International Journal of Human-Computer Studies 138 (2020) 102399

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Measuring subjectively experienced time in usability and user experience testing scenarios☆

Anna K. Trukenbrod[a,*], Nils Backhaus[b], Roland Thomaschke[c]

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

ARTICLE INFO

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

ABSTRACT

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 Tangmanee, 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; Pohlmeyer 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 Tangmanee, 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,

Abbreviations: PTR(s), Perceived Time Ratio(s); SXT, Subjectively Experienced Time; UX, User Experience

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.
E-mail address: anna.k.trapp@tu-berlin.de (A.K. Trukenbrod).

https://doi.org/10.1016/j.ijhcs.2020.102399
Received 14 November 2019; Accepted 26 January 2020
Available online 27 January 2020
1071-5819/

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 interaction 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). These items can be categorized into items asking for an assessment of the waited clock time, items asking for the passage of time during the wait and items concerning the affective state of the subject.

Table 1
Different items to measure perceived waiting time.

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

For the operationalization of SXT, items concerning the assessment of the waited clock time and the passage of time during the wait are good candidates to represent SXT. Items refering to the affective state of the user, however, are not. When including SXT in UX research, it should be distinct from other common UX measures such as emotions (Agarwal and Meyer, 2009; Hudlicka, 2003; Mahlke, 2008). Thus, those items should not be used to measure SXT to avoid the mingling of constructs.

Summoning up, we define SXT as the user's evaluation of temporal attributes of an interaction. It can be assessed in terms of subjective ratings on temporal attributes (e.g. long - short, unacceptable - acceptable, quick - not quick).

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. Intergrating SXT in a framework of UX

One important reductionist framework of UX is the CUE model (Components of User Experience, Minge and Thüring, 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; Minge and Thüring, 2018; Sonderegger and Sauer, 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. Ben-Bassat et al., 2006; van der Geest and van Dongelen, 2009; Hamborg et al., 2014; Minge and Thüring, 2018), but also the perception of non-instrumental qualities (e.g. Hamborg et al., 2014; Minge and Thüring, 2018; Sonderegger and Sauer, 2010; Tractinsky et al., 2000) and the emotions of the user (e.g. Minge and Thüring, 2018; Thüring and Mahlke, 2007). Hence, when varying the usability of a system, this should affect SXT and UX components in the same fashion. Such a finding would indicate that SXT indeed is a part of UX.

1.2.2. SXT and time perception

Based on the CUE model, we argued that the objective clock time of an interaction is perceived and evaluated by the user, but how are objective clock time, perceived clock time, and its evaluation connected? In timing research, two paradigms are discriminated: Prospective and retrospective timing. The distinction between the two paradigms is important because they are based on different cognitive processes and, hence, influenced by different factors. The focus of this paper lies in the prospective timing paradigm. Prospective timing emphasizes that an attentional focus lies on the timing task during the temporal interval (for a review and comparison of the two paradigms see e.g. Block and Zakay, 1997). The prospective timing paradigm fits situations in which time is critical, such as the interactions with technical devices this paper focuses on.

The prospective perception of time is often operationalized by a verbal estimate in minutes and seconds. For such estimates, it has been frequently shown that longer clock times will lead to longer estimations (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 testings (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 homogeneous 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.

Table 2
Overview of the dependent variables in Study 1.

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

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, $t(18) = 6.64$, $p < 0.001$, $d = 3.1$, 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 purpose. During the whole experiment, no temporal cues were given to the participants. Overall, the experiment lasted about 30 min.[1]

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.

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 ggplot2 package was used (Wickham, 2009). A linear mixed effect model was fitted for each dependent variable (DV) based on the following formula:

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

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.

Table 3
Within-subject correlation coefficients r between SXT and the other variables. Based on Cohen (1992) a correlation of $0.1 \le |r| < 0.3$ can be seen as a small effect, of $0.3 \le |r| < 0.5$ as a medium effect, and $|r| \ge 0.5$ as a large effect.

Within-subject correlations	SXT
Clock time	r = 0.27, p < 0.01
Estimated clock time	r = 0.41, p < 0.001
PTR	r = 0.21, p < 0.05
Subj. usability	r = -0.43, p < 0.001
Attractiveness	r = -0.34, p < 0.001
Valence	r = -0.56, p < 0.001
Overall UX	r = -0.40, p < 0.001

[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.

	Inter-cept	Think Aloud	Usab.	Usab.: Think Aloud	Task A vs. B	Task B vs. C
SXT (subjectively experienced time)						
Estimate	5.57	0.24	2.54	-0.28	0.56	-0.47
SE	0.24	0.48	0.48	0.96	0.37	0.37
t-value	**23.30**	0.49	**5.30**	-0.29	1.54	-1.29
Clock time [sec]						
Estimate	200.31	-30.57	90.97	21.75	-115.71	-30.69
SE	5.61	11.22	11.22	22.44	10.54	10.59
t-value	**35.70**	**-2.72**	**8.11**	0.97	**-10.98**	**-2.90**
Estimated clock time [log(sec)] (log-transformed)						
Estimate	5.23	-0.10	0.55	0.333	-0.46	-0.25
SE	0.063	0.13	0.13	0.25	0.085	0.084
t-value	**83.69**	-0.78	**4.41**	1.33	**-5.46**	**-2.98**
PTRs (perceived time ratios – log-transformed)						
Estimate	0.11	0.12	-0.06	0.02	0.15	0.03
SE	0.05	0.10	0.10	0.20	0.04	0.04
t-value	**2.14**	1.17	-0.62	0.11	**3.47**	0.72
Subjective usability (log-transformed)						
Estimate	1.27	0.03	-1.25	0.05	-0.07	0.06
SE	0.06	0.13	0.13	0.26	0.07	0.07
t-value	**19.71**	0.27	**-9.73**	0.19	-1.02	0.78
Attractiveness (log-transformed)						
Estimate	1.17	0.11	-1.06	0.12	-0.13	0.12
SE	0.07	0.14	0.14	0.29	0.07	0.07
t-value	**16.12**	0.76	**-7.33**	0.40	-1.95	1.77
Valence						
Estimate	4.89	0.25	-1.21	1.13	-0.19	0.35
SE	0.20	0.40	0.40	0.79	0.20	0.20
t-value	**24.73**	0.63	**-3.05**	1.42	-0.96	1.77
Overall UX (log-transformed)						
Estimate	1.21	0.06	-1.17	0.02	-0.11	0.06
SE	0.06	0.13	0.13	0.26	0.08	0.08
t-value	**18.91**	0.48	**-9.16**	0.07	-1.34	0.79

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 working with the low usability website. There was no significant effect of think aloud, of the interaction nor the different tasks. The intercept indicates that ratings on SXT were on average significantly different from zero with a medium level of $M = 5.57$ on the 11-point scale.

Clock time was on average about $M = 200.31$ s per task. The main effect of usability was mirrored in longer clock time (by $M = 90.97$ s see Table 4) and longer estimated clock time in the low- compared to the high-usability groups. Yet, clock time also significantly increased by $M = 30.57$ s (see Table 4) when performing think aloud compared to not performing think aloud. This effect did not translate into the estimated clock time indicating the expected distortion in time perception due to additional attentional demands when performing think aloud. However, we found no main effect of think aloud on PTRs indicating that think aloud did not lead to a distortion in estimated clock time. An additional explorative analysis of PTR with help of a directed post hoc *t*-test with a Bonferroni-Holm-correction revealed that think aloud showed the expected effect of temporal distortion during the first task only, $t(60) = 2.591$, $p = 0.018$, but not for the second and third task, *ts* $(60) < 1$, $ps > 0.5$. Hence, the differences in clock time between the usability groups and the think aloud groups were perceived rather accurately, but there were distortions due to think aloud during the first task. Moreover, the different tasks varied significantly in their total duration which was mirrored in the estimated clock time but not in SXT.

Regarding the components of UX and the overall UX evaluation, we found the expected main effect of usability on all four scales: The high-usability website was rated to be more usable, more attractive, to elicit a more positive valence and to have a better overall UX than the low-usability website. Similar to SXT, there was neither a main effect of think aloud nor a significant interaction effect resp. an effect of the different tasks on UX variables. Hence, the experimental design showed the same effects on UX as on SXT.

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 time 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 towards the components of UX.

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.

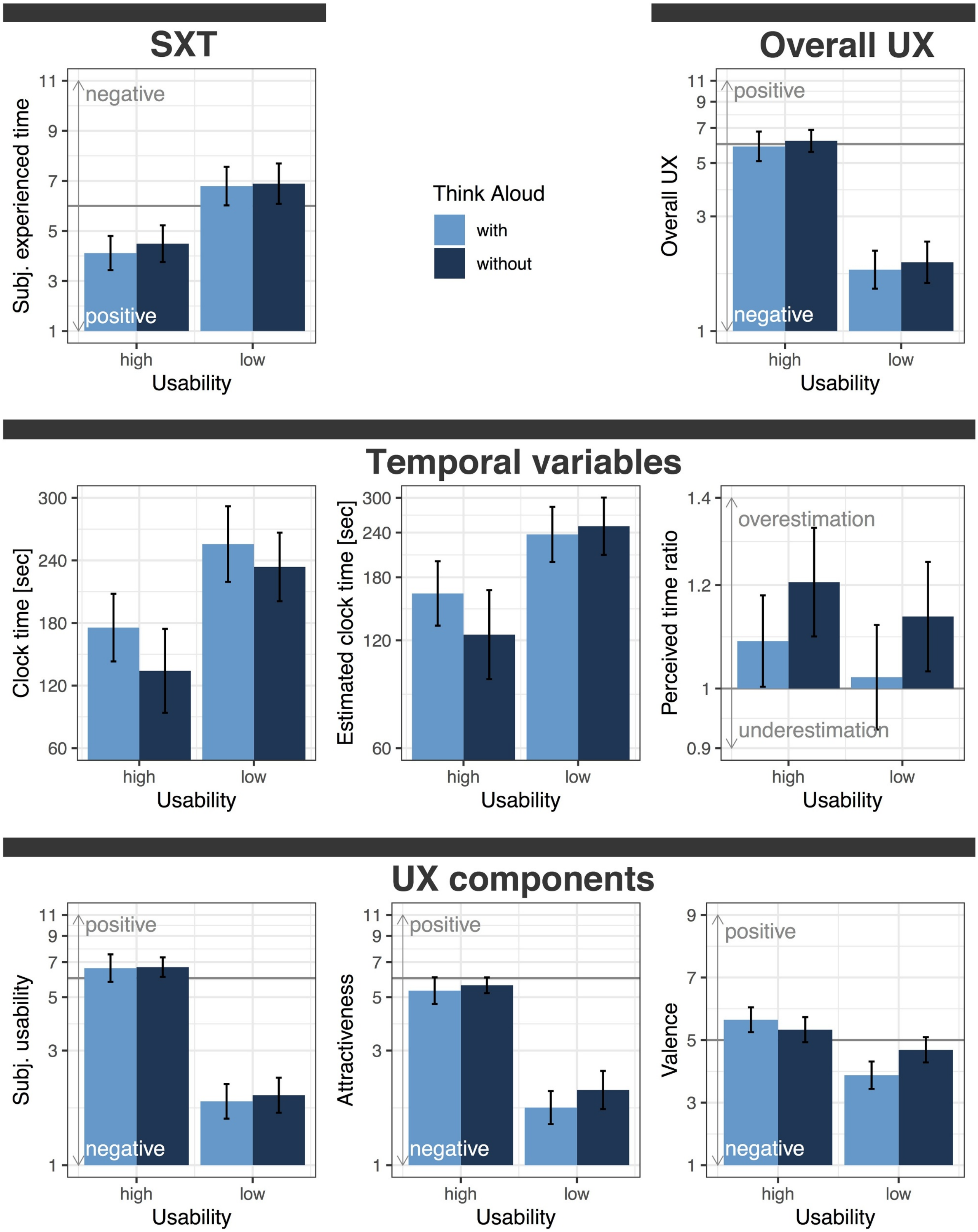

Fig. 1. Effects of Think Aloud and Usability on the dependent variables. PTRs, subjective usability, attractiveness, and overall UX were log-transformed for the analysis. The graphs show the transformed values with retransformed y-axes to allow an interpretation in the original scales. Horizontal lines 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).

Liikkanen 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 headsets 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 4*80 = 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.

Table 5
Within-subject correlation coefficients r between SXT and the other variables. Based on Cohen (1992) a correlation of $0.1 \le |r| < 0.3$ can be seen as a small effect, of $0.3 \le |r| < 0.5$ as a medium effect, and $|r| \ge 0.5$ as a large effect.

Within-subject correlations	SXT
Clock time	r = 0.47, p < 0.01
Estimated clock time	r = 0.48, p < 0.001
PTR	r = 0.07, not significant
Subj. usability	r = -0.46, p < 0.001
Attractiveness	r = -0.38, p < 0.001
Valence	r = -0.51, p < 0.001
Overall UX	r = -0.49, p < 0.001

Table 6
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.

	Inter-cept	Think Aloud	Usab.	Usab.: Think Aloud	Task A vs. B	Task B vs. C	Task C vs. D
SXT (subjectively experienced time)							
Estimate	5.18	-0.18	2.51	1.35	0.39	-1.52	0.57
SE	0.14	0.27	0.27	0.54	0.26	0.26	0.26
t-value	**37.99**	-0.67	**9.29**	**2.49**	1.50	**-5.86**	**2.18**
Clock time [log(sec)] (log-transformed)							
Estimate	4.86	-0.30	0.42	0.41	-0.29	0.09	0.25
SE	0.03	0.06	0.06	0.11	0.05	0.05	0.05
t-value	**173.65**	**-5.30**	**7.48**	**3.69**	**-6.46**	**2.10**	**5.62**
Estimated clock time [log(sec)] (log-transformed)							
Estimate	5.01	-0.09	0.37	0.53	-0.24	0.01	0.21
SE	0.05	0.09	0.09	0.18	0.06	0.06	0.06
t-value	**110.02**	-0.93	**4.08**	**2.92**	**-3.75**	0.16	**3.30**
PTRs (perceived time ratios – log-transformed)							
Estimate	0.15	0.21	-0.05	0.11	0.05	-0.09	-0.04
SE	0.04	0.08	0.08	0.16	0.05	0.05	0.05
t-value	**3.96**	**2.75**	-0.64	0.73	1.02	-1.68	-0.79
Subjective usability							
Estimate	6.93	0.26	-3.63	-0.80	-1.07	1.29	-0.12
SE	0.21	0.41	0.41	0.82	0.29	0.29	0.29
t-value	**33.65**	0.64	**-8.80**	-0.97	**-3.77**	**4.52**	-0.44
Attractiveness							
Estimate	6.79	0.06	-2.44	-1.19	-1.06	1.15	-0.30
SE	0.23	0.46	0.46	0.93	0.24	0.24	0.24
t-value	**29.30**	0.12	**-5.27**	-1.28	**-4.43**	**4.79**	-1.25
Valence							
Estimate	5.73	-0.33	-1.08	-0.26	0.19	0.29	-0.15
SE	0.14	0.27	0.27	0.54	0.17	0.17	0.17
t-value	**42.27**	-1.22	**-3.99**	-0.48	1.12	1.71	-0.89
Overall UX							
Estimate	6.70	0.20	-3.54	-0.83	-1.12	1.36	-0.15
SE	0.19	0.39	0.39	0.78	0.25	0.25	0.25
t-value	**34.56**	0.52	**-9.12**	-1.06	**-4.57**	**5.53**	-0.61

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

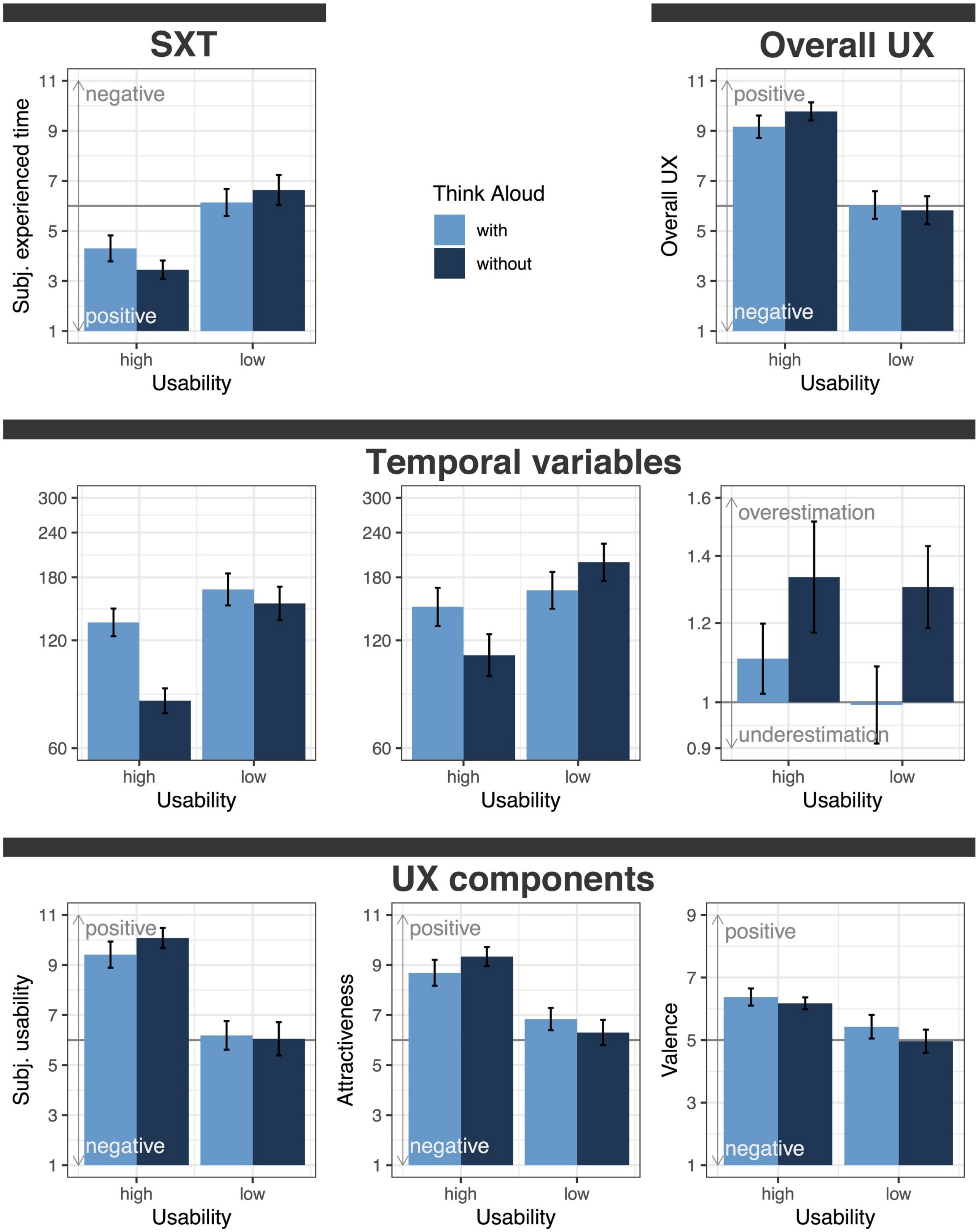

Fig. 2. Effects of Think Aloud and Usability on the dependent variables. Clock time, estimated clock time, and PTRs were log-transformed for the analysis. The graphs show the transformed values with retransformed y-axes to allow an interpretation in the original scales. Horizontal lines indicate perfect 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 distorion 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 concurrent 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 generalizability 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 focused on task-driven, non-taxing interactions with a website. Hence, future research needs to check and validate the results for other contexts of human-computer interaction such as free explorations of websites 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 limitation 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 characteristics 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 require 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 qualities, higher correlations were found for the construct representing instrumental qualities (subjective usability). But, the difference in the correlation coefficient was rather small. Hence, we can only derive the hypothesis that SXT could be seen as part of the perception of instrumental qualities of the interaction but not test it in the current design. Future research should test this hypothesis, preferably by pulling apart the measurement of instrumental qualities from measuring SXT and the other UX components to avoid a common method bias.

A third limitation could be seen in a possible confounding within the experimental design. We varied think aloud to add attentional demands in the think aloud groups. However, one could argue that confronting a user with low usability can also add attentional demands because low usability can lead to more uncertainty or longer search times during an interaction. Moreover, usability is often defined in terms of efficiency (e.g. DIN EN ISO 9241-11, 2018), which represents 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 problems 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 possibility 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 substantial 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 analysis 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 differently from how they would approach real-world tasks.

5. Conclusion

We argue that users perceive time during most of their daily interactions 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 connections 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 article 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 financial support for this work that could have influenced its outcome.

Acknowledgements

We thank Julien Tanardi, Florian Grieb, Kevin Nürck and Lena Mrowetz for their support in programming the website of study one and data collection within the scope of their Bachelor theses.

References

Abbasi, M.Q., Lew, P., Rafique, I., Weng, J., Wang, Y., 2012. Modeling User Experience. Proceedings of the ICCGI 2012: The Seventh International Multi-Conference on Computing in the Global Information Technology. pp. 63–69.
Agarwal, A., Meyer, A., 2009. Beyond usability: evaluating emotional response as an integral part of the user experience. CHI'09 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 2919–2930.
Albert, W., Tullis, T., Tedesco, D., 2009. Beyond the Usability Lab: Conducting Large-scale Online User Experience Studies. Morgan Kaufmann, Burlington, MA.
Antonides, G., Verhoef, P., A., 2002. Consumer perception and evaluation of waiting time : a field experiment. J. Consumer Psychol. 12 (3), 193–202. https://doi.org/10.1207/S15327663JCP1203_02.
Aranyi, G., van Schaik, P., 2015. Testing a model of user-experience with news websites. J. Assoc. Inf. Sci. Technol. 67 (7), 1555–1575. https://doi.org/10.1002/asi.23462.
Baddeley, A., 2003. Working memory: looking back and looking forward. Nat. Rev. Neurosci. 4 (10), 829–839. https://doi.org/10.1038/nrn1201.
Bakdash, J. Z., Marusich, L. R., 2018. rmcorr: Repeated Measures Correlation. R package version 0.3.0 https://CRAN.R-project.org/package=rmcorr.
Bates, D.M., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Ben-Bassat, T., Meyer, J., Tractinsky, N., 2006. Economic and subjective measures of the perceived value of aesthetics and usability. ACM Trans. Comput.-Human Interaction (TOCHI) 13 (2), 210–234. https://doi.org/10.1145/1165734.1165737.
Bevan, N., 2009. Extending quality in use to provide a framework for usability measurement. In: Kurosu, M. (Ed.), Human Centered Design Lecture Notes in Computer Science. 5619. Springer, Berlin, pp. 13–22. https://doi.org/10.1007/978-3-642-02806-9_2.
Blackler, A., Popovic, V., Mahar, D., 2010. Investigating users' intuitive interaction with complex artefacts. Appl. Ergon. 41 (1), 72–92. https://doi.org/10.1016/j.apergo.2009.04.010.
Bland, J.M., Altman, D.G., 1994. Correlation, regression, and repeated data. BMJ: Br. Med. J. 308 (6933), 896. https://doi.org/10.1136/bmj.308.6933.896.
Bland, J.M., Altman, D.G., 1995. Statistics notes: calculating correlation coefficients with repeated observations: part 1—correlation within subjects. BMJ 310 (6977), 446. https://doi.org/10.1136/bmj.310.6977.446.
Block, R.A., George, E.J., Reed, M.A., 1980. A watched pot sometimes boils: a study of duration experience. Acta Psychologica 46 (2), 81–94. https://doi.org/10.1016/0001-6918(80)90001-3.
Block, R.A., Hancock, P.A., Zakay, D., 2010. How cognitive load affects duration judgments: ameta-analytic review. Acta Psychologica 134 (3), 330–343. https://doi.org/10.1016/j.actpsy.2010.03.006.
Block, R.A., Zakay, D., 1996. Models of Psychological Time Revisited. In: Helfrich, H. (Ed.), Time and mind. Hogrefe & Huber, Kirkland, WA, pp. 171–195.
Block, R.A., Zakay, D., 1997. Prospective and retrospective duration judgments: ameta-analytic review. Psychon. Bull. Rev. 4 (2), 184–197. https://doi.org/10.3758/BF03209393.
Bolker, B., 2019. GLMM FAQ. Webseite.
Brown, S.W., 1985. Time perception and attention: the effects of prospective versus retrospective paradigms and task demands on perceived duration. Perception Psychophys. 38 (2), 115–124. https://doi.org/10.3758/BF03198848.
Brown, S.W., 1997. Attentional resources in timing: interference effects in concurrent temporal and nontemporal working memory tasks. PerceptionPsychophys. 59 (7), 1118–1140. https://doi.org/10.3758/BF03205526.
Brown, S.W., 2008. Time and Attention: Review of the Literature. In: Grondin, S. (Ed.), Psychology of time. Emerald, Bingley, pp. 111–138.
Cohen, J., 1992. A power primer. Psychol. Bull. 112 (1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155.
Cousineau, D., 2005. Confidence intervals in within-subject designs: asimpler solution to Loftus and Masson's method. Tutor. Quant. Methods Psychol. 1 (1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042.
Dellaert, B.G., Kahn, B.E., 1998. How tolerable is delay?: consumers evaluations of internet web sites after waiting. J. Interact. Market. 13 (1), 41–54. https://doi.org/10.1002/(SICI)1520-6653(199924)13:1<41::AID-DIR4>3.0.CO;2-S.
DIN, EN ISO 9241-11, 2018. Ergonomie der mensch-System-Interaktion – Teil 11: gebrauchstauglichkeit: Begriffe und konzepte. Beuth Verlag, Berlin.
DIN EN ISO 9241-210, 2010. Ergonomie der mensch-System-Interaktion – Teil 210: prozess zur gestaltung gebrauchstauglicher interaktiver systeme. Beuth Verlag, Berlin.
Droit-Volet, S., Bigand, E., Ramos, D., Bueno, J.L.O., 2010. Time flies with music whatever its emotional valence. Acta Psychologica 135 (2), 226–232. https://doi.org/10.1016/j.actpsy.2010.07.003.
Droit-Volet, S., Monceau, S., Berthon, M., Trahanias, P., Maniadakis, M., 2018. The explicit judgment of long durations of several minutes in everyday life: conscious retrospective memory judgment and the role of affects? PLoS ONE 13 (4), e0195397. https://doi.org/10.1371/journal.pone.0195397.
Droit-Volet, S., Ramos, D., Bueno, L.J., Bigand, E., 2013. Music, emotion, and time perception: the influence of subjective emotional valence and arousal? Emotion Sci. 4, 417. https://doi.org/10.3389/fpsyg.2013.00417.
Droit-Volet, S., Trahanias, P., Maniadakis, M., 2017. Passage of time judgments in everyday life are not related to duration judgments except for long durations of several minutes. Acta psychologica 173, 116–121. https://doi.org/10.1016/j.actpsy.2016.12.010.
Droit-Volet, S., Wearden, J.H., 2015. Experience sampling methodology reveals similarities in the experience of passage of time in young and elderly adults. Acta Psychologica 156, 77–82. https://doi.org/10.1016/j.actpsy.2015.01.006.
Droit-Volet, S., Wearden, J.H., 2016. Passage of time judgments are not duration judgments: evidence from a study using experience sampling methodology. Front. Psychol. 7, 176. https://doi.org/10.3389/fpsyg.2016.00176.
Fox, J., Weisberg, S., 2011. An R Companion to Applied Regression, 2. edition. Sage, Thousand Oaks CA.
Frøkjær, E., Hertzum, M., Hornbæk, K., 2000. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, pp. 345–352.
van der Geest, T., van Dongelen, R., 2009. What is beautiful is useful-visual appeal and expected information quality. Professional Communication Conference, IPCC 2009. IEEE International, pp. 1–5.
Gibbon, J., 1977. Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. 84 (3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279.
Gibbon, J., Church, R.M., Meck, W.H., 1984. Scalar timing in memory. Ann. N. Y. Acad. Sci. 423 (1), 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x.
Gil, S., Droit-Volet, S., 2011. "Time flies in the presence of angry faces"... depending on the temporal task used!. Acta Psychologica 136 (3), 354–362. https://doi.org/10.1016/j.actpsy.2010.12.010.
Gil, S., Droit-Volet, S., 2012. Emotional time distortions: the fundamental role of arousal. Cognit. Emotion 26 (5), 847–862. https://doi.org/10.1080/02699931.2011.625401.
Gorn, G.J., Chattopadhyay, A., Sengupta, J., Tripathi, S., 2004. Waiting for the web: how screen color affects time perception. J. Market. Res. 41 (2), 215–225. https://doi.org/10.1509/jmkr.41.2.215.28668.
Hamborg, K.-C., Hülsmann, J., Kaspar, K., 2014. The interplay between usability and aesthetics: more evidence for the "what is usable is beautiful" notion. Adv. Human-Comput. Interact. 2014, 15. https://doi.org/10.1155/2014/946239.
Hassenzahl, M., 2003. The Thing and I: Understanding the Relationship between User and Product. In: Blythe, M.A., Overbeeke, K., Monk, A.F., Wright, P.C. (Eds.), Funology: From Usability to Enjoymen. Springer, Netherlands, pp. 31–42. https://doi.org/10.1007/1-4020-2967-5.
Hassenzahl, M., 2008. User experience (UX): towards an experiential perspective on product quality. Proceedings of the 20th Conference on l'Interaction Homme-Machine. ACM, pp. 11–15.
Hertzum, M., Hansen, K.D., Andersen, H.H., 2009. Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? Behav. Inf. Technol. 28 (2), 165–181. https://doi.org/10.1080/01449290701773842.
Hertzum, M., Holmegaard, K.D., 2013. Thinking aloud in the presence of interruptions and time constraints. Int. J. Hum. Comput. Interact. 29 (5), 351–364.
Hertzum, M., Holmegaard, K.D., 2015. Thinking aloud influences perceived time. Hum. Factors 57 (1), 101–109. https://doi.org/10.1177/0018720814540208.
Hudlicka, E., 2003. To feel or not to feel: the role of affect in human–computer interaction. Int. J. Hum. Comput. Stud. 59 (1–2), 1–32. https://doi.org/10.1016/S1071-5819(03)00047-8.
Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B., 2010. Measuring the dynamics of remembered experience over time. Interact. Comput. 22 (5), 328–335. https://doi.org/10.1016/j.intcom.2010.04.003.
Kliegl, R., Masson, M.E., Richter, E.M., 2010. A linear mixed model analysis of masked repetition priming. Vis. Cognit. 18 (5), 655–681. https://doi.org/10.1080/13506280902986058.
Kurusathianpong, P., Tangmanee, C., 2018. Comparison of perceived waiting time between two lengths of progress indicator and two styles of graphics animation with perceived uncertainty as a covariate. Seventh ICT International Student Project Conference (ICT-ISPC). IEEE, pp. 1–6.
Lavie, T., Tractinsky, N., 2004. Assessing dimensions of perceived visual aesthetics of web sites. Int. J. Hum. Comput. Stud. 60 (3), 269–298. https://doi.org/10.1016/j.ijhcs.2003.09.002.
Law, E.L.-C., Roto, V., Hassenzahl, M., Vermeeren, A.P., Kort, J., 2009. Understanding, scoping and defining user experience: a survey approach. Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp. 719–728.
Lee, Y., Chen, A.N., Ilie, V., 2012. Can online wait be managed? the effect of filler interfaces and presentation modes on perceived waiting time online. MIS Q. 365–394.
Lesemann, E., Woletz, N., Koerber, S., 2007. Combining methods to evaluate mobile usability. Proceedings of the 9th international conference on Human computer interaction with mobile devices and services. ACM, pp. 444–447.
Liikkanen, L.A., Gómez Gómez, P., 2013. Designing interactive systems for the experience of time. Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces. ACM, pp. 146–155. https://doi.org/10.1145/2513506.2513522.
Mahlke, S., 2008. User experience of interaction with technical systems. Technische Universität Berlin, Berlin.
McDonald, S., Zhao, T., Edwards, H.M., 2013. Dual verbal elicitation: the complementary use of concurrent and retrospective reporting within a usability test. Int. J. Hum. Comput. Interact. 29 (10), 647–660. https://doi.org/10.1080/10447318.2012.758529.
Minge, M., Riedel, L., 2013. meCUE-Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. Mensch & Computer. pp. 89–98.
Minge, M., Thüring, M., 2018. Hedonic and pragmatic halo effects at early stages of user

experience. Int. J. Hum. Comput. Stud. 109, 13–25. https://doi.org/10.1016/j.ijhcs.2017.07.007.
Minge, M., Thüring, M., Wagner, I., 2016. Developing and validating an english version of the mecue questionnaire for measuring user experience. Proceedings of the Human Factors and Ergonomics Society 2016 Annual Meeting. SAGE Publications, Los Angeles, CA, pp. 2063–2067.
Morey, R.D., 2008. Confidence intervals from normalized data: acorrection to cousineau (2005). Tutor. Quant. Methods Psychol. 4 (2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061.
Pohlmeyer, A.E., Hecht, M., Blessing, L., 2009. User Experience Lifecycle Model ContinUE [Continuous User Experience]. Der Mensch im Mittelpunkt technischer Systeme. Fortschritt-Berichte VDI Reihe. 22. VDI-Verlag, Düsseldorf, pp. 314–317.
R Core Team, 2018. R: A Language and Environment for Statistical Computing, Vienna, Austria. (R version 3.5.1).
Russell, J.A., Weiss, A., Mendelsohn, G.A., 1989. Affect grid: a single-item scale of pleasure and arousal. J. Pers. Soc. Psychol. 57 (3), 493–502. https://doi.org/10.1037/0022-3514.57.3.493.
Sackett, A.M., Meyvis, T., Nelson, L.D., Converse, B.A., Sackett, A.L., 2010. You're having fun when time flies the hedonic consequences of subjective time progression. Psychol. Sci. 21 (1), 111–117. https://doi.org/10.1177/0956797609354832.
Seow, S.C., 2008. Designing and Engineering Time - The Psychology of Time Perception in Software. Pearson Education.
Sonderegger, A., Sauer, J., 2010. The influence of design aesthetics in usability testing: effects on user performance and perceived usability. Appl. Ergon. 41 (3), 403–410. https://doi.org/10.1016/j.apergo.2009.09.002.
Sucala, M., Scheckner, B., David, D., 2011. Psychological time: interval length judgments and subjective passage of time judgments. Current Psychology Letters. Behaviour, brain & cognition 26 (2). http://journals.openedition.org/cpl/4998
Szameitat, A.J., Rummel, J., Szameitat, D.P., Sterr, A., 2009. Behavioral and emotional consequences of brief delays in human–computer interaction. Int. J. Hum. Comput. Stud. 67 (7), 561–570. https://doi.org/10.1016/j.ijhcs.2009.02.004.
Thüring, M., Mahlke, S., 2007. Usability, aesthetics and emotions in human–technology interaction. Int. J. Psychol. 42 (4), 253–264. https://doi.org/10.1080/00207590701396674.
Tractinsky, N., Katz, A.S., Ikar, D., 2000. What is beautiful is usable. Interact Comput 13 (2), 127–145. https://doi.org/10.1016/S0953-5438(00)00031-X.
Treisman, M., 1963. Temporal discrimination and the indifference interval. implications for a model of the "internal clock". Psychol. Monographs: General. 77, 1–31. https://doi.org/10.1037/h0093864.
Tullis, T., Albert, W., 2010. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, 4th edition. Springer, New York, NY. http://www.stats.ox.ac.uk/pub/MASS4
Wickens, C.D., Hollands, J.G., Banbury, S., Parasuraman, R., 2012. Engineering psychology & human performance, (4th edition). Routledge, Abingdon, Oxon.
Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wood, R.T., Griffiths, M.D., Parke, A., 2007. Experiences of time loss among videogame players: an empirical study. Cyberpsychol. Behav. 10 (1), 38–44.
Zakay, D., Block, R.A., 1997. Temporal cognition. Curr. Dir. Psychol. Sci. 6 (1), 12–16. https://doi.org/10.1111/1467-8721.ep11512604.
Ziefle, M., Bay, S., 2005. How older adults meet complexity: aging effects on the usability of different mobile phones. Behav. Inf. Technol. 24 (5), 375–389. https://doi.org/10.1080/0144929042000320009.

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 Wüzburg (Germany), Regensburg (Germany), and Freiburg (Germany). His main research interests are human timing, human-machine interaction, empirical museology, and the psychology of architecture.