

Article

Interruption Cost Evaluation by Cognitive Workload and Task Performance in Interruption Coordination Modes for Human–Computer Interaction Tasks

Byung Cheol Lee ¹, Kwanghun Chung ² and Sung-Hee Kim ^{3,*}¹ Department of Engineering, Texas A&M University–Corpus Christi, Corpus Christi, TX 78412, USA; byungcheol.lee@tamucc.edu² College of Business Administration, Hongik University, Seoul 04066, Korea; khchung@hongik.ac.kr³ Industrial ICT Engineering, Dong-eui University, Busan 47340, Korea

* Correspondence: sh.kim@deu.ac.kr; Tel.: +82-51-890-2366

Received: 23 August 2018; Accepted: 27 September 2018; Published: 30 September 2018



Abstract: Interruption is a widespread phenomenon in human–computer interaction in modern working environments. To minimize the adverse impact or to maximize possible benefits of interruptions, a reliable approach to evaluate interruption cost needs to be established. In this paper, we suggest a new approach to evaluate the interruption cost by cognitive workload and task performance measures. The cognitive workload is assessed by pupil diameter changes and National Aeronautics and Space Administration (NASA) task load index. Task performance includes task completion time and task accuracy in a series of controlled laboratory experiments. This integrated approach was applied to three interruption coordination modes (i.e., the immediate, the negotiated, and the scheduled modes), which were designed based on McFarlane’s interruption coordination modes. Each mode consists of cognitive and skill tasks depending on the degree of mental demands providing four different task sets of interruptive task environments. Our results demonstrate that the negotiated mode shows a lower interruption cost than other modes, and primary task type and task similarity between primary and peripheral tasks are crucial in the evaluation of the cost. This study suggests a new approach evaluating interruption cost by cognitive workload and task performance measures. Applying this approach to various interruptive environments, disruptiveness of interruption was evaluated considering interruption coordination modes and task types, and the outcomes can support development of strategies to reduce the detrimental effects of unexpected and unnecessary interruptions.

Keywords: human–computer interaction; interruption cost; interruption coordination mode; task type; cognitive workload; task performance

1. Introduction

Interruption is an increasingly common occurrence in human–computer interaction in modern working environments. Due to the wide adoption of information technologies in workplaces, the employees experience increasing amounts of simultaneous and dynamic task demands. Their attention to one task cannot be maintained until a prior task is complete. Such multitasking and task-switching seems to be one of the required talents for modern information workers in a typical work environment. Gonzalez and Mark [1] observed that information workers work in an average of ten different working spheres per day and switch their working spheres because of both internal and external interruptions, on average every 11.5 min. Due to such highly fragmented, reactive, and multimodal features of modern tasks, unfavorable and inopportune interruption is frequent in current working environments.

Generally, interrupting peripheral tasks unfavorably affects human behavioral performance and work productivity. It is well known that unexpected, unavoidable interruption causes stress and frustration that can deter task performance, most notably in terms of task completion times and task accuracy during and after interruptions [2,3]. Interruption increases physical and mental workloads, degrades psychological symptoms, and decreases work efficiency [4–6]. Interruptions both negatively influence the quantitative outcome of task performance and degrade qualitative performance. Moreover, they may result in frustration, anxiety, frequent errors, delays, and hesitations in a decision-making process [7,8]. These are mainly due to limited cognitive resources that are not well allocated across multiple tasks [9–12]. Adamczyk and Bailey [13] reported the adverse effects of interruptions based on inopportune occurrence timing, and unexpected interruptions can unfavorably influence task performance and lead to unwanted emotional consequences [14,15].

To develop remedies for mitigating such adverse effects, a growing body of research explored manipulating the time and location of interruption occurrence. A new visualization approach [16], opportune timing [17], and multimodal information presentation [18,19] were attempted to find an alternate tasking mode or pattern that could reduce the user's cognitive demands upon interruption. However, there is no clear strategy to avoid or manage deleterious interruptions. Ideally, recognizing the exact spot when a task was suspended and retrieving the interrupted information are the best way to eliminate the negative effects of interruptions and to resume work properly, but this approach is far from realized [13].

The main goal of this study was to develop effective interruption management strategies by suggesting a new approach to evaluate interruption cost by cognitive workload and task performance measures and to evaluate the cost of various interruptive working environments. In particular, we examined three research questions: (1) how are interruption costs assessed in different interruption coordination modes? (2) What are the contextual circumstances that characterize the interruption costs? (3) How may we design successful interruption management strategies?

This study differs from prior studies on interruption effects in two aspects. Firstly, it suggests a computational interruption cost model as the combination of task performance and cognitive workload. While previous research mostly focused on identifying impacting factors such as interruption characteristics [6], presentation format and primary task characteristics [20], and user interface design development [21], this study provides a quantitative barometer based on cognitive workload through a physiological approach and subjective self-report with task performance measures, including task completion time and task accuracy. Secondly, this study evaluates the disruptiveness of interruption in three interruption coordinating modes and combinations of different cognitive demanding tasks. Not many earlier experiments evaluated the effects of interruption while considering task types or complexity, and various interruptive tasking environments.

In Section 2, we begin with a brief review of the literature on interruption coordination modes and cognitive workload measuring with eye gaze data and unique features of interruptive tasks. In Section 3, we provide the research framework, including the elements of the interruption cost evaluation approach. In Section 4, we describe our experiment participants, measures, and procedures. In Section 5, we present the results of the experiment and analysis. In Section 6, we discuss our findings, implications for interruption management strategies to decrease detrimental effects of interruptions, and suggestions for future interruption research. Finally, we offer our concluding remarks in Section 7.

2. Related Works

2.1. Temporal Progress of Task Interruption

An interruption can be defined as “an externally generated randomly occurring, discrete event that breaks continuity of cognitive focus on a primary task” [22]. Total discrete events of the interruption process are composed of primary task performance, interruption lag, peripheral task performance, resumption lag, and further primary task performance, as shown in Figure 1. When an interruption

occurs, switching from the primary task to a peripheral task requires an additional time, referred to as interruption lag, which is needed to initiate the new peripheral task. After completing the interrupting peripheral task, even more additional time is required to return to the suspended point of tasking, which is called resumption lag. Both time lags deprive attention and memory resources in information processing of ongoing primary task completion and result in increasing the risk of various types of failure modes. Altmann and Trafton [23] asserted that the interruption lag affects the resumption lag, which they manipulated to vary in length. The interruption lag may be critical in resuming suspended primary tasks after the interruption, because, during that period, the goals to be achieved at resumption may be prospectively encoded. The encoded retrieval cues may assist the resumption of the primary task processing.

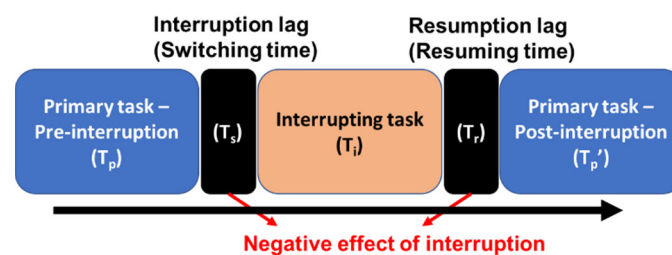


Figure 1. Time course of interruption.

Interruptions can cause decreased performance, and the demands of switching tasks and the resumption of the suspended task require additional effort [24]. This additional effort is likely to be the cause of the prolonged processing time, leading to more errors [7,23,25]. McFarlane and Latorella [21] found that most people are error prone when exposed to interruptions when they have to process cognitively demanding tasks. Additionally, interruptions force a delay in planned sequential actions that require prospective memory and working memory [26]. Without maintaining the activation level of these memories, interruptions cause the forgetting of intentions, and the greater workload caused by task switching reduces opportunities to recover mental resources [27]. Another cause of decreased performance upon interruption is the demand of attention. Generally, interruptions demand instant attention from the same sensory channel used by the primary task and generally hinder the accomplishment of task goals [28,29]. The proper allocation of attentional resources between the primary and the peripheral tasks is difficult to achieve.

2.2. Classification of Interruption Coordination Modes

How and when to deal with interruption is a major topic in interruption research and a criterion for classifying interruption types. For example, Trafton et al. [30] proposed an auditory notification interface support to find opportune moments to interrupt primary tasks and to facilitate the resumption of interrupted tasks. This interface reduces the cost of interruption by additional sensory inputs and prepared consequent activities at the moments of lower cognitive workload. Though this approach needed the consideration of the different effort levels in interruption handling, it may provide the possibility to understand the abstract level of interruption interactions. Latorella [31] suggested the interruption management stage model (IMSM), which suggests interruption management strategies on the basis of the information processing stages. IMSM acknowledges cognitive processes such as the requirement that the annunciation stimulus must be detected, and acknowledges the flexibility that the interrupting peripheral task may also be intentionally dismissed or scheduled to be completed at a later stage [21].

McFarlane introduced four modes to coordinate interruptions in human–computer interactions [32]. Among them, the immediate mode indicates that an interruption is promptly delivered regardless of the status or progress of the primary task and disrupts the recipient's primary task without warning or notification. In the negotiated mode, the user is aware of the onset of interruption by notification;

thus, the user can manipulate when to accept the interruption. The scheduled mode delivers the interruption according to a predetermined schedule. Users can expect when interruptions happen, but they cannot administer specific interventions to control them. In the mediated mode, a mediation agent is searching for an opportune moment for interrupting and suggests the timing to users, and the user decides and conducts the peripheral task processing.

Several studies examined and compared McFarlane's four interruption coordination modes, and the outcomes were inconsistent. The negotiated and the mediated modes in controlled laboratory experiments showed better efficiency and accuracy and less interruptivity [32,33]. Robertson et al. [34] claimed that the negotiated mode is better at debugging tasks than the immediate mode. However, Witt and Drugge [35] found that the immediate mode is favorable compared to the negotiated mode in their head-mounted display (HMD) experiment. However, these results were obtained without consideration of different types of tasks. Iqbal and Bailey [11] compared the immediate mode to the mediated mode, and confirmed that the effects of interruption varied by task type.

2.3. Task Features in Task Interruption

Diverse types of tasks were employed to evaluate the effects of interruptions. Different interface designs of calculators [36], number listing and book title searching tasks [37], and computer-based game tasks [6] are some example tasks used in interruption research. However, there was no specific rationale to select such tasks. Eyrolle and Cellier [7] employed more realistic tasks in an office environment. Speier, Valacich, and Vessey [20] also investigated the effects of interruptions on a decision-making process with college-level coursework in different information-presenting modes. Interruption effects based on these tasks were mainly explained by general task performance, and task performers' cognitive resource or information processing capability were not addressed much.

Different features of primary or peripheral tasks influence cognitive resources and the attention required to complete these tasks. Task features require different types of information-processing capabilities and various interacting modalities such as listening, memory, and verbalization. To monitor real-life interruption engagements, consideration of these features is imperative and may bring consistent changes in various measures of cognitive demands [38]. The tasks with clear cognitive demand properties are useful in developing a more detailed understanding of the different components of cognitive demands or information-processing capabilities in tasking environments prone to interruption.

2.4. Cognitive Workload Measurement—A Way of Interruption Cost Evaluation

The decline in task performance upon interruption is closely correlated with the increase of cognitive workload. Basically, cognitive workload can be feasibly measured by how many and how long critical pieces of information can be held during interruptive tasking environments [39]. This memory-dependent, task-related information, which is referred to as the problem state [39], was recognized as a key criterion for measuring cognitive workload: humans can retain limited amounts of problem-state information in single-task processing [40], and the task switching upon interruption incurs additional costs of increasing task completion time and the risk of error [41]. The concept of the problem-state indicates the best timing of engaging interruptions will be at moments of minimal problem state or at the lowest level of cognitive workload, which can be evaluated as the lowest interruption cost.

Numerous approaches were employed to evaluate cognitive workloads. Firstly, the subjective approach is traditionally used in various research domains. The workload is measured by directly asking how much workload individuals experience while performing a certain task. The representative example of this approach is the National Aeronautics and Space Administration (NASA) task load index (TLX) method [42]. It assesses and quantifies the degree of workload by combining perceived scores on six subscales and individual weighting of these subscales. This subjective, self-reporting measure assumes that individuals can distinguish the different amounts of workload among tasks. However, the consequent measures indicate a snapshot of recency effects at the end of a task rather than

a continuous measurement. Another limitation is that the workload estimate is based on self-reporting, which may not be reliable or accurate.

Secondly, behavioral task performance can measure cognitive workload [43,44]. This approach assumes that the external behavioral performance is a measurable outcome of an individual cognitive task and that poor performance is a good indicator of workload. The measured outcomes may be quantitative or qualitative, and indicate the individual's successful execution of the task. Though task performance and cognitive workload are related to each other, their levels are not necessarily matched. Task performance is maximized when the cognitive workload is optimal, and it can be degraded when the workload is too high or too low. The selection of right tasks and the consideration of individual difference are key aspects in this approach.

Thirdly, the physiological approach uses data from heart rate (HR) [45], heart-rate variability (HRV) [46], galvanic skin response (GSR) [47], skin temperature, and respiration rate to indirectly assess how a person uses mental resources to process information and identify the limits of cognitive capabilities. However, these data are obtained by rather obtrusive, invasive methods, such as the wearing or direct connecting of sensors, along with the risk of breaching personal health data. Among these physiological measures, pupil diameter data by eye-tracking technology is advantageous in cognitive workload measurement, because reliable data can be collected in a non-invasive way.

A main issue with using pupil diameter data to infer cognitive workload is that physiological changes in the eye are influenced by the difficulty of the engaged task [48–50]. Pupil diameter is highly responsive to changes in task difficulty, which is called the task-evoked pupillary response (TEPR) [48]. Studies confirmed this relationship between pupillary changes and cognitive workload, and this rapid change in pupil diameter enables the measurement of cognitive workload with minimal discomfort [51,52]. Recently, Gable et al. [53] asserted that pupil diameter might be a more reliable measure of cognitive workload than heart rate in driving simulator studies. Although the study had limitations due to the low number of samples, the study demonstrated that TEPR has enough feasibility to estimate cognitive workloads, and it provides consistent physiological reactions, which are independent of domain-specific contexts.

3. Research Framework

The interruption cost was examined by the interruption modes and task types with the hypotheses discussed below. The cost could be evaluated with three approaches: physiological measures (pupil diameter changes), subjective measures (self-reported cognitive workload), and behavioral measures (task performance). Considering data sources and methodologies, cognitive workloads measured by physiological and subjective approaches are taken as user-oriented interruption costs, and task performance measures, including task completion time and task accuracy, are regarded as task-oriented costs. We assumed that the interruption cost varies by both the interruption coordination modes and the task types. Table 1 shows variables and measures of interruption cost.

Table 1. Variables and measures of interruption cost. NASA TLX—National Aeronautics and Space Administration (NASA) task load index.

Interruption Cost	Variable	Measures
User-oriented	Cognitive workload	<ul style="list-style-type: none"> Physiological measure (pupil diameter variance) self-reported measure (NASA TLX)
Task-oriented	Task performance	<ul style="list-style-type: none"> Task completion time Task accuracy (the number of errors)

As described above, this study aims to identify the most effective approach to evaluate interruption cost using cognitive workload and task performance, and to apply the approach to three interruption coordination modes according to McFarlane's classification [32] with two types of tasks that require different levels of cognitive demands.

4. Experiment

4.1. Subjects

Forty participants from 20 to 28 years of age (Mean = 24.0 years, SD = 2.22 years) were recruited in the study. All participants were college students in their junior (44%) or senior years (56%); 17 participants were females, and 23 were males. All participants were familiar with general computer use and had enough competency in basic mathematical problem solving. The institutional review board protocol was reviewed and approved prior to testing to ensure the protection of human subjects in research. During data collection procedures, the pupil size data from two participants were removed due to technical issues. We also excluded data from one other participant who did not follow instructions in completing the deferred modes. Thus, our total sample size was 37.

4.2. Independent Variables: Types of Tasks and Interruption Coordination Modes

Because this study was focused on evaluating the interruption cost, selected tasks needed to be described more broadly as the individual's general activity, and needed to include distinctive cognitive processes to complete the task. A cognitive task requires a high level of cognitive demand and is sensitive to individual differences and errors. In this study, arithmetic word problems using only basic arithmetic operations were chosen. Reading comprehension skills and arithmetic operation capability were the main cognitive resources for completing these tasks. On the other hand, sentence copying tasks in which given sentences were typed with a computer keyboard were selected as the skill task. The skill task highly depends on motor skill ability and visual-spatial processing. The main performance criterion was how fast participants completed typing without typographical errors. In addition, the number of words in both tasks was controlled similarly to minimize the effects of individual differences. Table 2 shows some examples of cognitive and skill tasks used in this study.

Table 2. Sample questions for cognitive and skill tasks.

Cognitive Task: Mathematical Question Solving ¹			
A certain property doubled in value from 1950 to 1960 and tripled in value from 1960 to 1980. The value of the property in 1980 was how many times the value in 1950?			
a. 3	b. 5	c. 6	d. 8
Physical Task: Sentence Copying ²			
Type the following sentences in the given space. The specific and detailed structure is difficult to be obtained. Healthcare Information and Management System Society introduced a view about healthcare systems including the Electronic Health Record and the Electronic Medical Record.			

¹ All questions were selected from a Grade 7 math textbook [54]. Only word problems were selected. Calculators or other calculation aids were not allowed to extend cognitive process in task processing. ² Sentences were arbitrarily composed, but the number of words was limited from 40 to 45. This number was based on average typing words per minute for clerical workers [55].

As another independent variable, three interruption coordination modes based on McFarlane's four modes of interruption coordination [32] were designed and implemented. Each mode indicates different interfaces to interact with interruptions. In the immediate mode, an interrupting task appeared while a primary task was performed without any notice. An interruption was immediately delivered regardless of the status and progress of primary tasks. Participants could not control the occurrence of interruptions. In the negotiated mode, participants were notified that the interruption

would occur, and they could control the onset of an interruption by clicking a button on the screen. In the scheduled mode, participants acknowledged the onset of an interrupting task after a primary task was given. Though participants could expect when an interruption occurred, they could not manipulate the given schedule. In the experiment, once the primary task was given, the peripheral task popped up after 15 s. A countdown timer informed the participant of the remaining time before the task had to be completed.

The other mode, “mediated mode”, was not included in this study. In a mediated mode, the agent crucially decides when and how the interruption is initiated, and this controversial process requires interruptibility prediction, intelligent interfaces, precise cognitive workload measurement, human factors techniques, and a cognitive model to supervise the mediator [32]. Thus, the impact of the mediated mode on an actual user’s cognitive workload would be minimal, and this mode has a quite different mechanism for the interruption interaction from that of other modes.

4.3. Apparatus

The experiment was conducted in a small windowless room, and participants were tested individually. They were seated approximately 50 cm from a 20-inch liquid crystal display (LCD) monitor with a screen resolution of 1600×1200 pixels and a screen density of 64 pixels/inch. An SMI RED 250 remote eye-tracker was used to collect the eye-tracking data which was recorded at 120 Hz. Calibration was performed before the experiment started. A calibration accuracy of 0.8° was considered acceptable. The eye-tracker’s default parameters were used to convert gaze positions into fixations and saccades. The pupil diameters for both eyes were collected. All data were stored and analyzed using the SMI internal software using default parameters for the event detection.

4.4. Experiment Design and Dependent Variables

A within-subject experiment design was applied in a series of controlled laboratory experiments. Participants completed three interruption modes (immediate, mediated, and scheduled) with four different primary interrupting peripheral task sets: cognitive/cognitive, cognitive/skill, skill/cognitive, and skill/skill task sets. Both tasks used a visual modality presentation to minimize the effects of extraneous visual stimuli that may have influenced eye-tracking measurements [56]. Because a visual modality is sensitive to environmental conditions and other concurrent modality inputs, the experiment was conducted in a carefully controlled environment. Visual stimulus is only an input channel to collect pupil diameter data without the interference of other sensory inputs. The satisfactory counterbalance for the order of tasks among participants was designed. Figures 2 and 3 shows the experiment framework for this study.



Figure 2. Experiment setting.

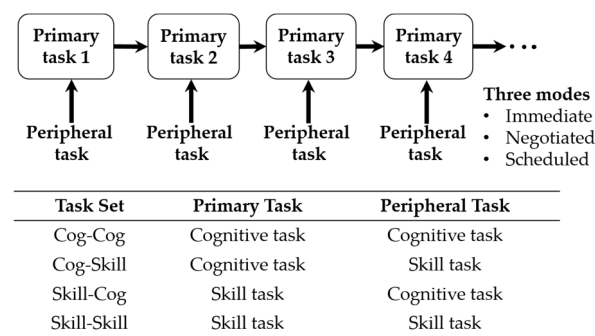


Figure 3. Experiment design.

For dependent variables, we recorded pupil diameters from an eye-tracker, task performance, and self-reported workload rates. Task performance measures were divided into a quantitative measure and a qualitative measure. As a quantitative measure, task completion time was recorded, and task accuracy (wrong answers for cognitive tasks and typographical errors for skill tasks) was counted as a qualitative measure. Task completion time included time to complete a primary task and an interrupting peripheral task, resumption lag time, and interruption lag time. Task accuracy measures reflected each task type's features. For example, the choice of "I do not know" was given to participants to minimize the bias of the right answer rate in cognitive tasks. The errors in skill tasks included both spelling errors and punctuation and capitalization errors. In addition, subjective workload ratings were measured using the simplified NASA TLX questionnaire.

4.5. Procedure

Each participant was asked to read and sign the consent form, and filled out a demographic questionnaire before the experiment; a ten-minute training session with several sample tasks was provided in order to fully familiarize with the tasks in different interruption coordination modes.

Before the experiment started, a nine-point calibration method that was provided with the eye-tracking software was performed as a calibration process of the eye-tracker. Participants then performed one set of reference task scenarios and twelve different task scenarios that were assigned in a predetermined order. The reference task scenario consisted of five cognitive tasks and five skill tasks without any interruptions, and task performance data were used as a reference for task performance ratio calculations. Each scenario had five interruptions and each participant conducted 12 scenarios. As explained above, a primary task appeared on the monitor screen. The interruption tasks were given by three interruption modes, popping up on a new screen which totally occluded suspended primary task. Once the subject completed the interruption task, he/she clicked an "OK" button and automatically returned to the suspended primary task. In the immediate mode, interruption tasks emerged without any precursors. In the negotiated mode, when the subject clicked the button that appeared on the screen, interruption tasks popped up. The button appeared in the right bottom corner of the screen and the countdown clock that went from 10 to zero while flashing from three onward was located in the top right corner of the screen. In the scheduled mode, participants had interruption tasks after fifteen seconds with the countdown clock and clicking sounds. Each scenario continued for 5 min, and the order of task sets was fully counterbalanced. Each participant also experienced a non-interruption scenario consisting of pairs of cognitive and skill tasks, and the results were used as a performance reference.

Participants were given one minute of resting time between the end of one trial and the start of the next trial. Previous studies indicated that just a few seconds were needed for pupil diameter after a cognitively demanding task to return to its pre-task level [57]. One minute of recovery time was enough to recover a baseline state of workload before the next trial.

At the end of the scenario, the participants were asked to fill out the simplified version of NASA TLX questionnaires to measure their subjective workload using a mobile application in a

tablet computer. In this study, a simplified unweighted version of the scale (raw TLX, RTLX) was used, and physical demand, one of the dimensions, was excluded. Each subjective workload rating represents the participants' average workload from 12 trials of that combination of task sets and interruption modes. Instead of weighted average estimation, overall workload scores were calculated by averaging participants' ratings on a scale of 1–20 over the five dimensions of the NASA TLX questionnaire [29].

5. Data Collection and Analysis

5.1. Physiological Cognitive Workload Using Pupil Diameter

Given that the participants experienced four primary interruption task sets with each of the three interruption modes, a total of 480 pupil diameters (40 participants with 12 task scenarios) averaged for both eyes in each task scenario were obtained. To achieve a uniform sampling rate of 120 Hz, a one-dimensional data interpolation method was employed for the raw measurements. From the collected data, we excluded 11 data that were noisy or failed to detect pupil diameter accurately, and the total number of calculated averages was 469 cases. Though the loss of data did not distort our counterbalanced design, a single imputation method was used to fill in the missing data [58]. Means and standard deviations for pupil diameters for all participants can be found in Table 3 and Figure 4.

We conducted a 4 (task set) \times 3 (interruption mode) repeated-measures ANOVA to examine the separate and combined influences of these factors on pupil diameter. Mauchly's test checked if the sphericity assumption was violated in our pupil data for the effect of task set (χ^2 (5) = 2.412, p = 0.061) and the effect of interruption mode (χ^2 (2) = 10.162, p < 0.001), and the interaction between task set and interruption mode (χ^2 (27) = 47.262, p < 0.001). Therefore, Greenhouse-Geisser corrections are reported here. We observed a significant main effect of task set on pupil diameter (F (3.921, 48.421) = 47.143, p < 0.05). Post hoc least significant difference (LSD) comparisons revealed significant differences in pupil diameter among task sets, except for the comparisons between the skill/cognitive set and skill/skill set (p > 0.05). Participants' pupil diameters were more dilated in cognitive primary task sets (cognitive/cognitive and cognitive/skill sets) than skill primary task sets (skill/cognitive and skill/skill sets) in all three interruption modes. However, there was no difference in the cognitive/cognitive and cognitive/skill sets in the negotiated mode. This omnibus test also revealed a significant main effect of interruption mode on pupil diameter (F (1.368, 25.990) = 71.903, p < 0.05). Post hoc LSD comparisons showed that pupil diameters decreased with the order of the immediate, the scheduled, and the negotiated modes. Another result observed within this omnibus test was that there was no significant interaction between interruption mode and task set on pupil diameter (F (4.356, 82.767) = 3.633, p > 0.05).

Table 3. Descriptive statistics of pupil diameters under each task set.

Pupil Diameter		Mean	Standard Deviation
(Average Value for Both Eyes)		(mm)	(mm)
Cognitive/cognitive task set	Immediate	4.553	0.573
	Scheduled	4.199	0.474
	Negotiated	3.520	0.529
Cognitive/skill task set	Immediate	4.035	0.538
	Scheduled	3.967	0.443
	Negotiated	3.479	0.394
Skill/cognitive task set	Immediate	3.700	0.388
	Scheduled	3.572	0.327
	Negotiated	3.001	0.327
Skill/skill task set	Immediate	3.646	0.362
	Scheduled	3.462	0.273
	Negotiated	3.013	0.255

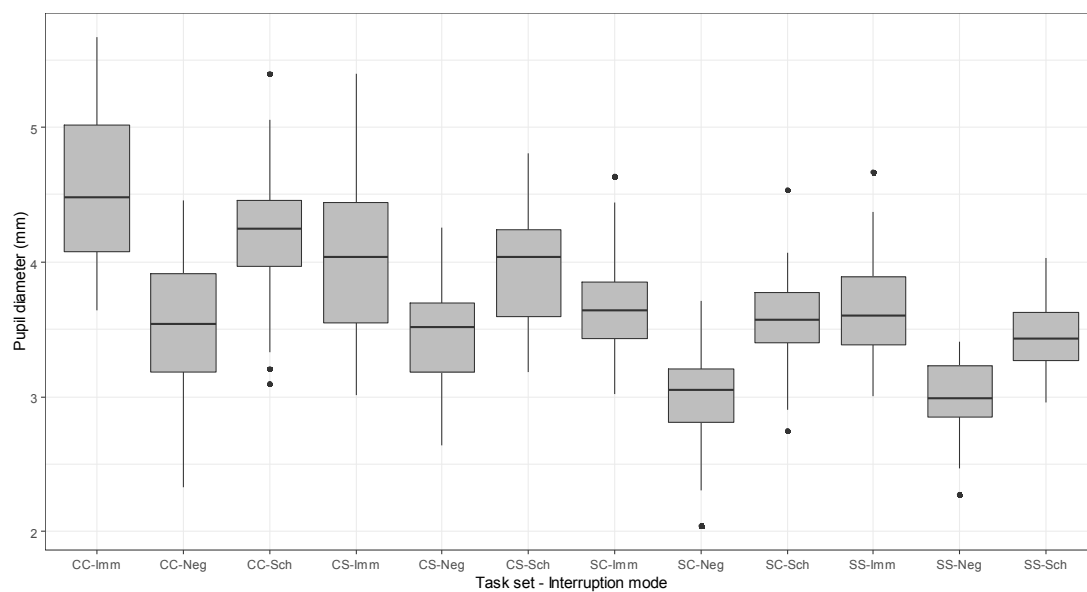


Figure 4. Pupil diameter variation by task set and interruption coordination mode.

5.2. Subjective Cognitive Workload Using NASA TLX

A two-way repeated-measures ANOVA was performed for participants' responses to the NASA TLX questionnaire. Mauchly's test confirmed the sphericity assumption in the subjective workload measures from NASA TLX for the effect of interruption mode ($\chi^2 (5) = 7.891, p > 0.05$) and the effect of task set ($\chi^2 (2) = 5.743, p > 0.05$), and the interaction between interruption mode and task set ($\chi^2 (27) = 35.621, p > 0.05$). Therefore, no corrections to degrees of freedom were needed. It also showed a significant effect of interruption modes ($F (2, 235) = 4.71, p < 0.01$) and task sets ($F (3, 235) = 5.29, p < 0.01$) on task completion time. No interaction effect was observed. Post hoc LSD pairwise comparisons showed that there was a difference between task sets and their subjective ratings of workload in the immediate mode, but those in the negotiated mode were not statistically differentiated. In other words, as participants delicately perceived the demands in task sets in the immediate mode, they felt that the task was less workload-inducing in the negotiated mode. The analysis results indicated that there was a significant statistical difference between interruption modes and overall subjective workload. Participants felt the highest workload in the immediate mode and the lowest in the negotiated mode. Descriptive statics of NASA TLX can be found in Table 4 and Figure 5.

Table 4. Descriptive statistics of NASA TLX data under each task set.

NASA TLX Measure		Mean	Standard Deviation
Cognitive/cognitive task set	Immediate	67	9.62
	Scheduled	35	6.94
	Negotiated	29	4.59
Cognitive/skill task set	Immediate	55	9.92
	Scheduled	26	5.56
	Negotiated	22	3.41
Skill/cognitive task set	Immediate	42	8.26
	Scheduled	26	5.85
	Negotiated	19	2.71
Skill/skill task set	Immediate	34	7.13
	Scheduled	22	3.95
	Negotiated	17	2.63

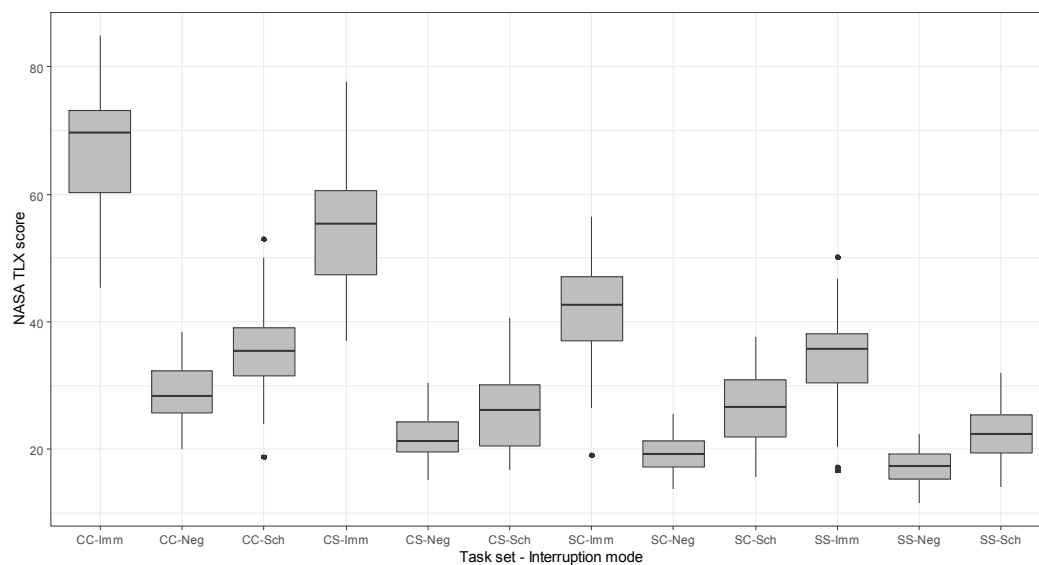


Figure 5. NASA TLX data by task set and interruption coordination mode.

5.3. Behavioral Task Performance Using Task Completion Time and Task Accuracy

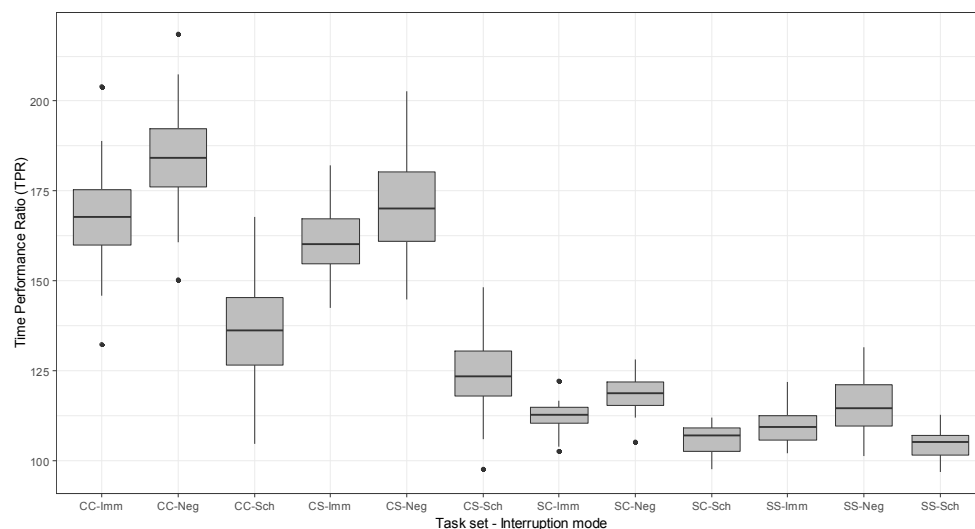
Task performance was assessed using two measures: task completion time and task accuracy. Task completion time was measured using a standardized ratio between task completion time with interruptions and without interruptions as a quantitative performance measure [59]. While task completion time was measured regardless of task type, task accuracy was measured with different outcomes for each type of task. In cognitive tasks, task accuracy could be assessed based on the answers per task, and, in skill tasks, the number of typographical errors represented the task accuracy. Thus, task accuracy was evaluated based on the percentage of false responses in cognitive tasks and typographical errors in skill tasks, and it was standardized using measures with interruptions and without interruptions [59].

Similar to cognitive workload analysis, a two-way repeated-measures ANOVA was conducted for task completion time measures. Mauchly's test checked if the sphericity assumption was violated in task completion time measures (time performance ratio, TPR) for the effect of interruption mode ($\chi^2(5) = 9.281, p > 0.05$) and the effect of task set ($\chi^2(2) = 5.124, p > 0.05$), and the interaction between interruption mode and task set ($\chi^2(27) = 37.262, p > 0.05$). Therefore, no corrections to degrees of freedom were needed. It also showed a significant effect of interruption modes ($F(2, 1225) = 4.71, p < 0.01$) and task sets ($F(3, 1224) = 5.29, p < 0.01$) on task completion time. No interaction effect was observed. Post hoc LSD comparisons showed that the TPR of cognitive primary task sets (cognitive/cognitive and cognitive/skill sets) was significantly higher than skill primary task sets (skill/cognitive and skill/skill sets) in the immediate mode. Similar patterns were shown in the scheduled mode; however, in the negotiated mode, the results of both task sets were not significantly different. Regarding the effect of interruption mode, the scheduled mode (mean = 116.28, SD = 12.01) took the shortest task completion time, and the negotiated mode took the longest time (mean = 168.84, SD = 9.97). This suggests our participants spent more time completing task sets in the immediate mode, and they spent less time in the scheduled and mediated modes. For task sets, skill primary task sets took a similar length of task completion time. Descriptive statics of task completion time can be found in Table 5 and Figure 6.

Table 5. Descriptive statistics of time performance ratio (TPR) under each task set.

Time Performance Ratio (TPR) ¹		Mean	Standard Deviation
(Primary/Peripheral Task)			
Cognitive/cognitive task set	Immediate	166.81	13.08
	Scheduled	135.90	13.61
	Negotiated	184.41	13.63
Cognitive/skill task set	Immediate	161.28	10.12
	Scheduled	124.81	10.28
	Negotiated	171.71	14.47
Skill/cognitive task set	Immediate	112.32	3.99
	Scheduled	106.18	3.98
	Negotiated	118.71	4.58
Skill/skill task set	Immediate	109.26	4.88
	Scheduled	104.67	4.16
	Negotiated	114.77	7.15

$$^1 \text{ Time performance ratio (TPR)} = \frac{\text{Task completion time with interruptions}}{\text{Task completion time without interruptions}}.$$

**Figure 6.** Task completion time by task set and interruption coordination mode.

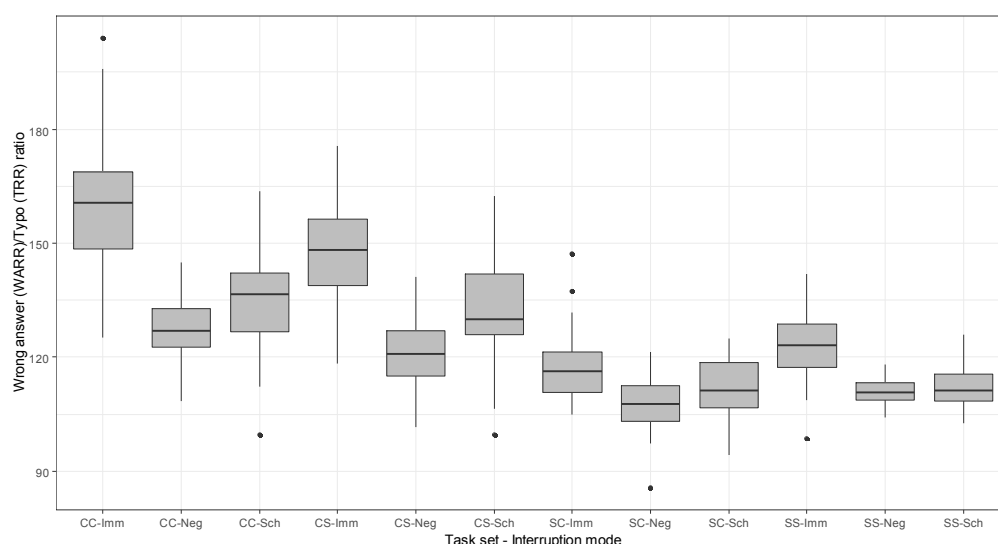
Interruption mode and task set significantly affected task accuracy. Similar to task completion time, the results of a two-way repeated-measures ANOVA revealed statistically significant differences in each subject's task accuracy among task sets, ($F(3, 274) = 27.300, p < 0.05$) and interruption modes ($F(2, 275) = 7.85, p < 0.001$). There was no interaction effect between interruption mode and task set. Mauchly's test showed that the sphericity assumption was well maintained in task accuracy (wrong answer rate ratio, WARR; and typo rate ratio, TRR) for the effect of interruption mode ($\chi^2(5) = 8.985, p > 0.05$) and the effect of task set ($\chi^2(2) = 5.743, p > 0.05$), and the interaction between interruption mode and task set ($\chi^2(27) = 31.641, p > 0.05$). Therefore, no corrections to degrees of freedom were needed. Descriptive statistics of task accuracy can be found in Table 6 and Figure 7.

According to post hoc LSD paired comparisons, task accuracy showed different patterns by task set and interruption mode. The task accuracy of skill primary task sets was lower than that of cognitive primary task sets in the immediate and the scheduled modes, but there was no difference between the skill/skill task set and skill/cognitive task set in the scheduled mode. In addition, in the negotiated mode, there was no statistical difference of task accuracy in all four task sets. Regarding the effect of interruption mode, the negotiated mode (mean = 116.85, SD = 9.57) had the lowest error rate, and the immediate mode had the highest rate (mean = 137.86, SD = 13.36).

Table 6. Descriptive statistics of typo rate ratio and wrong answer rate ratio under each task set.

Wrong Answer Rate Ratio (WARR) ¹		Mean	Standard Deviation
Typo Rate Ratio (TRR) ²			
(Primary/Peripheral Task)			
Cognitive/cognitive task set	Immediate	159.50	18.62
	Scheduled	134.84	12.30
	Negotiated	127.38	8.63
Cognitive/skill task set	Immediate	148.34	13.52
	Scheduled	131.91	15.00
	Negotiated	121.16	8.63
Skill/cognitive task Set	Immediate	116.88	8.81
	Scheduled	111.66	7.72
	Negotiated	107.49	7.07
Skill/skill task set	Immediate	122.77	8.53
	Scheduled	112.23	5.66
	Negotiated	110.80	3.45

¹ Wrong Answer Rate Ratio (WARR) = $\frac{\text{Wrong Answer Rate with Interruptions}}{\text{Wrong Answer Rate without Interruption}}$; ² Typo Rate Ratio (TRR) = $\frac{\text{Typo Rate with Interruptions}}{\text{Typo Rate without Interruption}}$.

**Figure 7.** Task accuracy by task set and interruption coordination mode.

In task performance measures, interruptions more negatively affected task completion time on the negotiated stage. Figure 6 shows that task completion time in the negotiated mode was higher than that in the immediate and the scheduled modes ($F(2, 331) = 7.11, p < 0.001$). In the negotiated mode, participants had not much perceived time pressure for completing the task, and they could manage the onset of interrupting peripheral tasks with control. This mode can be deemed as a consecutive process, not a parallel process. Resumption lags between primary and peripheral tasks in the negotiated mode are longer than those in other modes. Due to the effects of task type, shown in Figure 7, cognitive primary task sets took more task completion time than skill primary task sets ($F(1, 331) = 11.88, p < 0.001$). Since cognitive tasks need more mental demands for task completion, more interruption lags to initiate peripheral tasks and resumption lags to continue interrupted primary tasks are required. Such long switching times make it difficult to retrieve the cues of the interrupted timing and place, and impede the transition to the new task [23]. Accordingly, a memory load is the crucial aspect of the negative interruption effects on quantitative performance. Since cognitive tasks require more memory load than sensory information-based skill tasks, the degradation of task performance in

cognitive tasks is more severe. However, in the negotiated mode, there was no significant difference in task performance.

Interruption effects showed different results from task accuracy. The negotiated stage generated the fewest errors, and the immediate stage resulted in the most frequent errors in both primary task scenarios. Unlike the task completion time, task accuracy was sensitive to time pressure. In the negotiated stage, participants did not realize time pressure much in handling interruptions, and task accuracy in the negotiated mode was better than any other modes. This inference can be justified by the similar task accuracy performance in the four different task sets. Among the task sets, comparable task types (skill/skill and cognitive/cognitive task sets) resulted in more errors than incomparable task types (skill/cognitive and cognitive/skill task sets). One probable cause of these low task accuracy rates in these task sets would be task similarity. The same task types between primary and peripheral tasks need higher cognitive demand for switching the tasks and confuse the cue–goal connection [23,59]. In addition, well-trained consistent motor behaviors are more susceptible to interruptions, and retrieving and resuming interrupted cues require more effort [23,60].

6. Discussion

Subjective and physiological measures of cognitive workload and task performance demonstrated how interruptions detrimentally affect information or task processing. Three interruption coordination modes and two types of task provided specific tasking conditions to explore human and task interactions of interruption management.

6.1. Evaluation of Interruption Cost by Interruption Coordination Mode and Task Type

Cognitive workload data measured by pupil diameters and the NASA TLX method, as well as task performance data including task completion time and task accuracy. Indicated that the interruption costs varied with interruption mode and task set.

Two trends are visible in the data analysis. Firstly, the average interruption costs clearly vary with interruption mode. In user-oriented costs, both physiological and subjective workload measures indicated that the immediate mode showed the highest cost and the negotiated mode demonstrated the lowest cost. Interestingly, while the scheduled mode and the negotiated mode were significantly different in pupil diameter variation, they were not distinctive in subjective cognitive workload measures (NASA TLX scores). This may be due to limitations in the self-reported survey-based approach. Due to the feature of after-event data collection, participants hardly recalled and recognized workload variations in their perception of entire task experiences.

Task-oriented costs showed different patterns. The negotiated mode scored the lowest error rates, but the scheduled mode recorded shorter task completion times than the immediate mode. Because task accuracy counts the errors in primary tasks, participants in the negotiated mode could maintain full attention during the task, which resulted in the fewest number of errors in both cognitive and skill tasks. This may be due to the participants' ability to control the initiation of interruptions in the negotiated mode. This pattern is well matched with the lowest cognitive workload by the physiological measure. For task completion time, the time difference in the scheduled and the immediate modes can be explained by the roles of interruption lag [30]. Participants were fully aware of the onset timing of interrupting peripheral tasks in the scheduled mode, whereas participants were not ready to transfer to peripheral tasks when an interruption occurred in the immediate mode. This resulted in shorter interruption lag and total task completion time in the scheduled mode than in the immediate mode. Because the interruption coordination modes can be interpreted as guides for how to deliver interruptions, interruption lag depends more on the mode than the resumption lag required to resume suspended primary tasks. In the negotiated mode, participants did not experience any time pressure on interruption and recorded the longest time.

Secondly, the interruption costs also highly depended on the different type of primary task. Generally, compared to skill/skill sets, cognitive/cognitive sets showed the highest workload and the

longest task completion time. The cost by task type depended on the performance of the primary task. Cognitive primary task sets (cognitive/cognitive and cognitive/skill tasksets) had higher workload and worse task performance than skill primary task sets (skill/cognitive and skill/skill task sets). However, the cost by task type was affected by interruption coordination modes as well. In particular, in the negotiated mode, skill/skill task sets demonstrated the highest task accuracy, and similar levels of subjective workload were measured across all task sets. For task performance, as Lee and Duffy [59] reported, cognitive task sets (cognitive/cognitive and cognitive/skill task sets) need more time for task completion, and skill task sets (skill/cognitive and skill/skill task sets) generate more errors.

The memory for goals theory supports the explanation of the resumption process, where the most active goal in memory is pursued [23,30]. The goals of interrupted (suspended) tasks decay over time and lose activation. The base-level activation (task history) and associated problem state (current context) enable overcoming decay over time and keep relevant goals to be resumed. Generally, longer resumption lags make a goal less active, and therefore, more difficult to be resumed [61,62]. Base-level activation can be increased by task rehearsals, and associated problem state activation is caused by the priming of mental or physical cues [30]. However, depending on whether the goals of the interrupted tasks are stored in declarative or procedural memories, the cues will be more easily retrieved [41]. This may explain the effects of different cognitive demanding tasks in this study; the interrupting peripheral task will more disrupt the primary task if it is sufficiently complex.

6.2. Cognitive Workload Measurement—Interruption Management Strategies

The results analyzed in previous sections indicate that when peripheral tasks interrupt the ongoing processing of primary tasks, users experience more irritation and anxiety, need more time to complete the tasks, and make more mistakes across tasks. These adverse effects are worse when the types of tasks are dissimilar between the primary and peripheral tasks. Speier et al. [63] investigated the frequency effect of interruptions and content relevancy between primary and peripheral tasks. Then, they found that the negative impact of interruptions on task performance was more severe when the contents of both tasks were dissimilar. As more diverse notifications and interventions are implemented in modern workplaces, unavoidable immediate interruptions increasingly occur, and the collective effect of these interruptions can be quite notable.

The data showed that interruption costs are taken as a negative implication that increase workload and lower work performance; such an estimate of negative effects would be useful in modern working environments. Considering unavoidable and widespread features of interruptions, the evaluation of interruption cost might allow users to determine how interrupting peripheral tasks should be handled. The different ways of interruption interaction may help identify the optimal way of presenting the information or sequencing peripheral tasks based on the user's workload levels. Because the interruption coordination modes considered in this study manipulated timings and types of interruption onset, our results can lead to the development of temporal strategies for reducing the interruption cost. By deferring delivery of imminent peripheral tasks to lower mental demanding moments during task execution, the resulting outcomes of the interruption would be significantly less disruptive. As shown by our results, the control of interruption delivery could lead to a large alleviation of disruptive effects.

This study suggested three approaches to measure interruption costs in terms of workload assessment. Firstly, the performance-based measure assumes that successful behavioral outcomes mirror workload and that poor performance is a good indicator of workload. Performance measures with task completion time and task accuracy capture how well the user is performing a given task. Secondly, the self-reporting subjective perception measure using the NASA TLX questionnaire also assumes that an individual can correctly identify and report changes in workload. However, both approaches have limitations. Workload measured by behavioral and self-reported outcomes is a snapshot rather than a continuous measurement. Because the self-reported measure is conducted after the task is performed, the measurement is mainly reflected by most recent effects. In addition,

the performance-based measure is mainly limited to a laboratory environment; thus, the results would be difficult to demonstrate in a real-world experience.

To complement the limitations of subjective and task performance measures, a physiological approach based on pupil diameter data is suggested. With advances in sensor technology, physiological metrics can produce continuous estimates in real time and overcome the limitations inherent in subjective measurement or behavioral performance. In particular, as shown in this study, the pupillary dilations accompanying cognitive processes are taken as a representative function of mental effort in performing the cognitive task. However, physiological measures depend on many factors, including other aspects of the user's cognitive state, the user's physical activity, and environmental variables such as temperature and lighting [64]. While subjective measures are relatively simple to oversee, they cannot account for rapid changes in cognitive load that may be the result of changes in experimental conditions. In this work, we evaluated the interruption cost with the prospect of workload measures based on pupil diameter measurements, task performance measures with task completion time and task accuracy, and a subjective measure of NASA TLX.

7. Limitations

One limitation of our work is that our experimental design focused on discrete events of interruption modes and distinctive task types. Though this is inevitable and intentional to clarify the collected data, natural and mixed transitions between these discrete levels or task types are required to reflect actual tasking environments. In addition, though interruptions may also be incurred internally [65], we considered only external interruptions. For future work, we plan on introducing more realistic interruption processes and more diverse task examples. However, we see this model as a first step toward the comprehensive estimation of interruption cost through pupil diameter changes, task completion time and accuracy, and a subjective workload measure.

The lack of ecological validity in this study's tasks is another limitation. The tasks used in this study were designed based on contrived and laboratory-oriented tasks that are far from actual, authentic tasks in a real setting, and the tasks used lack generalizability and fidelity. Furthermore, suspended primary tasks were occluded by interrupting tasks in all interruption modes in this study. This also results in an ecological validity issue because total occlusions are not frequent in actual task settings. Such differences in task type affect not only experiment and study design, but also the disreputability of interruption. Arbitrary, contrived tasks can manipulate design factors such as task difficulty and interruption moments and durations, whereas actual tasks in natural settings are easily supported by various post-interruption tools and artifacts, which guide one back into the task in other ways. Several interruption studies were carried out with actual authentic tasks including the comparison of interruption modes in a manufacturing process [66] and the investigation of resumption lags in interrupted critical care tasks in the healthcare domain [67]; however, more interruption studies are still needed to understand how to minimize the negative effects of interruptions.

Another limitation is that the study tested only three out of four interruption coordination modes, as defined by McFarlane. This study ruled out the mediated mode because we mainly focused on direct interaction between users and interruption, where the mediated mode depends greatly on the agent function to determine when and how interruptions interfere. However, four interruption coordination modes reflect well the interruption patterns in user interface design guidelines [21], and we need to include the mediated mode in a more general application for the human-computer interaction (HCI) domain.

Individual differences and narrow participant distribution are other considerations in the estimation of interruption cost. Specifically, individual differences result in a widely varying degree of problem state and focus of attention, which are critical in information processing [68]. Furthermore, the interruption cost depends on the duration, the moments, and the complexity of interruptions. In addition, the experiments were limited to college students who major in engineering. A broader range of participants needs to be recruited in future studies. Finally, a number of environmental

variables, such as ambient lighting, as well as task modality and difficulty, will change continuously in time and value.

8. Conclusions

This study aimed to evaluate the interruption cost by interruption coordination modes and task types using user- and task-oriented measures. Many prior studies argued that interruptions may be disruptive and detrimental in complex multitasking environments. On the other hand, interruptions may be beneficial in initiating the reception of new information and alerts. Regardless of positive or negative, the reliable evaluation of interruption cost is valuable in interruption-ridden modern working environments. In this study, cognitive workload and task performance were taken as user-oriented and task-oriented cost components in the interruption cost evaluation approach. As specific measures of the cost components, pupil diameter changes and the NASA TLX questionnaire were used to assess cognitive workload, and task completion time and task accuracy were collected as task performance measures. As an application of the approach, three interruption coordination modes, namely the immediate, the negotiated, and the scheduled modes, were designed and tested, and task type depending on cognitive demands was considered as another factor for interruptive task environments. The results highlighted that the interruption cost in the negotiated mode is lower than that in other modes, and that primary task type is more critical in the evaluation of the cost. The results also validated the effects of task similarity of primary and peripheral tasks in interruptions. Moreover, the controllability of interruptions in the immediate mode and the interruption lags in scheduled mode were evaluated as the most cost-demanding processes. This study made several contributions toward understanding how the variations of workload and task performance with regards to the structure of interruptive task types and different interruption management modes are related. In the end, dissimilar types of tasks and optimal lag-time control can be suggested as potential strategies to reduce detrimental effects of unexpected and unnecessary interruptions.

Author Contributions: All authors conceived the research questions and the structure of this study. B.C. designed and conducted the experiment and data collection. All authors wrote the paper; B.C. wrote an original draft, and K.H. and S.H. revised the paper. S.H. secured funding to conduct the experiment and publish the paper.

Funding: This research was funded by the National Research Foundation of Korea, grant number NRF-2017R1C1B5076926.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gonzalez, V.M.; Mark, G. “Constant, Constant, Multi-tasking Craziness”: Managing Multiple Working Spheres. In Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, Vienna, Austria, 24–29 April 2004; pp. 113–120.
2. Carayon, P. Stressful jobs and non-stressful jobs: A cluster analysis of office jobs. *Ergonomics* **1994**, *37*, 311–323. [[CrossRef](#)] [[PubMed](#)]
3. Yeh, Y.Y.; Wickens, C.D. *An Investigation of the Dissociation between Subjective Measures of Mental Workload and Performance*; University of Illinois: Urbana-Champaign, IL, USA, 1984.
4. Carayon, P.; Wetterneck, T.B.; Hundt, A.S.; Ozkaynak, M.; DeSilvey, J.; Ludwig, B.; Ram, P.; Rough, S.S. Evaluation of nurse interaction with bar code medication administration technology in the work environment. *J. Patient Saf.* **2007**, *3*, 34. [[CrossRef](#)]
5. Cutrell, E.B.; Czerwinski, M.; Horvitz, E. Effects of instant messaging interruptions on computing tasks. In *CHI’00 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2000; pp. 99–100.
6. Gillie, T.; Broadbent, D. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychol. Res.* **1989**, *50*, 243–250. [[CrossRef](#)]
7. Eyrolle, H.; Cellier, J.-M. The effects of interruptions in work activity: Field and laboratory results. *Appl. Ergon.* **2000**, *31*, 537–543. [[CrossRef](#)]

8. Horvitz, E.C.M.C. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In Proceedings of the INTERACT'01: IFIP TC 13 International Conference on Human-Computer Interaction, Tokyo, Japan, 9–13 July 2001; p. 263.
9. Alm, H.; Nilsson, L. The effects of a mobile telephone task on driver behaviour in a car following situation. *Accid. Anal. Prev.* **1995**, *27*, 707–715. [\[CrossRef\]](#)
10. Anderson, J.R.; Bothell, D.; Byrne, M.D.; Douglass, S.; Lebiere, C.; Qin, Y. An integrated theory of the mind. *Psychol. Rev.* **2004**, *111*, 1036. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Bailey, B.P.; Iqbal, S.T. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2008**, *14*, 21. [\[CrossRef\]](#)
12. Card, S.K.; Moran, T.P.; Newell, A. *The Psychology of Human-Computer Interaction*; CRC: Boca Raton, FL, USA, 1983.
13. Adamczyk, P.D.; Bailey, B.P. If not now, when? The effects of interruption at different moments within task execution. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24–29 April 2004; pp. 271–278.
14. Bailey, B.P.; Konstan, J.A.; Carlis, J.V. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Interact*; University of Minnesota: Minneapolis, MN, USA, 2001; Volume 1, pp. 593–601.
15. Zijlstra, F.R.; Roe, R.A.; Leonora, A.B.; Krediet, I. Temporal factors in mental work: Effects of interrupted activities. *J. Occupat. Organ. Psychol.* **1999**, *72*, 163–185. [\[CrossRef\]](#)
16. Van Dantzich, M.; Robbins, D.; Horvitz, E.; Czerwinski, M. Scope: Providing awareness of multiple notifications at a glance. In Proceedings of the Working Conference on Advanced Visual Interfaces, Trento, Italy, 22–24 May 2002; pp. 267–281.
17. Monk, C.A.; Boehm-Davis, D.A.; Trafton, J.G. The attentional costs of interrupting task performance at various stages. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting; SAGE Publications Sage CA: Los Angeles, CA, USA, 2002; Volume 46, pp. 1824–1828.
18. Arroyo, E.; Selker, T.; Stouffs, A. Interruptions as multimodal outputs: Which are the less disruptive? In Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces, Pittsburgh, PA, USA, 16 October 2002; pp. 479–482.
19. Brewster, S.A.; Wright, P.C.; Edwards, A.D. The design and evaluation of an auditory-enhanced scrollbar. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 24–28 April 1994; pp. 173–179.
20. Speier, C.; Vessey, I.; Valacich, J.S. The Effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision-Making Performance. *Decis. Sci.* **2003**, *34*, 771–797. [\[CrossRef\]](#)
21. McFarlane, D.C.; Latorella, K.A. The scope and importance of human interruption in Human-Computer Interaction design. *Hum.-Comput. Interact.* **2002**, *17*, 1–61. [\[CrossRef\]](#)
22. Coraggio, L. *Deleterious Effects of Intermittent Interruptions on the task Performance of Knowledge Workers: A Laboratory Investigation*; The University of Arizona: Tucson, AZ, USA, 1990.
23. Altmann, E.M.; Trafton, J.G. Memory for goals: An activation-based model. *Cognit. Sci.* **2002**, *26*, 39–83. [\[CrossRef\]](#)
24. Brixey, J.J.; Robinson, D.J.; Johnson, C.W.; Johnson, T.R.; Turley, J.P.; Zhang, J. A concept analysis of the phenomenon interruption. *Adv. Nurs. Sci.* **2007**, *30*, E26. [\[CrossRef\]](#)
25. Bailey, B.P.; Konstan, J.A. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Comput. Hum. Behav.* **2006**, *22*, 685–708. [\[CrossRef\]](#)
26. McDaniel, M.A.; Einstein, G.O. Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Appl. Cognit. Psychol.* **2000**, *14*, S127–S144. [\[CrossRef\]](#)
27. Einstein, A. *The Meaning of Relativity*; Routledge: Abingdon, UK, 2003.
28. Baron, R.S. Distraction-conflict theory: Progress and problems. In *Advances in Experimental Social Psychology*; Elsevier: Amsterdam, The Netherlands, 1986; Volume 19, pp. 1–40.
29. Hacker, W. Action regulation theory: A practical tool for the design of modern work processes? *Eur. J. Work Organ. Psychol.* **2003**, *12*, 105–130. [\[CrossRef\]](#)
30. Trafton, J.G.; Altmann, E.M.; Brock, D.P.; Mintz, F.E. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *Int. J. Hum.-Comput. Stud.* **2003**, *58*, 583–603. [\[CrossRef\]](#)

31. Latorella, K.A. Investigating interruptions: An example from the flightdeck. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Philadelphia, PA, USA, 2–6 September 1996; Volume 40, pp. 249–253.
32. McFarlane, D. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Hum.-Comput. Interact.* **2002**, *17*, 63–139. [[CrossRef](#)]
33. Sasse, A.; Johnson, C. Coordinating the interruption of people in human-computer interaction. In *Human-Computer Interaction, INTERACT*; IOS Press: Amsterdam, The Netherlands, 1999; Volume 99, p. 295.
34. Robertson, T.J.; Prabhakararao, S.; Burnett, M.; Cook, C.; Ruthruff, J.R.; Beckwith, L.; Phalgune, A. Impact of interruption style on end-user debugging. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24–29 April 2004; pp. 287–294.
35. Witt, H.; Drugge, M. Hotwire: An apparatus for simulating primary tasks in wearable computing. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2006; pp. 1535–1540.
36. Kreifeldt, J.; McCarthy, M. Interruption as a test of the user-computer interface. In Proceedings of the 17th Annual Conference on Manual Control (SEE N 82-13665 04-54), Los Angeles, CA, USA, 16–18 June 1981; pp. 655–667.
37. Field, G.E. Experimentus Interruptus. *SIGCHI Bull.* **1987**, *19*, 42–46. [[CrossRef](#)]
38. Nunes, L.; Recarte, M.A. Cognitive demands of hands-free-phone conversation while driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2002**, *5*, 133–144. [[CrossRef](#)]
39. Salvucci, D.D.; Taatgen, N.A. *The Multitasking Mind*; Oxford University Press: Oxford, UK, 2010.
40. Anderson, J.R. How Can the Human Mind Occur in the Physical Universe? Oxford University Press: Oxford, UK, 2009.
41. Borst, J.; Taatgen, N. The costs of multitasking in threaded cognition. In Proceedings of the 8th International, ICCM'07, Paris, France, 17–19 September 2007.
42. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.
43. Brookhuis, K.A.; de Waard, D. Assessment of Drivers' workload: Performance and Subjective and Physiological Indexes. In *Stress, Workload and Fatigue*; Laurence Erlbaum Associates Inc.: Mahwah, NJ, USA, 2001.
44. Wickens, C.D.; Hollands, J.G.; Banbury, S.; Parasuraman, R. *Engineering Psychology & Human Performance*; Psychology Press: London, UK, 2015.
45. Roscoe, A.H. Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biol. Psychol.* **1992**, *34*, 259–287. [[CrossRef](#)]
46. Rowe, D.W.; Sibert, J.; Irwin, D. Heart rate variability: Indicator of user state as an aid to human-computer interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, CA, USA, 18–23 April 1998; pp. 480–487.
47. Healey, J.A.; Picard, R.W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [[CrossRef](#)]
48. Beatty, J. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* **1982**, *91*, 276. [[CrossRef](#)] [[PubMed](#)]
49. Katidioti, I.; Borst, J.P.; Bierens de Haan, D.J.; Pepping, T.; van Vugt, M.K.; Taatgen, N.A. Interrupted by your pupil: An interruption management system based on pupil dilation. *Int. J. Hum.-Comput. Interact.* **2016**, *32*, 791–801.
50. Katidioti, I.; Borst, J.P.; Taatgen, N.A. What happens when we switch tasks: Pupil dilation in multitasking. *J. Exp. Psychol. Appl.* **2014**, *20*, 380. [[CrossRef](#)] [[PubMed](#)]
51. Kahneman, D.; Beatty, J. Pupil diameter and load on memory. *Science* **1966**, *154*, 1583–1585. [[CrossRef](#)] [[PubMed](#)]
52. Kahneman, D.; Onuska, L.; Wolman, R.E. Effects of grouping on the pupillary response in a short-term memory task. *Q. J. Exp. Psychol.* **1968**, *20*, 309–311. [[CrossRef](#)] [[PubMed](#)]
53. Gable, T.M.; Kun, A.L.; Walker, B.N.; Winton, R.J. Comparing heart rate and pupil size as objective measures of workload in the driving context: Initial look. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, UK, 1–3 September 2015; pp. 20–25.
54. Linderman, B. *Word Problems Homework Booklet, Grades 7–8*; Instructional Fair: Greensboro, NC, USA, 1999; ISBN 0-88012-865-8.

55. Ostrach, T.R. *Typing Speed: How Fast is Average: 4,000 Typing Scores Statistically Analyzed and Interpreted*; Five Star Staffing, Inc.: Orlando, FL, USA, 1997.
56. Duchowski, A.T. A breadth-first survey of eye-tracking applications. *Behav. Res. Methods Instrum. Comput.* **2002**, *34*, 455–470. [[CrossRef](#)] [[PubMed](#)]
57. Kun, A.L.; Palinko, O.; Medenica, Z.; Heeman, P.A. On the feasibility of using pupil diameter to estimate cognitive load changes for in-vehicle spoken dialogues. In *Interspeech*; ISCA: Singapore, 2013; pp. 3766–3770.
58. Baraldi, A.N.; Enders, C.K. An introduction to modern missing data analyses. *J. School Psychol.* **2010**, *48*, 5–37. [[CrossRef](#)] [[PubMed](#)]
59. Lee, B.C.; Duffy, V.G. The effects of task interruption on human performance: A study of the systematic classification of human behavior and interruption frequency. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2015**, *25*, 137–152. [[CrossRef](#)]
60. Wickens, C.D.; McCarley, J.S. *Applied Attention Theory*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2007; ISBN 0-8058-5983-7.
61. Hodgetts, H.M.; Jones, D.M. Resuming an interrupted task: Activation and decay in goal memory. In Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006), Vancouver, UK, 26–29 July 2006; p. 2506.
62. Monk, C.A.; Trafton, J.G.; Boehm-Davis, D.A. The effect of interruption duration and demand on resuming suspended goals. *J. Exp. Psychol. Appl.* **2008**, *14*, 299. [[CrossRef](#)] [[PubMed](#)]
63. Speier, C.; Valacich, J.S.; Vessey, I. The influence of task interruption on individual decision making: An information overload perspective. *Decis. Sci.* **1999**, *30*, 337–360. [[CrossRef](#)]
64. Pflieger, B.; Fekety, D.K.; Schmidt, A.; Kun, A.L. A model relating pupil diameter to mental workload and lighting conditions. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 5776–5788.
65. Miyata, Y.; Norman, D.A. Psychological issues in support of multiple activities. In *User Centered System Design: New Perspectives on Human-Computer Interaction*; CRC: Boca Raton, FL, USA, 1986; pp. 265–284.
66. Kolbeinsson, A.; Thorvald, P.; Lindblom, J. Coordinating the interruption of assembly workers in manufacturing. *Appl. Ergon.* **2017**, *58*, 361–371. [[CrossRef](#)] [[PubMed](#)]
67. Grundgeiger, T.; Sanderson, P.; MacDougall, H.G.; Venkatesh, B. Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *J. Exp. Psychol. Appl.* **2010**, *16*, 317. [[CrossRef](#)] [[PubMed](#)]
68. Borst, J.P.; Taatgen, N.A.; van Rijn, H. What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 2971–2980.

