking task. The participant is at her sixth guess of an untimed task for which the code is 1442 (color figure available online).

strengthen participants' motivation and help avoid fatigue. In addition, the task allowed for introducing distinctions between two time-constraint levels and two task-success levels.

*Time constraint (timed, untimed).* During timed tasks, participants had a maximum of 25 s for each guess. This time limit was set on the basis of pilot tests. The passing of the 25 s was impressed upon participants by a progress bar that visualized how the elapsed time progressed toward the time limit; participants were not told that the time limit corresponded to 25 s. If a participant did not make a guess within the time limit, that guess was lost, the participant was moved forward to the next guess, and the progress bar restarted. During untimed tasks no time limits were enforced and participants could spend as much time as they needed on each guess. There was no progress bar during untimed tasks.

*Task success (solved, unsolved).* A task was solved if the participant broke the code, that is, if the participant's guess exactly matched the code. We chose against using the number of guesses to differentiate between multiple levels of solvedness because we felt that such distinctions would be minor compared to the main distinction between solving the task at all and not solving it. A task was unsolved if a participant had used all eight guesses and not broken the code.

Each of the 16 participants performed three blocks of eight tasks. Participants alternated between timed and untimed tasks. Half of the participants started with a timed task, and the other half with an untimed task. Task success was logged during the experimental sessions because it is a measure of participants'

actual performance and thus could not be controlled in the experimental setup.

### 3.3. Interruptions

We assumed the code-breaking task was sensitive to interruptions because it involved keeping track of how feedback from new guesses fitted with or forced revision of the understanding built from the feedback from earlier guesses. Thus, to manipulate mental workload further the code-breaking task was interrupted every 15 to 25 s.

The interruptions occupied an area in the left-hand part of the screen (see Figure 1). Between interruptions this area contained an empty bar. When an interruption occurred, participants were notified in different ways depending on the *interruption type* (auditory, visual, audiovisual, none). Auditory interruptions were indicated by a 1-second sound. Visual interruptions were indicated by the appearance of a white square in the interruption bar. Audiovisual interruptions were indicated by the 1-second sound and the white square. A final interruption type involved no interruptions; participants performed the code-breaking task without interruptions and the interruption bar was not present.

From the onset of a notification participants had 5 s to acknowledge the interruption by clicking the interruption bar, otherwise the interruption was cancelled. The acknowledgment caused the interruption bar to expand and reveal two target figures that differed in shape and color and a reference figure that matched one target figure in shape and the other in

color. A text below the reference figure instructed participants to “Match by shape” or “Match by color.” Participants completed the interruption by clicking the target figure consistent with the instruction. The interruption task is loosely based on the Stroop effect (MacLeod, 1991) and was adopted from McFarlane (2002), who noted that it cannot be automated and thus requires attention.

In each block of eight tasks, participants received interruptions of the same type for two consecutive tasks (i.e., a timed and an untimed) and then proceeded to the next of the four interruption types. The order of the interruption types was selected using four balanced Latin squares, one for each group of four participants. The target figures, reference figure, and instruction differed across interruptions.

### 3.4. Procedure

Participants were initially introduced to the experiment and asked questions about their background. Then participants were explained the task and the interruptions, followed by some training tasks, during which participants performed both timed and untimed tasks and experienced the different types of interruptions. To add weight to the workload manipulations, participants were instructed to attempt to complete all tasks as well as to react to all interruptions. Participants were also informed that they had 5 s to acknowledge interruptions. In the experiment participants acknowledged 96% of the interruptions and, thus, complied with the instruction. Next, participants were introduced to the TLX (Hart & Staveland, 1988) and to the estimation of perceived time. The preparations for the experimental tasks were completed by setting up and calibrating the eye tracker so that it accurately captured the participant’s pupil.

Participants performed three blocks of eight tasks, each block consisting of one instance of every combination of time constraint and interruption type. The time constraint (i.e., timed or untimed) and interruption type (i.e., auditory, visual, audiovisual, or none) were indicated on the screen ahead of each task. Upon completing a task participants rated their mental workload on the six TLX subscales and then estimated perceived time. After each block, participants were allowed a break before they commenced on the next block. After the third block, participants were debriefed.

To minimize noise in the pupil-diameter measurements, the experiment was run in a laboratory with controlled lighting conditions. External sunlight was blocked, and it was ensured that the internal light sources did not produce glare in the computer screen. The experiment lasted an average of 2.1 hr per participant. As a token of our appreciation participants received a gift certificate of DKK 350.

### 3.5. Dependent Variables

We measured mental workload in three ways. In addition to the perceived time ratio, mental workload was measured subjectively by means of TLX and physiologically by means of the participants’ pupil diameter.

*Perceived time ratio.* We express the relationship between perceived time and clock time by the commonly used ratio of perceived time to clock time (Block et al., 2010; Hancock & Weaver, 2005). With this definition, unity (100%) indicates that perceived time equals clock time. Participants estimated perceived time on a pop-up screen that appeared after each task. The estimate was made by positioning a slider on a scale from 0 to 10 min, with the possibility (used in only three instances) of shifting the scale to the interval from 10 to 20 min. The estimates were recorded in 6-second increments. Clock time was the interval between the log events recording the start and end of a task.

*TLX.* The administration of TLX comprised ratings of its six subscales: Mental Demand, Physical Demand, Temporal Demand, Effort, Performance, and Frustration. Each subscale was rated on a scale from low (0) to high (100) in increments of 5, except for Performance, where the anchors were *good* (0) and *bad* (100). The subscale ratings were made with sliders on the same pop-up screen as perceived time. We left out the weighting procedure for combining the six TLX subscales into a single measure of mental workload and, instead, report participants’ answers to the six subscales. This is done to increase the diagnostic information acquired from the TLX ratings and because the weighting procedure has been discouraged (Hendy, Hamilton, & Landry, 1993; Nygren, 1991).

*Pupil diameter.* Participants’ pupil diameter was recorded with a remote eye tracker from SMI, sampling at 50 Hz. The participants were seated so that the distance from the eye tracker, mounted below the stimulus screen, to the participants’ eye was approximately 70 cm, as recommended for the eye tracker. A calibration process, repeated for each block of tasks, ensured that the eye tracker accurately captured the participant’s pupil. As in previous studies (e.g., Bailey & Iqbal, 2008; Bernhardt, Dabbs, & Riad, 1996), the changes in pupil diameter were calculated by converting the pupil-diameter measurements to percentages of the participant mean. Hence, a value below 100% represents a constriction and a value above 100% represents a dilation of the pupil, relative to its average diameter across the 24 tasks.

## 4. RESULTS

In the following we first analyze the data for block effects. Then we analyze the effects of time constraint and task success on, in turn, perceived time ratios, TLX ratings, and pupil diameters. After these main parts of the section, the effects of interruptions on perceived time ratios are analyzed. We conclude the analysis by modeling the relationship of the perceived time ratios with TLX ratings and pupil measures.

### 4.1. Effects of Block

We first analyzed the 384 tasks for effects of learning and fatigue across the three blocks. Table 1 shows for each block the average perceived time ratio, perceived time, and clock time of a task and the average success rate for a participant. Across

TABLE 1  
Learning and Fatigue Across Blocks

	Block 1		Block 2		Block 3		Total	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Perceived time ratio	1.85	0.86	1.96	0.64	2.02	0.65	1.94	0.64
Perceived time (s)	208	83	192	74	195	78	198	75
Clock time (s)	136	58	115	40	116	47	122	44
Success rate (%)	45	25	51	27	53	24	49	22

Note.  $N = 384$  tasks.

all three blocks, participants overestimated time by an average of 94%. We note that 15 of the 16 participants overestimated time; only one participant had an average perceived time ratio of less than 1.

There was no effect of block for any of perceived time ratio, perceived time, clock time, and success rate,  $F_s(2, 14) = 0.66, 1.44, 3.08, 1.32$ , respectively (all  $ps > .08$ ). We therefore collapsed the data across blocks and included the data from all three blocks in the remainder of our analyses. In addition, task success very nearly divided the tasks into equally many solved and unsolved tasks. It should, however, be noted that (as indicated by the standard deviation) some participants successfully solved more of the 24 tasks than others. We accept this imbalance in our study as a near unavoidable consequence of studying how perceived time is affected by a performance aspect such as task success.

#### 4.2. Effects of Time Constraint and Task Success on Perceived Time Ratio

Table 2 shows perceived time ratio divided onto untimed and timed tasks and onto unsolved and solved tasks. To assign equal weight to each participant, we first calculated the average perceived time ratio, and the averages of the other time measures, for each participant and made the analysis on these averages. For perceived time ratio, we found significant effects of time constraint and task success,  $F_s(1, 15) = 30.86, 55.35$ , respectively (both  $ps < .001$ ). These

results show that the perceived time ratio was higher for timed than untimed tasks and for solved than unsolved tasks. We also found a significant interaction between time constraint and task success,  $F(1, 15) = 6.27, p < .05$ , indicating that the increase in perceived time ratio from untimed to timed tasks was lower for unsolved tasks (a 27% increase) than for solved tasks (a 37% increase). The perceived time ratio was lowest for untimed, unsolved tasks ( $M = 1.38, SD = 0.38$ ) and highest for timed, solved tasks ( $M = 2.72, SD = 1.01$ ).

For perceived time, there were significant effects of time constraint and task success,  $F_s(1, 15) = 33.10, 7.73$ , respectively (both  $ps < .05$ ), but no interaction between time constraint and task success,  $F(1, 15) = 0.28, p = .6$ . These results show that participants gave shorter time estimates for timed than untimed tasks and for solved than unsolved tasks. For clock time, there were significant effects of time constraint and task success,  $F_s(1, 15) = 39.91, 38.22$ , respectively (both  $ps < .001$ ), but no interaction between time constraint and task success,  $F(1, 15) = 2.28, p = .2$ . These results show that timed tasks were shorter than untimed tasks and that solved tasks were shorter than unsolved tasks. We note that if perceived time had changed proportionally to clock time, there could have been significant differences in perceived time and clock time without significant differences in the perceived time ratio. The significant difference in the perceived time ratio indicates that the participants' estimates of perceived time changed more than proportionally to the changes in clock time.

TABLE 2  
Effects of Time Constraint and Task Success on Perceived Time Ratios

	Time Constraint				Task Success			
	Untimed		Timed		Unsolved		Solved	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Perceived time ratio	1.78	0.55	2.11	0.76	1.64	0.48	2.23	0.76
Perceived time (s)	231	106	165	54	201	51	182	95
Clock time (s)	157	74	88	19	133	34	98	58

Note.  $N = 384$  tasks.

**4.3. Effects of Time Constraint and Task Success on TLX Ratings**

To investigate whether time constraint and task success affected subjective mental workload in ways similar to perceived time ratio, we analyzed participants' TLX ratings. Table 3 shows the average rating for each TLX subscale and for an overall TLX rating computed by averaging the six subscales. Generally, Mental Demand and Effort were high and Physical Demand was low, whereas there appeared to be more variation in Temporal Demand, Performance, and Frustration.

Before the statistical analysis we transformed the TLX ratings to avoid that differences in time constraint or task success went unnoticed because some participants consistently gave higher TLX ratings than others. The transformation consisted of subtracting the participant mean on each TLX subscale from all the participant's ratings on that subscale. To clarify the data on which the analysis was made, Table 4 shows the transformed ratings. A multivariate analysis of the six subscales showed significant effects of time constraint, Wilks's  $\lambda = 0.12$ ,  $F(6,$

$10) = 12.14$ ,  $p < .001$ , and task success, Wilks's  $\lambda = 0.03$ ,  $F(6, 10) = 51.57$ ,  $p < .001$ , with higher TLX ratings for timed than untimed tasks and for unsolved than solved task. These effects were moderated by a significant interaction between time constraint and task success, Wilks's  $\lambda = 0.25$ ,  $F(6, 10) = 5.15$ ,  $p < .05$ . With the experiment-wide error thus protected we proceeded with analyses of the individual TLX subscales.

There were significant effects of time constraint on Temporal Demand,  $F(1, 15) = 58.38$ ,  $p < .001$ , and Physical Demand,  $F(1, 15) = 5.54$ ,  $p < .05$ , but not on any of Mental Demand, Effort, Performance, and Frustration,  $F_s(1, 15) = 0.004, 0.02, 0.61, 1.58$ , respectively (all  $p_s > .2$ ). Temporal Demand was higher for timed tasks, and Physical Demand was, surprisingly, higher for untimed tasks. The average of the six TLX subscales was higher for timed than untimed tasks,  $F(1, 15) = 18.53$ ,  $p < .001$ .

There were significant effects of task success on Mental Demand, Temporal Demand, Effort, Performance, and Frustration,  $F_s(1, 15) = 17.17, 34.47, 17.39, 118.71, 90.14$ ,

TABLE 3  
Effects of Time Constraint and Task Success on Raw TLX Ratings

	Time Constraint				Task Success			
	Untimed		Timed		Unsolved		Solved	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Mental Demand	60	11	64	11	69	12	53	14
Physical Demand	9	10	8	9	9	9	9	9
Temporal Demand	19	16	64	14	57	17	30	11
Effort	54	16	58	13	62	13	48	15
Performance	36	15	45	14	62	12	17	14
Frustration	43	17	53	15	62	15	33	12
Average TLX	37	8	49	8	54	8	32	7

Note.  $N = 384$  tasks. TLX = task load index.

TABLE 4  
Effects of Time Constraint and Task Success on Transformed TLX Ratings

	Unsolved				Solved			
	Untimed		Timed		Untimed		Timed	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Mental Demand	7	6	7	8	-9	10	-10	19
Physical Demand	0	3	0	2	1	5	-2	3
Temporal Demand	-18	13	29	14	-25	11	10	15
Effort	5	7	7	9	-8	10	-9	16
Performance	19	12	22	13	-21	14	-27	16
Frustration	10	12	16	9	-15	10	-15	15

Note.  $N = 384$  tasks. TLX = task load index.



respectively (all  $ps < .001$ ), but not on Physical Demand,  $F(1, 15) = 1.13, p = .3$ . All differences indicated higher ratings for unsolved than solved tasks (note that for performance higher ratings indicate worse performance).

We also found significant interactions between time constraint and task success for Temporal Demand and Performance,  $F_s(1, 15) = 6.56, 4.73$ , respectively (both  $ps < .05$ ), but not for any of the four other TLX subscales (all  $ps > .2$ ). The increase in temporal demand from untimed to timed tasks was larger for unsolved than solved tasks (see Table 4). The improvement in performance ratings from timed to untimed tasks was larger for solved than unsolved tasks.

In sum, the timed tasks resulted in higher perceived time ratios and higher subjective workload, whereas the solved tasks resulted in higher perceived time ratios but lower subjective workload.

#### 4.4. Effects of Time Constraint and Task Success on Pupil Diameter

To further cross-validate the use of perceived time ratio as a measure of mental workload, we analyzed how time constraint and task success affected participants' pupil diameter. Before analyzing the pupil measurements we removed 87,807 (3.8%) measurements during which the pupil could not be located, for example, due to blinks. We also removed 6,639 (0.3%) outlier measurements that had a pupil diameter more than 50% above or below the participant's mean pupil diameter. The remaining 2,190,908 pupil measurements were converted to percentages of the participant mean. Table 5 shows the minimum, maximum, and average pupil diameter for a task and the average pupil diameter for the last 5% of the clock duration of a task. A multivariate analysis of the four pupil measures showed significant effects of time constraint, Wilks's  $\lambda = 0.18, F(4, 12) = 13.62, p < .001$ , and task success, Wilks's  $\lambda = 0.33, F(4, 12) = 6.06, p < .01$ , but no interaction between time constraint and task success, Wilks's  $\lambda = 0.56, F(4, 12) = 2.33, p = .1$ . We therefore analyzed the individual pupil measures with respect to main effects only.

There were significant effects of time constraint on the minimum, maximum, and average pupil diameter during a task and on the average pupil diameter during the last 5% of a task,  $F_s(1, 15) = 24.09, 14.71, 5.56, 5.16$ , respectively (all  $ps < .05$ ). The average pupil diameter was larger during timed than untimed tasks, indicating higher mental workload during timed tasks. The minimum pupil diameter and the average pupil diameter during the last 5% of a task were also larger during timed than untimed tasks, corroborating the result for the average pupil diameter. Conversely, maximum pupil diameter was larger during untimed than timed task. Together the results for the minimum and maximum pupil diameter suggest less variation in mental workload during timed tasks.

We found significant effects of task success on minimum pupil diameter, maximum pupil diameter, and average pupil diameter during the last 5% of a task,  $F_s(1, 15) = 15.48, 9.95, 7.86$ , respectively (all  $ps < .05$ ), but not on average pupil diameter,  $F(1, 15) = 3.04, p = .1$ . Minimum pupil diameter was larger and maximum pupil diameter smaller during solved than unsolved tasks, suggesting less variation in mental workload during solved tasks. In addition, the average pupil diameter was larger during the last 5% of solved than unsolved tasks, indicating that solved tasks ended at a higher level of mental workload.

In sum, the larger pupil diameters for timed tasks co-occurred with higher perceived time ratios. For task success, the average pupil diameter across the entire task was unaffected but the pupil diameter was larger at the end of solved than unsolved tasks. It should be noted that though significant, the differences in pupil diameter were generally small.

#### 4.5. Effects of Interruptions on Perceived Time Ratio

We also investigated whether the perceived time ratio was affected by periodic interruptions of the tasks. Table 6 shows the results for tasks that were subject to either no, auditory, visual, or audiovisual interruptions. We found no effects of interruption on any of perceived time ratio, perceived time, and clock time,  $F_s(3, 13) = 0.42, 2.95, 0.82$ , respectively (all  $ps > .08$ ).

TABLE 5  
Effects of Time Constraint and Task Success on Pupil Diameter

	Time Constraint				Task Success			
	Untimed		Timed		Unsolved		Solved	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Minimum	62.52	6.37	69.48	6.28	64.50	6.02	68.06	7.22
Maximum	130.10	5.91	126.57	4.85	129.29	6.43	126.89	4.56
Average	99.51	1.15	101.40	2.04	100.49	1.58	100.94	2.46
Average of last 5%	101.16	2.85	103.64	2.87	101.71	3.02	103.49	3.03

Note. Pupil diameter given in percent of the participant mean.  $N = 384$  tasks.

TABLE 6  
Effects of Interruptions on Perceived Time Ratios

	None		Audio		Visual		Audiovisual	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Perceived time ratio	1.95	0.81	1.90	0.52	1.90	0.58	2.02	0.86
Perceived time (s)	180	67	211	93	208	89	194	67
Clock time (s)	113	38	128	57	126	63	121	40

Note.  $N = 384$  tasks.

Thus, the ratio of perceived to clock time was unaffected by the different interruptions (and so was perceived time itself), suggesting that the absence or presence of the interruptions did not change participants' mental workload. Consistent with the results for perceived time ratio, a multivariate analysis showed no effect of interruptions on the six TLX subscales, Wilks's  $\lambda = 0.53$ ,  $F(18, 113.62) = 1.58$ ,  $p = .08$ . Similarly, another multivariate analysis showed no effect of interruptions on the four pupil diameter measures, Wilks's  $\lambda = 0.33$ ,  $F(12, 4) = 0.69$ ,  $p = .7$ .

#### 4.6. Relationship of Perceived Time Ratio With TLX and Pupil Measures

To investigate the perceived time ratio without a priori assigning importance to time constraint, task success, and interruptions, we used multiple regression to analyze the extent to which the variation in perceived time ratio could be explained by the variation in TLX ratings and pupil measures. For these analyses each of the 384 tasks provided associated values of perceived time ratio and, for example, the four pupil-diameter measures. We determined the regression models by backward elimination (Thompson, 1978); that is, we initially included all candidate predictor variables and then sequentially eliminated the variable that contributed the least to explaining the variation

in perceived time ratio. This elimination process continued as long as the  $F$  test of the eliminated variable was not significant. Table 7 summarizes the three regression models.

For TLX ratings, we regressed the perceived time ratio against the six TLX subscales and average TLX. The resulting regression model was significant,  $F(2, 381) = 23.34$ ,  $p < .001$ , and included Temporal Demand and average TLX. The variation in temporal demand and average TLX explained 11% of the variation in perceived time ratio. For Temporal Demand the  $\beta$  coefficient was positive, indicating that when Temporal Demand increased so did perceived time ratio. For average TLX the  $\beta$  coefficient was negative, indicating that when average TLX increased, perceived time ratio decreased. The elimination of the other subscales from the model indicated that they did not contribute appreciably to explaining the variation in perceived time ratio.

The regression of the perceived time ratio against the four pupil-diameter measures resulted in a significant model,  $F(3, 380) = 19.98$ ,  $p < .001$ , that included the minimum, maximum, and average pupil diameter. The variation in these three pupil measures explained 14% of the variation in perceived time ratio. The  $\beta$  coefficients show that when the minimum and average pupil diameter increased so did the perceived time ratio, whereas when the maximum pupil diameter increased then the perceived time ratio decreased.

TABLE 7  
Regression Models

Predicted Variable	Candidate Predictors	Resulting Predictors	$\beta$	$R^2$
Perceived time ratio	TLX measures	Temporal demand	0.42	11%
		Average TLX	-0.48	
Perceived time ratio	Pupil measures	Minimum	0.22	14%
		Maximum	-0.24	
		Average	0.16	
Perceived time ratio	TLX and pupil measures	Performance	-0.22	18%
		Minimum pupil diameter	0.21	
		Maximum pupil diameter	-0.21	
		Average pupil diameter	0.14	

Note.  $N = 384$  tasks. TLX = task load index.

We also regressed the perceived time ratio against TLX ratings as well as pupil measures. The resulting regression model was significant,  $F(4, 379) = 21.37, p < .001$ , and included the TLX subscale Performance and the minimum, maximum, and average pupil diameter. The variation in these four variables explained 18% of the variation in perceived time ratio. When the minimum and average pupil diameter increased, so did the perceived time ratio, whereas when performance and the maximum pupil diameter increased, the perceived time ratio decreased.

## 5. DISCUSSION

The average perceived time ratio across all tasks is 1.94, indicating a considerable overestimation of 94%. Considerable overestimation is, however, not exceptional in studies of perceived time. For example, Loftus, Schooler, Boone, and Kline (1987, Experiment 2) find an average overestimation of 407% across their 266 participants. In the following, we discuss the relationship of perceived time to mental workload, its implications for evaluation and design, and the limitations of our study.

### 5.1. The Relationship of Perceived Time to Mental Workload

For time constraint, we find that timed tasks result in higher perceived time ratios, higher TLX ratings, and larger pupil diameters than untimed tasks. The difference in overall TLX rating is accompanied by a difference in the subscale on temporal demand. As expected, timed tasks impose higher temporal demand than untimed tasks. In short, the subjective and physiological measures are in agreement and show that the manipulation of time constraint imposed the planned change in mental workload. Our results for perceived time ratios are contrary to models of prospective timing, which predict lower perceived time ratios for increased task demands. In these models, task demands have been manipulated in ways unrelated to time, for example, by varying the number of digits in mental arithmetic tasks (Brown, 1997), the number of sorting rules in card-sorting tasks (Zakay & Shub, 1998), the presence or absence of a proof-reading task (Brown & Stubbs, 1992), and the number of items in a pattern to be held in memory during visuospatial tasks (Benuzzi, Basso, & Nichelli, 2005). Thus, prospective timing models cannot be extended to timed tasks. This has previously been noted by Brown (2008), who stated that time perception is strongly affected if attention is directed to time. In our study, timed and untimed tasks were sufficiently different to impose a difference in mental workload, and we find that when time is made an explicit concern in the execution of a task by means of a time constraint then the perceived time ratio increases. This result suggests that the reason Liu and Wickens (1994) did not find an effect of time constraint on perceived time was that their time-constraint conditions were insufficiently different. To use perceived time as a measure of mental workload it is, thus,

necessary to distinguish between time-unrelated tasks and time-constrained tasks because an increase in mental workload for these two types of task has opposite effects on perceived time ratios.

For task success, we find that solved, as opposed to unsolved, tasks result in higher perceived time ratio, lower TLX ratings, and unchanged average pupil diameter. The changes in the TLX subscales of Performance and Frustration toward better performance and less frustration for solved tasks are intuitively right. In addition, the absence of a difference in average pupil diameter is consistent with task success being an end-of-task aspect of performance. The lower overall TLX rating and the lower ratings on the subscales of Mental Demand, Temporal Demand, and Effort show that the end result of a task in terms of success or failure affects the subjective experience of how demanding the task was. This is consistent with previous findings of considerable covariation between the TLX subscales (Hart & Staveland, 1988). We also find larger pupil diameter during the last 5% of the duration of solved tasks, indicating higher mental workload during the final part of solved tasks. This result is opposite to the lower TLX ratings but may indicate that success in the code-breaking task is contingent on managing an amount of information that builds up to become quite demanding, whereas task failure may be associated with not investing this amount of mental effort during the final part of tasks. Thus, an interpretation of the perceived time ratios must take into account that task success reduces subjective mental workload and has no effect on the average physiologically measured mental workload except near the end of tasks.

We offer three observations regarding the relationship of the perceived time ratio to task success. First, the perceived time ratio is sensitive to task success. This finding is consistent with the proposal by Czerwinski et al. (2001) to extend the Zeigarnik effect from task interruption to task success. We find, however, that task success increases the perceived time ratio, whereas the Zeigarnik effect predicts the opposite relationship. A possible explanation could be that the extension of the Zeigarnik effect from task interruption to task success is untenable, for example, because interruption is an external event for which the user is not accountable, whereas task success concerns the user's ability and motivation to perform. Second, it appears that the effect of task success on perceived time ratios cannot be explained by assuming that users keep track of time by attending to the pulses from an internal clock. Rather, the number of attended pulses will be largely similar for solved and unsolved tasks because the pupil-diameter measurements show that task success has no physiological effect, except near the end of tasks. Third, motivation may suggest an explanation of the larger perceived time ratios for solved tasks. Participants may retrospectively associate solved tasks with concerted efforts during which they, for some reason, tried harder, and they may associate unsolved tasks with having invested less in solving the task. Further work is required to investigate whether a perception of trying harder leads to longer time estimates. If so, motivation-related

explanations of the relationship between perceived time and task success may be an alternative to explanations in terms of mental workload.

For time constraint and task success combined, 18% of the variation in the perceived time ratio is explained by the variation in mental workload as measured by TLX ratings and pupil diameters. Only four of the 11 TLX and pupil measures contribute significantly to explaining the variation in the perceived time ratio. The four measures are the TLX subscale Performance, which is sensitive to task success but not time constraint, and the minimum, maximum, and average pupil diameter, which are sensitive to time constraints but, in the case of the average pupil diameter, not task success. Notably, the pupil-diameter measures reflect time constraints so well that including the TLX subscale Temporal Demand would not increase the amount of variation explained. In sum, the perceived time ratio captures a significant amount of the variation in mental workload.

## 5.2. Implications for Evaluation and Design

The usability (Hertzum, 2010) of a system is affected by the mental workload it imposes, and the measurement of mental workload is therefore important to evaluation and design. This study shows that perceived time is a reliable and easy-to-administer alternative or supplement to other measures of mental workload in the various situations where time constraints or task success may be important sources of mental workload. Time constraints are prominent when tasks are driven by external events, success whenever tasks are goal directed. In addition, previous work has shown that perceived time also reliably measures mental workload that is imposed by time-unrelated task demands. Unless the sources of mental workload can be reliably divided into time-unrelated task demands, time constraint, and performance-related task success it is, however, a nontrivial task to interpret the perceived time ratio.

Compared to subjective workload measures, perceived time is a less overt way of asking about mental workload. It is probably not obvious to users how their time estimates are translated into a measure of mental workload, reducing the possibilities to game the measurements and thereby the risk of response bias. However, perceived time lacks the diagnosticity of TLX and other subjective workload measures with multiple subscales. Whereas TLX provides information about whether the experienced mental workload is mainly driven by, for example, temporal demand or performance, the interpretation of perceived time as a measure of workload requires such information. This suggests that perceived time should supplement rather than replace subjective workload measures.

Compared to physiological workload measures, perceived time is a more lightweight workload measure in that it is easier to administer and less costly in equipment. These advantages resemble those of subjective workload measures and, for example, make perceived time applicable in situations

where physiological measures are considered too intrusive. Physiological measures can be made continually and are, thus, diagnostic with respect to when workload increases and decreases during a task, even for brief changes in workload. Although perceived time mostly lacks this diagnosticity, it is sensitive to end-of-task changes induced by whether tasks succeed or fail. This suggests that perceived time and physiological workload measures are to a considerable extent complementary.

The principal result of the research on perceived time is that users' experience of time comprises more than clock time. In this study we have investigated how differences beyond those in clock time affect perceived time. It is, however, equally the case that differences in clock time are experienced in terms of characteristics other than time. For example, Ramsay, Barbesi, and Preece (1998) found that a web page is experienced as more interesting if it downloads faster. In terms of implications for interface design, this means that time is an important concept to understand and that users' experience of time can be engineered. As indicated by the perceived time ratios in this study, perceived time may differ considerably from clock time, suggesting potentially large effects of engineering time well. Efforts to engineer time must, however, pay close attention to the means used to engineer users' perception of time because task demands, time constraints, and successes affect perceived time in different and interrelated ways. For the design of games, in particular, the association of perceived time to mental workload provides an alternative to the dominant linking of perceived time to flow and immersion.

## 5.3. Limitations

Four limitations should be remembered in interpreting the results of this study. First, we did not have full control over task success because it is a property of participants' performance. Although the overall number of solved and unsolved tasks were almost the same, individual participants differed somewhat in the number of tasks they solved successfully. As previously noted, we acknowledge this limitation but see it as a near unavoidable consequence of studying how a performance measure, such as task success, affects perceived time. Solved tasks also had shorter clock times than unsolved tasks, but we have accounted for this by analyzing the ratio of perceived to clock time. Second, participants indicated perceived time after rating the six TLX subscales. This ordering ensures that the TLX ratings are unaffected by the mental processes involved in estimating time but entails that time was estimated after participants had been sensitized to mental workload. Third, participants experienced the code-breaking task 24 times. Although our analysis of participants' performance across blocks does not suggest that fatigue is a problem, participants may be subject to an anchoring effect (Tversky & Kahneman, 1974) and fail to adjust their time estimates sufficiently from one instance of the task to the next. The alternation of timed and untimed tasks aimed to minimize such anchoring. Fourth, the 0-to-10-minute



scale used for rating perceived time may have biased participants toward overestimation because the midpoint of the scale (300 s) was well above the average clock time for a task (122 s). We chose the 10-min endpoint on the basis of pilot tests and to accommodate all tasks but acknowledge that participants' general tendency to overestimate time may in part be due to the high value of the scale endpoint.

## 6. CONCLUSION

The measurement of mental workload is important to system design and evaluation because high workload induces fatigue, leads to behavior changes associated with increased risk of error, and leaves less capacity for other tasks. As a measure of mental workload, perceived time is known to be sensitive to changes in time-unrelated task demands. This study has investigated how perceived time is affected by time constraints, which is a time-related task demand, and task success, which is a performance-related aspect. We find that

- Time constraints significantly affect the perceived time ratio with higher ratios for timed than untimed tasks. According to TLX ratings and pupil-diameter measurements, timed tasks impose higher mental workload. Notably, an increase in mental workload due to either time constraints or time-unrelated task demands changes the perceived time ratio in opposite directions.
- Task success significantly affects the perceived time ratio with higher ratios for solved than unsolved tasks. Solved tasks impose lower mental workload according to TLX ratings but only affect average pupil diameters near the end of tasks, indicating that task success is an end-of-task aspect of performance. Thus, an increase in the perceived time ratio indicates higher mental workload, if caused by time constraints, but lower mental workload, if caused by task success.

Whereas a change in the perceived time ratio appears a reliable and easy-to-administer indicator of a change in mental workload, it is more difficult to interpret how the direction of a change in perceived time ratios relates to mental workload. Specifically, the relationship is reversed for time constraints, which direct the user's attention to time, compared to task success and time-unrelated task demands. This reduces the diagnosticity of perceived time as a measure of mental workload and suggests that it is a supplement rather than an alternative to other workload measures, unless it is known in advance that the source of any change in mental workload is either time constraints or unrelated to time. In the absence of such knowledge, this study finds that the variation in TLX ratings and pupil-diameter measurements explains 18% of the variation in perceived time ratios. It is also worth noting that the perceived time ratio was consistent with TLX ratings and pupil-diameter measurements in indicating no differential effect of interruptions depending on whether users were notified of the

interruptions by an auditory, visual, or audiovisual signal or worked uninterrupted. Thus, the perceived time ratio appears to agree with subjective and physiological workload measures in indicating both the presence and the absence of a difference in mental workload.

More research is needed to understand, in particular, the mechanisms through which task success affects perceived time and the use of perceived time as a measure of mental workload in the frequent real-world situations in which a number of sources—time-related as well as time-unrelated—contribute to mental workload.

## REFERENCES

- Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction, 14*, 21:01–21:28.
- Baldauf, D., Burgard, E., & Wittmann, M. (2009). Time perception as a workload measure in simulated car driving. *Applied Ergonomics, 40*, 929–935.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276–292.
- Benuzzi, F., Basso, G., & Nichelli, P. (2005). Temporal production and visuo-spatial processing. *Perceptual and Motor Skills, 101*, 737–758.
- Bernhardt, P. C., Dabbs, J. M., & Riad, J. K. (1996). Pupillometry system for use in social psychology. *Behavior Research Methods, Instrumentation, & Computers, 28*, 61–66.
- Block, R. A. (2003). Psychological timing without a timer: The roles of attention and memory. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 41–59). Göttingen, Germany: Hogrefe & Huber.
- Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*, 330–343.
- Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 656–665.
- Block, R. A., & Zakay, D. (1996). Models of psychological time revisited. In H. Helfrich (Ed.), *Time and mind* (pp. 171–195). Kirkland, WA: Hogrefe & Huber.
- Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics, 59*, 1118–1140.
- Brown, S. W. (2008). Time and attention: Review of the literature. In S. Gronin (Ed.), *Psychology of time* (pp. 111–138). Bingley, UK: Emerald.
- Brown, S. W., & Stubbs, D. A. (1992). Attention and interference in prospective and retrospective timing. *Perception, 21*, 545–557.
- Conti, R. (2001). Time flies: Investigating the connection between intrinsic motivation and the experience of time. *Journal of Personality, 69*, 1–26.
- Csikszentmihalyi, M. (1990). *Flow—The psychology of optimal experience*. New York: Harper.
- Czaja, S. J., & Sharit, J. (1993). Stress reactions to computer-interactive tasks as a function of task structure and individual differences. *International Journal of Human-Computer Interaction, 5*, 1–22.
- Czerwinski, M., Horwitz, E., & Cutrell, E. (2001). Subjective duration assessment: An implicit probe for software usability. In *Proceedings of the IHM-HCI 2001 Conference* (Vol. 2, pp. 167–170). Toulouse, France: Cépaduès.
- Eggemeier, F. T., & Wilson, G. F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 217–278). London, UK: Taylor & Francis.
- Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 207–216). London, UK: Taylor & Francis.

- Fisk, A. D., Derrick, W. L., & Schneider, W. (1986). A methodological assessment and evaluation of dual-task paradigms. *Current Psychological Research & Reviews*, 5, 315–327.
- Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, 35, 1–36.
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, 72, 561–582.
- Hancock, P. A., & Weaver, J. L. (2005). On time distortion under stress. *Theoretical Issues in Ergonomics Science*, 6, 193–211.
- Hart, S. G. (1975). Time estimation as a secondary task to measure workload. In *Proceedings of the 11th Annual Conference on Manual Control* (pp. 64–77). Washington, DC: Government Printing Office.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA–TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, the Netherlands: North-Holland.
- Hendy, K., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors*, 35, 579–601.
- Hertzum, M. (2010). Images of usability. *International Journal of Human–Computer Interaction*, 26, 567–600.
- James, W. (1890). *The principles of psychology*. New York, NY: Dover.
- Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 5–39). Amsterdam, the Netherlands: North-Holland.
- Jung, H. S., & Jung, H.-S. (2001). Establishment of overall workload assessment technique for various tasks and workplaces. *International Journal of Industrial Ergonomics*, 28, 341–353.
- Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, 4, 56–88.
- Kokini, C. M., Lee, S., Koubek, R. J., & Moon, S. K. (in press). Considering context: The role of mental workload and operator control in users' perceptions of usability. *International Journal of Human–Computer Interaction*, 28, 543–559.
- Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279–328). London, UK: Taylor & Francis.
- Larsen, E., & von Eye, A. (2006). Predicting the perceived flow of time from qualities of activity and depth of engagement. *Ecological Psychology*, 18, 113–130.
- Liu, Y., & Wickens, C. D. (1994). Mental workload and cognitive task automaticity: An evaluation of subjective and time estimation metrics. *Ergonomics*, 37, 1843–1854.
- Loftus, E. F., Schooler, J. W., Boone, S. M., & Kline, D. (1987). Time went by so slowly: Overestimation of event duration by males and females. *Applied Cognitive Psychology*, 1, 3–13.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human–computer interaction. *Human–Computer Interaction*, 17, 63–139.
- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33, 17–33.
- Ramsay, J., Barbesi, A., & Preece, J. (1998). A psychological investigation of long retrieval times on the world wide web. *Interacting with Computers*, 10, 77–86.
- Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam, the Netherlands: North-Holland.
- Renaud, P., & Blondin, J.-P. (1997). The stress of Stroop performance: Physiological and emotional responses to color-word interference, task pacing, and pacing speed. *International Journal of Psychophysiology*, 27, 87–97.
- Schiffman, N., & Greist-Bousquet, S. (1992). The effect of task interruption and closure on perceived duration. *Bulletin of the Psychonomic Society*, 30, 9–11.
- Thompson, M. L. (1978). Selection of variables in multiple regression: Part I. A review and evaluation. *International Statistical Review*, 46, 1–19.
- Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 243–268). New York, NY: Wiley.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Weybrew, B. B. (1984). The Zeigarnik phenomenon revisited: Implications for enhancement of morale. *Perceptual and Motor Skills*, 58, 223–226.
- Wierwille, W. W., & Casali, J. G. (1983). A validated rating scale for global mental workload measurement application. In *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 129–133). Santa Monica, CA: HFS.
- Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 329–360). London, UK: Taylor & Francis.
- Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *International Journal of Cognitive Ergonomics*, 4, 213–242.
- Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, 30, 111–120.
- Yeh, Y.-Y., Wickens, C. D., & Hart, S. G. (1985). The effect of varying task difficulty on subjective workload. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 765–769). Santa Monica, CA: HFS.
- Zakay, D., & Shub, J. (1998). Concurrent duration production as a workload measure. *Ergonomics*, 41, 1115–1128.
- Zeigarnik, B. (1927). Über das behalten von erledigten und unerledigten handlungen [On finished and unfinished tasks]. *Psychologische Forschung*, 9, 1–85.

## ABOUT THE AUTHORS

**Morten Hertzum** is associate professor in computer science at Roskilde University. His research interests are in usability evaluation, assessments of how information systems affect human activities and organizational settings, and pilot implementation as a means of learning in information systems development. Currently, much of his empirical work concerns healthcare technologies.

**Kristin Due Holmegaard** is a Ph.D. student in computer science at Roskilde University. Her research interest is methods for the evaluation of usability and mental workload and how these methods interact with contextual variables such as task complexity and time constraints.