Measuring subjectively experienced time in usability and user experience testing scenarios

Anna K. Trukenbrod\textsuperscript{ab,*}, Nils Backhaus\textsuperscript{b}, Roland Thomaschke\textsuperscript{c}

\textsuperscript{a} Technische Universität Berlin, Marchstr. 23 Sekr. 3-2, Berlin 10587, Germany
\textsuperscript{b} Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Friedrich-Henkel-Weg 1-25, Dortmund 44149, Germany
\textsuperscript{c} Albert-Ludwigs-Universität Freiburg, Engelbergerstr. 41c, Freiburg 79085, Germany

A R T I C L E   I N F O

Keywords:
Subjectively experienced time
Time perception
User experience
Usability testing
Think aloud

A B S T R A C T

Duration perception influences decision and evaluation processes in daily life. Nevertheless, subjectively experienced time is not commonly considered in usability and user experience (UX) testing scenarios. In this paper, we introduce an easy way to measure subjectively experienced time (SXT) and integrate it into a framework of UX, both theoretically and empirically. In two studies \((N = 62\) and \(N = 80)\), we show that SXT can be evaluated by users after solving tasks on a website. Moreover, we show that it correlates substantially with components of UX but that it is not redundant to common UX measures. Additionally, we show that varying aspects of the website (usability) and the evaluation task (think aloud) show similar effects on both SXT and UX measures. Our results suggest that SXT is partly based on the estimated clock time as distortions in estimations translate into subjectively experienced time. We conclude that users perceive time during an ongoing interaction with a technical device, that they evaluate it, and integrate this evaluation into their UX evaluation. Hence, the presented way of measuring SXT can be handy for UX practitioners to understand how objective temporal characteristics of the interaction translate into UX, especially when designing temporal aspects of interactions to elicit a positive UX.

1. Introduction

In user experience (UX) research, subjectively experienced temporal aspects of an interaction are often considered to be important (e.g. Abbasi et al., 2012) but seldom put at the center of attention. In contrast, objective temporal aspects such as task completion time or time on task are frequently measured in UX testings to infer the efficiency of the system (Albert et al., 2009; Bevan, 2009; Frøkjær et al., 2000; Lesemann et al., 2007), its intuitivity (Blackler et al., 2010), or its expectation conformity (DIN EN ISO 9241-11, 2018). The subjective experience of these temporal aspects, however, is often reduced to perceived waiting times (Antonides et al., 2002; Dellaert and Kahn, 1998; Kurusathianpong and Tangmamee, 2018; Lee et al., 2012), even though delays are not part of every interaction with a technical device. In addition, “temporal aspects” is a very unspecific term. Some authors refer to the temporal aspects of UX when they talk about the dynamics of UX over a repetitive use (Karapanos et al., 2010; Minge and Thüring, 2018; Pohlmeier et al., 2009). Other authors emphasize the importance of temporal aspects of single interactions such as system responsiveness (Liikkanen and Gómez Gómez, 2013), brief delays (Szameitat et al., 2009) or waiting-intervals (Kurusathianpong and Tangmamee, 2018). To the best of our knowledge, how users perceive and evaluate temporal attributes of an interaction without delays has not been studied yet.

We argue that users perceive time during most of their daily interactions with computers independently whether the interaction comprises a delay or not. Therefore, research on UX and UX measures can benefit from understanding how users perceive and evaluate time during any continuous, ongoing interaction. On the one hand, being able to understand the user’s subjective experience of time can help to design interactions for a better temporal experience as well as a better overall UX (Liikkanen and Gómez Gómez, 2013; Seow, 2008). On the other hand, time is a mundane concept and can be evaluated and verbalized very quickly and intuitively. Hence, measuring the subjective experience of time might offer a way to approach UX with very simple everyday terminology making it easy for users to understand. However,
it is not clear how to measure the temporal experience nor which characteristics of an interaction will influence the subjective experience of time.

In this paper, we define *subjectively experienced time (SXT)* as the user's subjective evaluation of temporal attributes of an interaction. We ask how SXT can be measured and how it relates to temporal aspects of the experience as well as to the user experience (UX) during a typical interaction with a computer. To this end, we first build upon an earlier definition of the term SXT as coined by Liikkanen and Gómez Gómez (2013) and discuss a possible operationalization (chapter 1.1). Second, we discuss SXT and its relations to UX and prospective time perception (chapter 1.2). In our following empirical studies, we offer a way to measure SXT in a simple fashion. Moreover, we show, firstly, that characteristics of the interaction affect SXT consistent with UX and time perception theories and, secondly, how these different domains interrelate to SXT.

We present two studies to provide a reliable basis for measuring SXT and for connecting SXT with adjacent areas of research. Overall, we aim to show that SXT offers a valuable addition to the UX methodological toolkit that can be easily applied in UX testing scenarios.

1.1. How to define and measure SXT?

We define subjectively experienced time (SXT) as the user's evaluation of temporal attributes of an interaction. Such temporal attributes comprise system responsiveness, overall duration (clock time), delays, and waiting intervals. The user perceives these attributes of the interaction and evaluates them creating a SXT (Liikkanen and Gómez Gómez, 2013). Hence, SXT is a context- and user-dependent variable influenced by temporal attributes of the interaction but also by past experience, affective state, and attribution of the experienced time (Liikkanen and Gómez Gómez, 2013). As such, SXT is similar to UX as it represents a subjective construct (Hassenzahl, 2003; 2008; Law et al., 2009). Liikkanen and Gómez Gómez (2013) argued that SXT mediates global UX and, therefore, should be considered when designing for a positive UX. However, the authors did not test this claim empirically nor present a way to measure it.

Regarding the question of how to measure SXT, one can look at similar constructs. One of these is perceived waiting time. Perceived waiting time can be understood as an estimate of the waited clock time (e.g. Antonides et al., 2002) or as a subjective construct reflecting an evaluation of the waiting time (e.g. Kurusathianpong and Tangmanee, 2018; Lee et al., 2012). The later is similar to SXT as defined above, even though it is specific for waiting. It has mostly been measured via seven- to ten-point rating scales on different items (see Table 1).

Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Item Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of the waited clock time</td>
<td></td>
</tr>
<tr>
<td>long/short</td>
<td>Antonides et al. (2002)</td>
</tr>
<tr>
<td>unacceptable/acceptable</td>
<td>Antonides et al. (2002); Kurusathianpong and Tangmanee (2018)</td>
</tr>
<tr>
<td>poor/good</td>
<td>Delfert and Kahn (1998)</td>
</tr>
<tr>
<td>Passage of time during the wait</td>
<td></td>
</tr>
<tr>
<td>fast/slow</td>
<td>Gorn et al. (2004); Tangmanee (2018); Lee et al. (2012)</td>
</tr>
<tr>
<td>speedy/not speedy quick/not quick</td>
<td>Gorn et al. (2004); Lee et al. (2012)</td>
</tr>
<tr>
<td>Affective state of the subject</td>
<td></td>
</tr>
<tr>
<td>annoying/pleasant</td>
<td>Antonides et al. (2002)</td>
</tr>
<tr>
<td>boring/varied</td>
<td>Antonides et al. (2002); Tangmanee (2018)</td>
</tr>
<tr>
<td>unsatisfactory/satisfactory</td>
<td>Antonides et al. (2002); Tangmanee (2018)</td>
</tr>
<tr>
<td>irritating/not</td>
<td>Antonides et al. (2002); Kurusathianpong and Tangmanee (2018)</td>
</tr>
<tr>
<td>irritating</td>
<td>Tangmanee (2018)</td>
</tr>
</tbody>
</table>

1.2. How does SXT relate to UX and time perception research?

In the introduction, we asked how SXT relates to the adjacent areas of research, UX and time perception. An answer can be found based on two strategies. First, one can focus on correlations between SXT and dimensions of UX resp. time perception. Second, one can check whether SXT is influenced by the same factors as these dimensions. To put it more precisely, if SXT is indeed a valuable addition to established UX measures, it should share some variance (hence, correlate) with UX measures but not be fully congruent. Moreover, it should be influenced by the same factors as UX measures. Following this line of argumentation, we will discuss how SXT could fit into an existing UX framework in the following chapter. The same line of argumentation can be applied on SXT and time perception. If SXT is indeed based upon the user's perception of temporal attributes, it should both correlate with the user's time perception and be influenced by the same factors. Here, we aim to link SXT to basic research and basic theories about cognitive processes when perceiving time.

1.2.1. Integrating SXT in a framework of UX

One important reductionist framework of UX is the CUE model (Components of User Experience, Thüringer, 2018; Thüring and Mahlke, 2007). The CUE model suggests that UX evolves based on the user's perception of the objective characteristics of the interaction. These characteristics emerge based on the system, the user, and the context of the interaction including the task. The perception of the objective characteristics includes three major components: First, the user's subjective perception of instrumental qualities such as usability and usefulness, second, the user's perception of non-instrumental qualities such as aesthetics and other hedonic aspects, and third, the user's emotions during the interaction. The three components are seen as the core components of UX. They affect each other and build the basis for creating a global UX including an overall evaluation and a possible intention to (re)use the system. Several studies support this framework by, for example, showing an effect of the system's objective usability and its visual appearance on all three components of UX (Ben-Bassat et al., 2006; Hamborg et al., 2014; Thüring and Mahlke, 2018; Sonderegger and Sauер, 2010; Tractinsky et al., 2000).

Temporal aspects are not named specifically in this framework. However, they can be seen as part of the characteristics of the interaction. The duration of an interaction (clock time), for example, is an objective characteristic. As argued by the CUE model, it should be influenced by the user who is conducting the interaction, by the system and by the task. Such influences have already empirically been shown: Low usability of a system increases the time on task (e.g. Tractinsky et al., 2000; Ziefle and Bay, 2005). Also, a more complex task has been shown to prolong the duration of the interaction (e.g. adding think aloud to an interaction, see Hertzum et al., 2009).

Based on the CUE model, we expect the user to perceive this objective duration and to evaluate this perception (SXT). The evaluation is then integrated into the overall experience. Hence, SXT can be seen as
part of UX in the framework of the CUE model. We can test this assumption based on correlations between SXT and UX variables. In UX research, the different components of UX normally show medium to strong correlations between each other (e.g. Aranyi and van Schaik, 2015; Lavie and Tractinsky, 2004). Therefore, SXT should also be correlated to UX components with similar strength. Additionally, SXT should be influenced by similar factors as UX variables. A very strong and heavily researched influencing factor on UX is the usability of a system. It has been shown to affect the perception of instrumental qualities (e.g. Hamborg et al., 2014; 2018; 2017; Sucala et al., 2011). However, the correlations reported of or correlates with a passage of time judgment (Droit-Volet et al., 2010; 2018; 2013; Gil and Droit-Volet, 2011; 2012). Regarding the passage of time items, several authors have shown that the estimated clock time is a significant predictor of or correlates with a passage of time judgment (Droit-Volet et al., 2010; 2018; 2013; Gil and Droit-Volet, 2011; 2012). SXT, on the other hand, can be measured with the help of subjective rating scales as discussed earlier. Even though such items have not been researched extensively in the context of human-computer interactions, they can be found in other contexts. Antonides et al. (2002) used items regarding the assessment of the interaction’s clock time to measure the SXT of a waiting interval. They report only for one of two experiments that the estimated clock time was a significant predictor for SXT. However, the authors mixed those items with items concerning the affective state of the subject. Therefore, it is not clear whether this mingling of constructs affected the relationship between estimated clock time and SXT. Regarding the passage of time items, several authors have shown that the estimated clock time is a significant predictor of or correlates with a passage of time judgment (Droit-Volet et al., 2018; 2017; Sucala et al., 2011). However, the correlations reported here are rather small to medium (r = .33, Sucala et al., 2011).

Summoning up, we can expect SXT to covary with perceived time: a longer interaction duration (clock time) should not just lead to longer estimated clock time but also to a more negative SXT. This would support the connection between the interaction characteristic clock time, its perception and SXT as argued based on the CUE-model in chapter 1.2.1. However, we expect that SXT will not only be based on the estimated clock time as evidence only shows a relation of medium strength. Here, appraisal processes as measured with UX scales might be another important part of SXT.

Another way to check how SXT relates to the estimated clock time is to observe whether both variables are affected by the same factors. One heavily researched factor that influences prospective estimates of clock time is the attentional demand (Block et al., 2010). When less compared to all attention is available to manage the timing task, verbal estimates of clock time tend to be shorter. The effect has been shown for events redirecting attention away from the timing task (Block et al., 1980) and for secondary tasks (Brown, 1997). Moreover, the strength of temporal distortion due to secondary tasks can be amplified by the secondary task’s inherent mental workload (Block et al., 2010; Brown, 1985). The effect of additional attentional demands on prospective time estimates can be assigned to cognitive processes based on an “internal clock” that helps to monitor and process time as it passes (Gibbon, 1977; Gibbon et al., 1984; Treisman, 1963; Zakay and Block, 1997).

Though support for this effect comes mainly from basic research, it has also been shown in applied settings. Hertzum and Holmegaard (2015), for example, confronted users with additional attentional demands by asking them to think aloud while solving a task at a computer. Think aloud is a frequently used method in UX testing (Hertzum et al., 2009), in which users are asked to verbalize their thoughts during an ongoing interaction (Albert et al., 2009). These thoughts typically concern what users are doing, how they try to solve tasks, what expectations they have, and why they take certain actions during the interaction. The method offers a reliable way to detect usability errors and causes for errors (McDonald et al., 2013). However, Hertzum et al. (2009) showed that both classical think aloud (level 1 to 2 verbalization without verbalizing reasons or feelings, for definition see Hertzum and Holmegaard, 2013) as well as relaxed think aloud (level 3 verbalization with verbalizing reasons and feelings) increases mental demand and, hence, confront the user with additional attentional demands. In their study, Hertzum and Holmegaard (2015) showed that adding classical think aloud as a secondary task decreases verbally estimated clock time as expected.

Taken together, we can expect that a longer clock time will be perceived by the user and lead to longer estimated clock time. This estimation should correlate with SXT. However, adding attentional demands via think aloud should lead to a distortion in estimates. Hertzum and Holmegaard (2015) showed this effect for classic think aloud but the same should hold true for relaxed think aloud which is used far more often in UX testing scenarios (Hertzum et al., 2009). It is, however, not clear how such additional attentional demands in the form of relaxed think aloud affect SXT and whether distortions in estimated clock time translate into distortions in SXT. We aim to answer this question with the help of the following two studies.

1.3. Objectives

In this paper, we investigate subjectively experienced time (SXT) during an interaction with a personal computer. In contrast to earlier research, we do not focus on waiting but on ongoing interactions solving a specific task. We define SXT as the user’s evaluation of the temporal attributes of an interaction. We argue that users perceive the interaction’s duration (clock time) leading to an estimate of clock time. This estimate is evaluated resulting in SXT. We further argue that SXT can be seen as an important part of UX. Therefore, it might be a valuable addition to the UX toolkit. In the following two studies, we ask how SXT can be measured and how it relates to adjacent constructs (subjective perception of objective temporal aspects as well as components of UX). To this end, we choose to study typical interactions between a user and a website (e.g. searching for or entering information) and measure SXT, objective and estimated clock time as well as UX dimensions. In two studies, we experimentally vary characteristics of the task and the system by using think aloud as a secondary task for half of the participants and by manipulating the objective usability. This setting allows us to analyze SXT in the following ways:

- How does SXT covary with measures of prospective time perception and with different components of UX? Here, we check whether changes in the adjacent constructs go along with changes in SXT. This would support the line of argumentation presented in chapter 1.2.1 regarding the integration of SXT in a framework of UX.
- Is SXT influenced by similar factors as the adjacent constructs? On
the one hand, think aloud is a way to induce additional attentional demands of the task which should lead to a distortion in prospective time estimates. On the other hand, the objective usability of the system should have an effect on UX-components. Hence, the experimental design allows evaluating if SXT is influenced by the same factors as the adjacent constructs. Again, a similarity in influencing patterns would support the notion that SXT is connected to both time perception and UX as argued above.

2. Study 1

2.1. Method

2.1.1. Subjects

A study with $N = 65$ participants was conducted. Three participants were excluded from the analysis due to misinterpretation of the instructions or technical malfunction. Of the remaining $N = 62$ participants, 39 were male and 23 were female. Most participants (91.9%) were students at the Technische Universität Berlin and could earn course credit or money (8 Euros). All participants gave their informed consent in accordance with the declaration of Helsinki before participating. The final sample was on average $M = 25.16$ years old ($SD = 4.30$) and ranged from 18 to 38 years.

2.1.2. Setting and design

Participants were informed that they would participate in a UX test of a website. The website was prepared to confront participants with either low or high usability. The two versions of the website did not differ in content but the low usability website was more difficult to use due to bad readability, shifting links and disruptive pop-ups. Participants were instructed to complete three tasks with the website. Besides, they were asked to pay attention to the passing time during the task to ensure the prospective timing paradigm. Half of the participants were additionally instructed to perform think aloud while interacting with the website. This instruction included the request to describe as much and as precise as possible what they were doing, thinking, seeing, and feeling, as well as to verbalize their goals and decisions. This type of think aloud can be compared to relaxed think aloud (level 3 verbalization) which has been shown to increase cognitive workload (Hertzum et al., 2009). Thus, the study consisted of a two-factorial design with the independent between factors usability (high vs. low) and think aloud (with vs. without think aloud). The dependent variables comprised SXT, temporal dimensions, and UX. The measured constructs are shown in Table 2. They were measured after each task.

2.1.3. Operationalization of the dependent variables

The items for the variables SXT, subjective usability, attractiveness, and overall UX were based on 11-point semantic differential slider bars with the bipolar items presented in Table 2. SXT was based on operationalizations of perceived waiting times in terms of wording (Antonides et al., 2002) and their bipolar nature (Antonides et al., 2002; Gorn et al., 2004; Lee et al., 2012). For the UX-related items, single items were preferred over multiple item questionnaires because of the repetitive use during the experiment. The item for overall UX was adapted from a validated scale from the German questionnaire meCUE (Minge and Riedel, 2013; Minge et al., 2016). We used its 11-point scale for the other items as well to create a homogenous scaling.

The clock time of the interaction was measured automatically by the experimental setting. The estimated clock time was assessed verbally in minutes and seconds by the participants after each task. It was used to compute perceived time ratios (PTR) (e.g. Block et al., 2010) by dividing estimated clock time by clock time for each subject and task. PTRs allow comparing estimations independently from individual and task-related variations in durations. Thus, they reflect relative over- and underestimation while avoiding absolute values. They were analyzed in addition to estimated clock time because they can show whether a distortion in estimated clock time was present, even when clock time is not constant as it is in this study. However, as they are not as easily understandable as estimated clock time, analysis on both estimated clock time and PTRs will be presented.

To measure emotions, we used the Affect Grid (Russell et al., 1989) which measures both subjective valence and arousal on a single item in a 9-by-9 grid ranging from extremely negative to extremely positive (valence, x-axis) and from extremely sleepy to extremely aroused (arousal, y-axis). In the following analysis, however, we will focus on the subjective valence only. All scales were presented in German.

2.1.4. Procedure and material

We assigned participants randomly to one of four groups (low/high usability × with/without think aloud). After a short introduction, the experiment started with a free exploration of the website. We used an existing information website of a student consulting firm in Berlin (http://www.cct-ev.de, website from 2013) to create a low and a high usability version. In a pilot study ($N = 20$) subjective ratings measured with the German version of the meCUE questionnaire (Minge and Riedel, 2013; Minge et al., 2016) showed that the two versions differed significantly regarding perceived usability, $t(18) = 6.64, p < 0.001, d = 3.11$, perceived usefulness, $t(18) = 3.61, p < 0.01, d = 1.7$, and overall evaluation, $t(18) = 8.25, p < 0.001, d = 3.9$

After an exploration of maximal four minutes, participants had to solve three tasks in a counterbalanced order. In Task A, participants had to complete a contact form simulating an online application. Tasks B and C were typical usability tasks in which participants had to extract certain information from the website (B: Finding a specific project and its final product, C: Extracting the number of persons on the alumni board). To navigate to the contact form in Task A and to find the page describing the final product in Task B, two clicks were necessary. In Task C, three clicks were required to see the alumni board. If participants failed to complete the task after a certain time (6 min for Task A, 4 min for Task B, and 4 min for Task C) they were kindly instructed to stop their attempt to solve the task and proceed with the questionnaire. For each task, objective duration was measured and subjects were instructed to give prospective task duration estimates as well as to rate their SXT and UX including their emotional state. During the whole experiment, the investigator was present. In the think aloud group, he/she monitored whether the participants were performing think aloud and reminded them to continue if they stopped. The experiment concluded with a validated UX questionnaire (meCUE, Minge et al., 2016), the demographic questionnaire (regarding age, sex, education, technophilia, and course of study), and a debriefing about the study’s true

<table>
<thead>
<tr>
<th>Variable</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>SXT</td>
<td>The duration of the task was...</td>
</tr>
<tr>
<td></td>
<td>... short (1) or long (11)</td>
</tr>
<tr>
<td></td>
<td>... appropriate (1) or inappropriate (11)</td>
</tr>
<tr>
<td></td>
<td>... acceptable (1) or unacceptable (11)</td>
</tr>
<tr>
<td>(Objective) clock time [sec]</td>
<td>ratio between estimated and objective clock time</td>
</tr>
<tr>
<td>Estimated clock time [sec]</td>
<td>Subjective usability (perception of instrumental qualities)</td>
</tr>
<tr>
<td>PTR (perceived time ratio)</td>
<td>The website was usable (1) or not usable (11)</td>
</tr>
<tr>
<td>Attractiveness (perception of non-instrumental qualities)</td>
<td>The website was unattractive (1) or attractive (11)</td>
</tr>
<tr>
<td>Valence (emotion)</td>
<td>extremely negative (1) or extremely positive (9)</td>
</tr>
<tr>
<td>Overall UX (global evaluation)</td>
<td>The website was bad (1) or good (11)</td>
</tr>
</tbody>
</table>

Table 2
Overview of the dependent variables in Study 1.
purpose. During the whole experiment, no temporal cues were given to the participants. Overall, the experiment lasted about 30 min.2

2.1.5. Data preparation

Prior to averaging the three items measuring SXT for each task and participant, we conducted a principal component analyses for each point of measurement (first to third task) and for each task (Task A to C). All six analyses showed one underlying factor for the items indicating that the reduction of the three items into one measurement was justified.

To assure the validity of the single-UX-items, we correlated the three single items measured after the last task with the corresponding scales from the meCUE. The single item subjective usability correlated highly with the meCUE scale subjective usability, $r = 0.87$, $p < 0.001$. Attractiveness was correlated with visual aesthetics, $r = 0.64$, $p < 0.001$, and the single item overall evaluation with the meCUE scale overall evaluation, $r = 0.78$, $p < 0.001$. As all correlations were substantial, we concluded that the single items were valid measurements.

The time limits of the tasks were reached in 18 of 62 cases for Task A, in 24 of 62 cases for Task B, and in 15 of 62 cases for Task C (57 of all 3*62 = 186 tasks). For the analysis, we included all 186 tasks. However, an additional analysis with only those tasks within the limit was conducted and will be presented at the end of the results section.

2.1.6. Hypothesis

SXT was the central dependent variable in Study 1. Here, we expected to find positive correlations with clock time as well as with estimated clock time based on the earlier argued connection between estimated clock time and SXT (e.g. Antonides et al., 2002). In addition, we also expected participants to evaluate the duration more positively when they experienced a more usable and attractive website and when they experienced a more positive valence.

Concerning the effects of the independent variables on SXT, we looked at all effects exploratory to compare the results to the effects of the independent variables on the other variables. Regarding these other variables, we expected an effect of think aloud on duration estimates (PTRs) due to additional attentional demands based on prior research (Hertzum and Holmegaard, 2015), but no effect of usability on PTRs. Regarding the components of UX and the overall UX evaluation, we expected to find an effect of the usability of the system, as such effects have been frequently shown before (e.g. Ben-Bassat et al., 2006; van der Geest and van Dongelen, 2009; Hamborg et al., 2014; Minge and Thüring, 2018; Thüring and Mahlke, 2007). In addition to these main effects, we analyzed all interaction effects exploratively.

2.1.7. Data analysis

To analyze the correlations, we used within-subject correlations based on Bland and Altman (1994, 1995). This type of correlation is a measure for how much two variables covary within-subjects, that is, how much one variable will adjust when the other variable changes within a person. Thus, it is suitable for personal ratings as used here because such ratings are subject- and situation-related values and do not have one true value per person. Hence, averaging per subject to allow a classical correlation analysis would have been inappropriate. To compute the within-subject correlation coefficient $r$, we used the rmcorr package (Bakdash and Marusich, 2018) in R (R Core Team, 2018) version 3.5.0.

Table 3

<table>
<thead>
<tr>
<th>Within-subject correlations</th>
<th>SXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock time</td>
<td>$r = 0.27$, $p &lt; 0.01$</td>
</tr>
<tr>
<td>Estimated clock time</td>
<td>$r = 0.41$, $p &lt; 0.001$</td>
</tr>
<tr>
<td>PTR</td>
<td>$r = 0.21$, $p &lt; 0.05$</td>
</tr>
<tr>
<td>Subj. usability</td>
<td>$r = 0.43$, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>$r = 0.34$, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Valence</td>
<td>$r = 0.56$, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Overall UX</td>
<td>$r = 0.40$, $p &lt; 0.001$</td>
</tr>
</tbody>
</table>

To analyze the effects of the independent variables, we used linear mixed models (LMMs) with maximum likelihood estimation from the lme4 package (Bates et al., 2015). For visualization, the ggpubr package was used (Wickham, 2009). A linear mixed effect model was fitted for each dependent variable (DV) based on the following formula:

$$DV \sim \text{Think Aloud} + \text{Usability} + \text{Usability: Think Aloud} + \text{Task A vs. B} + \text{Task B vs. C} + \text{(1|ID)}$$

As fixed effects, the formula contains the main effects of usability and think aloud and their interaction as well as a sliding difference contrast for the tasks (Task A vs. B resp. Task B vs. C). Usability and think aloud were also coded as sliding difference contrasts (based on the MASS package, Venables and Ripley, 2002) to allow an easy interpretation of the intercept as the grand mean and the estimates as the difference between the respective groups. A significant effect of the intercept indicates that the grand mean is significantly different from zero, while a significant estimate indicates that the two groups compared with each other differ significantly from each other. For hypothesis testing, only the effects of usability and think aloud are of relevance. As random effects, we added a random intercept over participants (ID) to control for dependencies between repeated measures within-subjects. Other random effects were not included because each participant delivered only three points of measurement which was too little data to estimate random intercepts over task or random slopes for tasks over ID (Bolker, 2019).

Even though the lme4 package does not provide $p$-values, the provided t-values can be interpreted in terms of significance. In accordance with Kliegl et al. (2010), model terms with absolute t-values above 2 ($|t| > 2$) were interpreted as a significant predictor. Before the data analysis, we tested whether residuals of the LMM would be normally distributed based on a Box-Cox power transformation (based on the car and the MASS packages, Fox and Weisberg, 2011; Venables and Ripley, 2002). Based on this test, we used logarithmic transformations as a standard transformation for dependent variables when necessary. In the following analysis, all transformations will be declared.

2.2. Results

2.2.1. Correlations

The correlations between SXT and the other dependent variables are presented in Table 3. They show how much SXT covaries with the other variables within one subject. As expected, we found positive correlations for both clock time and estimated clock time. With increasing clock time and estimated clock time, participants reported a more negative SXT. We also found a small correlation with PTR-values indicating stronger overestimations for more negative SXT. Regarding UX components, the more negative SXT was assessed by a participant, the less positive this participant rated the subjective usability, attractiveness, valence, and overall evaluation. Note that each subject only experienced one of the four groups. Thus, the within-subject correlation coefficients are independent of the condition as they show how strong the correlations are within-subjects.

1 There were additional variables measured during the experiment to answer two more research questions that are unrelated to the presented topic. Due to one of these research questions, all participants had to complete two baseline measures of their emotional state and their voice before the interaction with the website and their statements were recorded in the think aloud conditions. To this account, all participants wore headsets during the experiment.
Table 4
The table shows the effects of the independent variables (see Eq. 1) on the dependent variables. t-values with an absolute value above 2 can be interpreted as significant (Kliegl et al., 2010) and are highlighted with bold font.

<table>
<thead>
<tr>
<th>SXT (subjectively experienced time)</th>
<th>Think Aloud</th>
<th>Usab.</th>
<th>Usab.: Think Aloud</th>
<th>Task A vs. B</th>
<th>Task B vs. C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate 5.57</td>
<td>0.24</td>
<td>2.54</td>
<td>-0.28</td>
<td>0.56</td>
<td>-0.47</td>
</tr>
<tr>
<td>SE</td>
<td>0.24</td>
<td>0.48</td>
<td>0.48</td>
<td>0.96</td>
<td>0.37</td>
</tr>
<tr>
<td>t-value</td>
<td>3.30</td>
<td>0.49</td>
<td>5.30</td>
<td>-0.29</td>
<td>1.54</td>
</tr>
<tr>
<td>Clock time (sec)</td>
<td>Estimate 200.31</td>
<td>-30.57</td>
<td>90.97</td>
<td>21.75</td>
<td>-115.71</td>
</tr>
<tr>
<td>SE</td>
<td>5.61</td>
<td>11.22</td>
<td>11.22</td>
<td>22.44</td>
<td>10.54</td>
</tr>
<tr>
<td>t-value</td>
<td>35.70</td>
<td>-2.72</td>
<td>8.11</td>
<td>0.97</td>
<td>-10.98</td>
</tr>
<tr>
<td>Estimated clock time (log[sec]) (log-transformed)</td>
<td>Estimate 5.23</td>
<td>-0.10</td>
<td>0.55</td>
<td>0.333</td>
<td>-0.46</td>
</tr>
<tr>
<td>SE</td>
<td>0.063</td>
<td>0.13</td>
<td>0.13</td>
<td>0.25</td>
<td>0.065</td>
</tr>
<tr>
<td>t-value</td>
<td>83.69</td>
<td>-0.78</td>
<td>4.41</td>
<td>1.33</td>
<td>-5.46</td>
</tr>
<tr>
<td>PTRs (perceived time ratios – log-transformed)</td>
<td>Estimate 0.11</td>
<td>0.15</td>
<td>0.06</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td>SE</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>0.04</td>
</tr>
<tr>
<td>t-value</td>
<td>2.14</td>
<td>1.17</td>
<td>-0.62</td>
<td>0.11</td>
<td>3.47</td>
</tr>
<tr>
<td>Subjective usability (log-transformed)</td>
<td>Estimate 1.27</td>
<td>0.03</td>
<td>-1.25</td>
<td>0.05</td>
<td>-0.07</td>
</tr>
<tr>
<td>SE</td>
<td>0.06</td>
<td>0.13</td>
<td>0.13</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>t-value</td>
<td>19.71</td>
<td>0.27</td>
<td>-9.73</td>
<td>0.19</td>
<td>-1.02</td>
</tr>
<tr>
<td>Attractiveness (log-transformed)</td>
<td>Estimate 1.17</td>
<td>0.11</td>
<td>-1.06</td>
<td>0.12</td>
<td>-0.13</td>
</tr>
<tr>
<td>SE</td>
<td>0.07</td>
<td>0.14</td>
<td>0.14</td>
<td>0.29</td>
<td>0.07</td>
</tr>
<tr>
<td>t-value</td>
<td>16.12</td>
<td>0.76</td>
<td>-7.33</td>
<td>0.40</td>
<td>-1.95</td>
</tr>
<tr>
<td>Valence</td>
<td>Estimate 8.89</td>
<td>0.25</td>
<td>-1.21</td>
<td>1.13</td>
<td>0.19</td>
</tr>
<tr>
<td>SE</td>
<td>0.20</td>
<td>0.40</td>
<td>0.40</td>
<td>0.79</td>
<td>0.20</td>
</tr>
<tr>
<td>t-value</td>
<td>24.73</td>
<td>0.63</td>
<td>-3.05</td>
<td>1.42</td>
<td>-0.96</td>
</tr>
<tr>
<td>Overall UX (log-transformed)</td>
<td>Estimate 1.21</td>
<td>0.06</td>
<td>-1.17</td>
<td>0.02</td>
<td>-0.11</td>
</tr>
<tr>
<td>SE</td>
<td>0.06</td>
<td>0.13</td>
<td>0.13</td>
<td>0.26</td>
<td>0.08</td>
</tr>
<tr>
<td>t-value</td>
<td>18.91</td>
<td>0.48</td>
<td>-9.16</td>
<td>0.07</td>
<td>-1.34</td>
</tr>
</tbody>
</table>

2.2.2. Effects of the experimental design
Table 4 and Fig. 1 summarize the effects of the independent variables think aloud and usability. Regarding the main dependent variable SXT, we found a main effect of usability; SXT was more negative in the groups with lower usability. We found no main effect of think aloud on SXT indicating that think aloud did not lead to a distortion in estimated clock time. An additional analysis of the only those tasks within the time limit showed only slight divergences from the analysis of all tasks. For PTR, the intercept failed to reach the significance level indicating that participants on average showed clock time estimates closed to the true duration. For attractiveness, the differences between the tasks became significant, while the main effect of usability slightly failed the significance level for valence. As these differences are rather small and do not affect the main conclusions, we will not discuss them further.

2.3. Discussion
In Study 1, we tested how subjectively experienced time (SXT) during an ongoing interaction contributes to UX. We introduced an easy way to measure SXT by using three simple rating items. Based on the UX framework of the CUE-model, we tested, first, whether this new measure would correlate with the user's perception and UX after an interaction with a website and, second, whether characteristics of the system (high vs. low usability) and the task (with vs. without think aloud) would show similar effects on SXT as on measures of time perception and UX. The results regarding the correlations were mostly substantial indicating that SXT is connected with both the users' perception of time and their UX. Results regarding the effects of the system and the task indicate that SXT can indeed be seen as an additional dimension of UX that fits into existing frameworks such as the CUE-model.

On the one side, this conclusion is based on the connection between SXT and time perception. As predicted by the CUE-model, variations in both the system's usability and the task led to changes in the objective interaction characteristic duration (clock time). Clock time was perceived and estimated quite adequately by the users. However, changes in clock time due to the varying task were not always properly represented by the user. This led to a distortion in estimated clock time, at least during the first task (see additional analysis of PTRs). Regarding the connection to SXT, we found similar effects of the experimental design on SXT and estimated clock time. More precisely, low compared to high usability caused a longer clock time which was perceived by the user and resulted in a less positive SXT. The increase of clock time due to the additional task of think aloud, however, was not properly perceived by the user (no effect of think aloud on estimated clock time and partial effect of think aloud on PTR) and, therefore, not translated into differences in SXT. These results are in line with the idea that SXT is built upon the user's perception of temporal attributes of the interaction, such as the total duration.

However, there were some discrepancies between SXT and the estimated clock time. First, the correlation between estimated clock time and SXT was only of medium strength suggesting that the SXT is based on more than just the perceived clock time. Second, differences in clock time due to the different tasks (task A to C) were perceived and mostly correctly estimated by the users. However, they were not represented in SXT. These discrepancies point towards other factors and processes influencing SXT such as integration into UX with an adjustment of SXT to high usability caused a longer clock time which was perceived by the user (no effect of think aloud on estimated clock time and partial effect of think aloud on PTR) and, therefore, not translated into differences in SXT. These results are in line with the idea that SXT is built upon the user's perception of temporal attributes of the interaction, such as the total duration.

On the other side, the notion that SXT can be seen as an additional dimension of UX is supported by its clear connection to the UX components. It correlated with all three components of UX and the overall UX. In addition, the experimental setting affected SXT and the UX components as well as the overall UX in the same fashion: A low compared to high usability website led to a less positive SXT and a decreased level of subjective usability, attractiveness, valence, and overall UX. Here, the different tasks had no effect on either SXT nor on UX measures.
Fig. 1. Effects of Think Aloud and Usability on the dependent variables. PTRs, subjective usability, attractiveness, and overall UX were log-transformed for the analysis. The graphs show the transformed values with retransformed y-axes to allow an interpretation in the original scales. Horizontal lines indicate perfect estimation for PTRs and the middle of the scale for SXT and UX-scales. All Error bars ( ± 2 SE) are corrected for dependent measures if applicable (Cousineau, 2005; Morey, 2008).
Liiikkanen and Gómez Gómez (2013) argued that SXT is a subjective assessment that reflects an overall experience and mediates global UX. The presented study provides empirical support for this claim by showing that, first, SXT evolves during an interaction in accordance with UX frameworks based on interaction characteristics and their perception. Second, SXT is integrated into UX components and the overall UX which was supported by both correlations and effects of the experimental design. Based on the results, SXT should be taken into account when designing for a positive UX as Liikkanen and Gómez Gómez (2013) already argued on a theoretical basis. Moreover, the study shows that SXT can be seen as a part of UX even when no delays or waiting times are included in an interaction.

2.3.1. Limitations
The generalizability of this conclusion, however, is limited due to two major points in Study 1. First, we introduced a new operationalization of SXT. Therefore, a second study is needed to validate the findings above. Moreover, SXT was measured with items concerning only the assessment of clock time. There are also other possible ways to operationalize SXT, for example, passage of time judgments. Such judgments ask for an evaluation of the passage of time that was experienced throughout the interaction. Several researchers showed that passage of time judgments are connected to both time perception as well as to measures of experience. As discussed earlier, passage of time judgments correlate at a medium strength with estimated clock time (Sucala et al., 2011). Other authors showed connections between passage of time judgments and emotions (e.g. Droit-Volet et al., 2017; Droit-Volet and Wearden, 2015; 2016) and between passage of time judgments and enjoyment (Sackett et al., 2010): With a faster passage of time, positive emotions and the enjoyment of the situations seem to increase. Taken together, earlier evidence points out that passage of time judgments might be a good addition to SXT. Therefore, we added passage of time judgments to the operationalization of SXT in Study 2.

The second limitation of Study 1 lay in the inconsistent results regarding a distortion in time perception due to think aloud. More precisely, we expected to find a main effect of think aloud on PTRs indicating that participants overestimate clock time less strongly in the groups performing think aloud compared to the ones not performing think aloud. Even though other authors already showed that classic think aloud leads to the expected distortion in time perception (Hertzum and Holmegaard, 2015), we could show this effect for relaxed think aloud only for the first task. The disappearance of the effect after the first task could be, however, due to an adaptation to the think aloud task. Such diminished distortions due to practice or adaptation have already been described in basic literature as the attenuation effect (Brown, 2008). The attenuation effect is ascribed to the fact that a trained secondary task will require less attentional resources than an untrained one. Hence, the temporal distortion due to a secondary task grows weaker when the secondary task is practiced more. In Study 1, thinking aloud was relatively new and unfamiliar during the first task. In the following tasks, the participants had grown more familiar with it. Moreover, the website was the same for all three tasks and it had the same usability bugs in case of the low usability website. Thus, the same difficulties had to be verbalized for each task leading to an increasing level of practice. Hence, the second and third tasks might have confronted the participants performing think aloud with a lower level of attentional demands compared to the first task. This could explain why the difference in temporal distortion between the think aloud groups was diminished after the first task. To test this explanation we conducted Study 2 where we included a wider variability in the interaction itself to prevent participants to verbalize the same issues during all tasks.

3. Study 2

Study 2 was designed to validate and expand the conclusions from Study 1. Therefore, the design of Study 2 was similar to Study 1 to allow replication. Such replication can be seen as further evidence for the connection between SXT and time perception resp. UX. Also, we adjusted the operationalization of SXT by including one more item regarding passage of time. Moreover, we increased both website and task variety to eliminate an attenuation effect in the think aloud condition.

3.1. Method

3.1.1. Subjects
Study 2 consisted of a sample of N = 80 participants (40 male, 40 female) at the age of M = 25.13 years (SD = 3.35) between 18 and 33 years. As in Study 1, most participants were students at the Technische Universität Berlin and could earn course credit or money (8 Euros) by participating. Again, all participants gave their informed consent before participating. Prior to Study 2, the ethical committee of the Technische Universität Berlin reviewed the research project and gave a positive statement.

3.1.2. Setting and design
The experimental design was the same as in Study 1. Hence, the study consisted of a two-factorial design with the independent between-variables usability (high vs. low) and think aloud (with vs. without think aloud). The dependent variables measured after each task were the same as in Study 1. They comprised SXT, clock time, estimated task duration, PTRs, subjective usability, attractiveness, valence, and overall UX. A passage of time judgment ranging from Time during the task passed by quickly (1) to slowly (11) was added as an additional item measuring SXT.

3.1.3. Procedure and material
The procedure was the same as in Study 1, but the website and the tasks were replaced. To this end, we created two versions of a website for an online cloud-service differing in usability but not in content. Usability problems in the low-usability website were tested beforehand iteratively during implementation (N = 12). To increase the length of the interaction, a total of four tasks was implemented. The tasks had to be solved in a counterbalanced order. Due to the new website, the tasks differed to the tasks in Study 1.

During the introduction to the study, participants were told to imagine working in a big firm where a colleague had asked them to take over some tasks. On this account, they were logged in on the website of an alleged cloud service with the colleague’s account. The tasks to be accomplished were A finding a file, extracting specific information from it and change another files name, B finding out who was responsible for the support of the alleged cloud service, C changing the colleague’s personal information (e.g. telephone number) stored on the cloud service, and D deleting specific pictures from a folder before sharing the folder. All tasks required participants to interact with different parts of the website and for the low usability condition, different usability problems were implemented for each task. These were, for example, for Task A nontransparent folder structures and hardly accessible information, for Task B automated forwarding to another page hindering the participants to conclude their search, for Task C small fonts with ambiguous names, interchanged positions of links (e.g. save and cancel), misleading colors (green cancel button), and the occurrence of an error, and for Task D relevant information at uncommon places and inappropriate sorting of the pictures. If participants failed to complete a task after four minutes they were kindly instructed to stop their attempt and proceed with the questionnaire before attending to the next task.

Overall, the experiment lasted about 40 min. Again, there were additional variables measured during the experiment which will not be reported here. Due to one of them, all participants wore headphones during the experiment. Age, sex, education, technophilia, and course of study were again measured at the end of the experiment.
3.1.4. Data preparation and analysis
The data preparation and analysis were the same as in Study 1. Similar to Study 1, we conducted principal component analyses for each point of measurement (first to fourth task) and for each task (task A to D) to reassure that all four items measuring SXT (short/long, appropriate/inappropriate, acceptable/unacceptable, quickly/slowly) could be summarized in one component. All eight analyses showed one underlying factor indicating that averaging over all four items were justified. Also in the present study, subjective usability and overall evaluation showed substantial correlation with the corresponding scales from the meCUE (Minge et al., 2016) which was administered at the end of the experiment, \( r = 0.84, p < 0.001 \), resp. \( r = 0.87, p < 0.001 \). Attractiveness and visual aesthetics only correlated at a medium effect size, \( r = 0.47, p < 0.001 \).

Again, we checked how many tasks were interrupted due to the time limit of four minutes. The limit was reached 18 times at Task A, 4 times at Task B, 7 times at Task C, and 14 times at Task D leading to a total of 43 of all 480 = 320 tasks. For the analysis, we included all 320 tasks. Again, an additional analysis with only those tasks that did not reach the limit yielded only small differences in the significant effects which will be discussed later.

3.2. Results
3.2.1. Correlations
Similar to Study 1, we first looked at correlations between SXT and the other dependent variables. The within-subject correlations are shown in Table 5. Again, they describe how two variables covary within participants. As expected based on Study 1, SXT showed positive correlations to clock time and estimated clock time. With increasing clock time and estimated clock time, the participants evaluated SXT more positively. The small correlation between SXT and PTR from Study 1 was not replicated. Regarding UX variables, we found negative correlations of similar strength to Study 1 between SXT and subjective usability, attractiveness, valence, and overall UX. Hence, the more positive participants assessed their SXT, the more positive ratings they gave on the UX scales.

3.2.2. Effects of the experimental design
The effects of the independent variables think aloud and usability on the dependent variables are summarized in Table 6 and Fig. 2. SXT was influenced by the usability of the website. Comparable to the results of Study 1, low usability led to a more negative SXT. There was also a significant interaction between think aloud and usability indicating that the effect of usability on SXT was not as strong for participants performing think aloud compared to participants not performing think aloud.

Concerning the temporal variables clock time, estimated clock time, and PTRs a log-transformation was necessary to adjust for a violation of normally distributed residuals. Usability showed a similar effect on clock time and estimated clock time as in Study 1: Low usability led to an increase in both variables. Performing think aloud also increased clock time but did not affect estimated clock time. This finding was supported by the expected main effect of think aloud on PTRs. While participants performing think aloud showed a rather accurate estimation, participants not performing think aloud overestimated clock time strongly. Study 2 also revealed significant interactions between think aloud and usability for both clock time and estimated clock time. The pattern of this effect on estimated clock time closely mirrored the interaction-effect on SXT (see Fig 2).

Regarding UX variables, we replicated the strong effect of usability on UX components and on overall UX. In the low usability groups, participants gave lower ratings for subjective usability, attractiveness, valence, and overall UX. There was neither an effect of think aloud nor of the interaction on these variables. The effects of the experimental design on UX variables and overall UX were overall rather similar to the effects on SXT with differences only in the interaction and the effects of the different tasks.

The additional analysis of the only those trials that were concluded within the time limit, showed very similar results and conclusions to the analysis of all trials. For SXT, the interaction between think aloud and usability slightly failed the significance level while the difference between task A and B became significant. For clock time, the difference between tasks B and C was not significant, while this difference reached significance for valence.

3.3. Discussion
The aim of Study 2 was to replicate the results from Study 1 regarding the correlations between SXT, time perception, and UX variables as well as the effects of the factors usability and think aloud on
Fig. 2. Effects of Think Aloud and Usability on the dependent variables. Clock time, estimated clock time, and PTRs were log-transformed for the analysis. The graphs show the transformed values with retransformed y-axes to allow an interpretation in the original scales. Horizontal lines indicate perfect estimation for PTRs and middle of the scale for total duration assessment and UX-scales. All Error bars (± 2 SE) are corrected for dependent measures if applicable (Cousineau, 2005; Morey, 2008).
these variables. Moreover, the stimulus material was adjusted to hinder participants to grow too familiar with the presented website. The results mainly replicated the correlations between SXT and time perception resp. UX as well as the effects of usability and think aloud indicating that the conclusions of Study 1 also hold for Study 2. The results are discussed in detail in the general discussion.

4. General discussion

Perceiving time is a constant experience during everyday life and, by implication, during the interaction with technical devices. Even though objective temporal characteristics of an interaction, such as clock time, have long been used as objective measures for the evaluation of the quality of an interaction (Albert et al., 2009; Blackler et al., 2010; DIN EN ISO 9241-210, 2010; Frøkjær et al., 2000; Lesemann et al., 2007), the subjective perception of time during an interaction has hardly been studied. In this paper, we argue that temporal aspects of an interaction are constantly perceived by users. Therefore, time perception offers easily accessible information about the interaction to the user who integrates this information into his/her user experience (UX). We claim that measuring and understanding subjectively experienced time (SXT) offers a valuable addition to the methodological UX toolkit as it describes how an objective characteristic of the interaction is translated into UX.

We tested this claim based on two questions: First, how can SXT be measured during a typical interaction with a computer? An easy and economical way to measure SXT is a precondition to use SXT in UX research. Based on the literature on perceived waiting time, we derived four items to measure SXT regarding assessments of total clock time and passage of time. Hence, the two presented studies provide an easy way to include SXT measures in any UX testing scenario. Future research can rely on this four-item scale, for example, to further understand how temporal perception contributes to UX or to evaluate design choices regarding designing for a positive experience of time during an interaction.

Regarding the second question, we aimed to answer how SXT relates to temporal aspects of the interaction as well as to UX? Based on the CUE-model (Minge and Thüring, 2018; Thüring and Mahlke, 2007), SXT should be based on the perception of temporal characteristics of the interaction and integrated into UX. We tested this claim with the help of two studies in which participants interacted with either a low or a high usability website. In addition, half of the participants performed think aloud during the interaction. Hence, we experimentally varied both the system and the task to create differences in the objective characteristics of the interaction. We analyzed how SXT covaried with both time perception and UX within-subjects and whether SXT was influenced by the same experimental factors as the adjacent constructs.

Concerning time perception, we found the expected correlations between SXT and the temporal variables clock time and estimated clock time. More precisely, if participants experienced a longer task compared to a shorter task, they would not only estimate this task to be longer but also give a more negative SXT rating. This relation was further supported by the effect of the system’s usability: Low compared to high usability increased clock time in both studies. The increase in clock time was perceived by the participants and translated into a change in SXT regarding both dimensions. However, the estimation of clock time was not always correct and this error in estimation was also translated into SXT. More precisely, think aloud also caused an increase in clock time. This was, however, neither reflected in the participants’ estimated clock times nor in their SXT. This distortion in time perception was expected based on cognitive models of time perception. The attentional-gate-model (Block and Zakay, 1996) predicts that additional attentional demands redirect attention away from perceiving time and thereby shorten prospective time estimations. In both studies, we increased additional attentional demands via think aloud and found the expected effect, in Study 1 partially and in Study 2 fully. Regarding our research question, we conclude that SXT is based upon the perception of the temporal aspects of the interaction. This perception, however, is not always reflecting the objective “truth” as it can be influenced by distorting factors, such as attentional demands.

Overall, the results support the notion that SXT is based on the perception of temporal characteristics of the interaction as argued by the CUE-model. However, none of the correlations between SXT and temporal variables could be categorized as strong. This is in line with weak correlations or effects regarding perceived waiting time and estimated clock time (Antonides et al., 2002) or passage of time and estimated clock time (Sucala et al., 2011). Hence, the data suggest that SXT is not purely based on time perception but also influenced by other aspects of the interaction. Such influences would show in correlations and similar effect patterns on SXT with UX variables. Indeed, the results of both studies showed correlations between SXT and UX components as well as overall UX. A more positively assessed SXT was associated with higher ratings on subjective usability, attractiveness of the system, valence, and overall UX within-subjects. Moreover, the strong effect of the experimental usability manipulation showed a similar effect pattern on SXT and UX variables in both studies.

To answer the second research question regarding the relation between SXT and temporal aspects respectively UX, we can summarize that SXT fits in existing frameworks of UX. That means that our data is consistent with the idea that SXT is at least partially based on the perception of temporal characteristics of an interaction. Moreover, our data show that SXT is more than just a mere estimation of clock time. It reflects the subjectivity of UX but is distinct from other UX measures. Hence, it is a valuable addition to the UX methodological toolkit.

4.1. Implications

Overall, the results indicate that SXT can be useful as a UX evaluation tool. Such an evaluation tool may have several benefits. First, when designing for a positive UX, a special focus may be put on the temporal experience of the user. This could be relevant, for example, when users should not be aware of how much time they spend on a website. Our results show that a mere focus on objective clock time is not enough to portray subjectively experienced time. In contrast, evaluating temporal aspects of UX needs to be subjective in nature. The provided items for SXT offer an easy way to evaluate temporal UX. Second, SXT is very easy to understand for users. Whereas the term subjective usability or visual attractiveness may be complex or unknown to naive users, time is a mundane concept and can be evaluated and verbalized very quickly and intuitively. Moreover, SXT does not need a specific time unit and is widely applicable without adaptation to the length of the interaction period. Third, SXT can be used when the evaluation of the technology should not be obvious to the user. The perception of time may act as a cover story for the UX evaluation when participants are not to be aware of the evaluation or if an acquiescence bias is anticipated. Future research, however, needs to show further evidence to support these claims and areas of application.

A second implication concerns the use of think aloud in human–computer studies. The results of both studies revealed that thinking aloud increased the objective clock time. The evidence supports the recommendation of Tullis and Albert (2010) not to pair think aloud with objective temporal measures in usability testing to avoid biases. Moreover, the results indicated that relaxed think aloud as commonly used in UX testing scenarios (Hertzum et al., 2009) increased workload. Thus, the effect is in line with predictions from basic research (attentional-gate model, Block and Zakay, 1996) and prior research on the effect of classic thinking aloud on time perception (Hertzum and Holmegaard, 2015). The results strongly support the notion that think aloud should always be seen as a secondary task that can interfere with the primary task of the interaction. Such interference might be especially relevant when the primary task shares the same resources as the think-aloud task (Baddeley, 2003; Wickens et al., 2012), for example,
speech production or verbal understanding. Researchers have to take
this into account when they interpret the results of studies using con-
current think-aloud protocols. Even though this finding is not in the
center of the presented research question, it is of great importance for
the field.

4.2. Limitations

The presented studies have several limitations regarding the gen-
eralizability of the results, the missing assignment of SXT to a specific
component of UX, and the possible confounding of the independent
variables usability and think aloud.

The generalizability of the results is limited by the fact that we fo-
cused on task-driven, non-taxing interactions with a website. Hence,
future research needs to check and validate the results for other con-
texts of human-computer interaction such as free explorations of web-
sites or programs and games. Computer games, for example, might
involve different modes of time processing due to their immersive
nature (Wood et al., 2007). Moreover, the experimental design limits
the generalizability as well, as each subject was assigned to one group
in our two-by-two between-subject design. To put it more precisely,
each subject experienced only one level of usability that could vary
between tasks but did not represent the whole usability spectrum.
Hence, when looking at within-subject correlations, the correlations are
probably based on a small variance of usability which might have led to
an underestimation of correlations between SXT and UX variables. In
future research, participants should be confronted with a wider range of
usability levels. The usability manipulation also yielded another lim-
itation to the interpretation of the results. The system’s usability showed
effects on both clock time and UX components. Hence, the effects of
usability on SXT could be based on either the different temporal char-
acteristics of the interaction (e.g. clock time) or on other characteristics
of the interaction (e.g. visual aesthetics, usefulness) or even both. Based
on the presented experimental design, we cannot separate these effects.
To do so, one would need to vary the one without the other. UX, for
example, could be varied by visual components of the device that show
no effect on clock time but should affect UX. Keeping the usability of
the system as one independent variable, on the other side, would re-
quire to find a task that can vary in usability without varying in clock
time. Here, a tracking task might be helpful. However, such a tracking
task would be rather different from typical tasks in usability testing
scenarios. Future research is needed to separate the effects of temporal
characteristics of the interaction from the effects of other evaluating
dimensions such as UX components on SXT.

The second major limitation lies in a missing assignment of SXT to a
specific component of UX. The studies leave open to which component
of UX SXT could be ascribed. In both studies, SXT correlated with all
three UX components, that is with the perception of instrumental
qualities, the perception of non-instrumental qualities, and emotions.
The highest correlations were found to the valence which represents the
emotional component. However, SXT is not an emotion and should
therefore not be ascribed to this component. Regarding the components
perception of instrumental and perception of non-instrumental quali-
ties, higher correlations were found for the construct representing in-
strumental qualities (subjective usability). But, the difference in the
relation between the outcome and the required resources including
cognitive effort (Bevan, 2009). Thus, the usability manipulation might
have confounded the effects of the think aloud groups. Additionally,
another confounding might stem from an effect of think aloud on the
usability manipulation. When performing think aloud, usability pro-
blems have to be verbalized which can make these problems more
lucid. Hence, adding think aloud might have confounded the effects of
the usability manipulation. Even though we cannot exclude the possi-
bility that usability and think aloud confounded each other, there is no
evidence in the data pointing towards such a relevant confounding. This
conclusion is based on the two non-significant effects. First, usability
showed no effect on PTRs indicating that the usability manipulation did
not affect the participant’s ability to perceive time significantly. The
ability to perceive time is, however, very strongly and reliably affected
by additional attentional demands (Brown, 2008). Hence, in our
sample, the difference in attentional demands between the usability
groups must have been rather small. Second, we found no effect of the
interaction between usability and think aloud on subjective usability or
any other UX-based variable. If the usability manipulation had been
perceived differently in the think aloud groups, such an interaction
effect should have been found. Overall, even though there might have
been a confound in our experimental design, the data suggest no sub-
stantial effects of such a confound.

A fourth limitation could lay in the assumed motivation of the
participants. In both experiments, we set a time limit for each task. If
participants did not finish the task within the given limit, they would be
asked to abort the task and continue with the questionnaire. In our
analysis, we did not exclude those tasks. However, an additional ana-
lysis with only those tasks within the time limit did not show different
conclusions. Moreover, the UX was mostly unaffected by whether a task
was finished within the time limit or not. Hence, one could argue that
participants were quite detached and approached the study tasks dif-
fently from how they would approach real-world tasks.

5. Conclusion

We argue that users perceive time during most of their daily inter-
actions with technology and that the subjectively experienced time
(SXT) contributes to UX. In this paper, we asked two research questions
regarding, first, how SXT can be measured and, second, how it relates to
temporal aspects of the interaction as well as to UX. The results showed
that, first, SXT can be measured via bipolar items based on total
duration assessments and passage of time judgments. Second, our data
support that SXT is based on the perception of temporal characteristics
of the interaction and integrated into UX. Hence, basic cognitive models
of time perception help to predict how users are going to estimate time
during an interaction. However, evaluating SXT is more than just a
mere representation of estimated clock time as it shows close connec-
tions to different components of UX. Therefore, research on UX can gain
further insight into how UX evolves by looking at SXT. Moreover, SXT
enriches the methodological toolbox for UX practitioners as it allows an
easy way to measure attributes of UX that have been mostly neglected
in UX research so far.

Data availability

The dataset and the analysis supporting the conclusions of this ar-
cative can be accessed via osf.io: https://osf.io/3c9xk/.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest
associated with this publication and there has been no significant fi-
nancial support for this work that could have influenced its outcome.
Anna K. Trapp earned her Ph.D. at the chair of cognitive psychology and cognitive ergonomics at the Technische Universität Berlin. Her research interests include applying cognitive psychology to human computer interactions and understanding how users evaluate experiences with technical devices. Her main research interests are time perception, user experience, and visual search in the context of human-computer interactions.

Nils Backhaus studied psychology at Ruhr-Universität Bochum and Human Factors at Technische Universität Berlin. He is with the unit "Changing World of Work" at the Federal Institute for Occupational Safety and Health in Dortmund. His Ph.D. thesis focused on psychological aspects in human-computer interaction like trust, emotions, and user experience. He combines research approaches based in psychology, social sciences, and ergonomics.

Roland Thomaschke is head of the research group "Time, Interaction, Self-determination" at Albert-Ludwigs-Universität Freiburg (Germany). He studied Cognitive Sciences, Computation & Linguistics, Mathematics, and Psychological Research Methods. After obtaining his PhD in Psychology from Lancaster University (United Kingdom), he worked as a postdoc researcher at universities of Würzburg (Germany), Regensburg (Germany), and Freiburg (Germany). His main research interests are human timing, human-machine interaction, empirical museology, and the psychology of architecture.