Effects of Mediating Notifications Based on Task Load

Rahul Rajan
Carnegie Mellon University
Pittsburgh, USA
rahulraj@cmu.edu

Ted Selker
Aarhus University
Aarhus, Denmark
ted.selker@gmail.com

Ian Lane
Carnegie Mellon University
Pittsburgh, USA
lane@cmu.edu

ABSTRACT
We know that talking or texting while involved in a complex sensorimotor task like driving is dangerous. In such situations, interruptions from notifications can negatively impact primary task performance as well. This paper investigates the impact of attending to such notifications. In our study, participants were engaged in a primary task with varying task loads. Notifications were presented to them aurally and visually, and were mediated relative to the primary task load. Our results show that a) attending to notifications were distracting regardless of modality, b) mediation helped users comprehension of visual notifications, but did not effect their comprehension of audio notifications, and c) even though mediated notifications reduced performance degradation, users did not notice or choose it differentially.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Mobile Notifications; Driver Distraction; Cognitive Load; Mediation.

INTRODUCTION
The effects of notifications have been studied in the office-desktop environment when engaged in primary tasks like editing or programming. However, less has been done to understand the nature and effect of mobile notifications in everyday life. Even as we begin to study them, notification strategies are evolving; wearable devices, like smartwatches and head-mounted displays, are being developed that aim to focus our attention on notifications while touting seamless integration. Despite their interruptive nature, notifications might be our only way to keep abreast of time critical requests. Thus, there is a need to develop effective strategies to manage people's attention without subjecting them to undue risk.

The asynchronous nature of notifications afford the user the ability to decide when to take action on a secondary or new activity. While in some cases immediate action is taken by the user, in other cases notifications must be ignored depending on the user's current context. The task of *attending* to these notifications, and making a decision on whether to take an action or not, is currently left up to the user. Interuption manager [2] calls the notification an interuption, and reacting to it a disruption. We want to understand the cost associated with the attending to a notification, and the role played by the modality, i.e. audio or visual. The primary task could be any immersive task involving complex sensorimotor skills, like driving, cooking, or even surgery. We present our experiments with ConTRe (Continues Tracking and Reaction) [21], which requires continuous tracking and episodic reactions in a driving-like task.

Automotive cockpits have been gaining attention because of the real danger that driver distraction can have on road traffic safety. A number of studies have shown how operating mobile devices and other in-vehicle infotainment systems is critically 'impacting' driving performance and is a major factor in automotive accidents. Results of testing these effects in simulator studies has been shown to be replicable in field studies [11]. A broad literature has demonstrated that interacting with telephones and similar secondary activities in the car can adversely affect the primary driving task. This paper explores how even simply attending to tasks that do not require a response might impact performance. Such tasks might be as simple as attending to notifications, which is the subject of our study.

RELATED WORK
We start off by describing work that has studied the interruptive nature of notifications, and associated risks. We then present strategies that others have prescribed to mitigate these risks, particularly through the use of mediation to manage attention allocation.

The Interruptive Nature of Notifications
Iqbal and Bailey [16] define a *notification* as a visual cue, auditory signal, or haptic alert generated by an application or service that relays information to a user outside their current focus of attention. A majority of the research on notifications is focused on information workers in a desktop computing environment. Its detrimental effects on primary task performance and efficiency have been highlighted through numerous studies [3,9,20]. This effect was shown to be more pronounced when the primary task is cognitively demanding [8].

Effects of Multitasking
Repeated task switching during an activity may lead to completion of the primary task with lower accuracy and longer

Automotive'UI 16, October 24-26, 2016, Ann Arbor, MI, USA

DOI: http://dx.doi.org/10.1145/3003715.3005413

duration, in addition to increased anxiety and perceived difficulty of the task [4]. Multiple studies have shown how cell phone conversations, texting, or interacting with In-Vehicle Information Systems (IVIS) can be detrimental to driving safety [5, 12, 17, 24, 25]. Drivers engaging in such activities have been shown to have increased brake reaction time [1, 19], failure in scanning for potential hazards in the driving environment [26], and to have accidents with higher likelihood [23].

Mediating Interruptions & Notifications

Successful dual-task scenarios depend on the availability and requirements of cognitive resources for the secondary task given resource consumption by the primary task [28]. It also depends on the context of the secondary task, and its relationship to the users goals [2]. This presents opportunities to increase people's ability to successfully handle interruptions, and prevent expensive errors. McFarlane's seminal work proposed four methods for coordinating interruptions [22], including immediate, negotiated, mediated and scheduled. Mediation has been widely studied in the desktop computing domain [15, 16], but has not been adequately explored in post-desktop, mobile situations.

OVERVIEW OF STUDY

Our study had multiple goals. First, we wanted to determine how notifications impacted a driving-like primary task. We then wanted to establish how mediating them relative to task load could improve a user's performance on both the primary and secondary tasks. Finally, we wanted to understand these effects for both audio and visual notifications. Our research questions include:

1. Mediation: How is primary task performance effected when the user is attending to notifications? Can mediation reduce this impact?
2. Modality: How is primary task performance effected by modality? Does mediation have the same effect across both audio and visual modes?
3. How do both of these conditions, i.e. mediation and modality, effect a user's ability to comprehend a notification?

Pilot explorations with a driving simulator, while promising, had a number of limitations. The interaction of full driving experience made it difficult to replicate task loads and added unnecessary dependent variables to data collection. For these reasons we chose to use the well studied and widely used ConTRe (Continuous Tracking and Reaction) task [21], which provides a highly controlled yet unpredictable task load for the participant. This allows for consistent and replicable analysis.

The study was setup so that the primary ConTRe task would randomly switch between low and high workloads. This was done to simulate a typical driving scenario where drivers episodically experience high workload when they are entering/exiting highways, changing lanes, following navigation instructions, etc. For the secondary task, participants attended to notifications that were presented to them, as they performed the primary ConTRe task. Audio notifications were delivered via speakers, while visual notifications appeared through a reality augmentinging Heads Up Display (HUD). The audio notifications were created using Apple's text-to-speech engine on OS X Yosemite (Speaking voice: Alex; Speaking rate: Normal). The augmented reality HUD device used was a Google Glass, which projects the screen at a working distance of 3.5 m, approximately 35°elevated from the primary position of the eye.

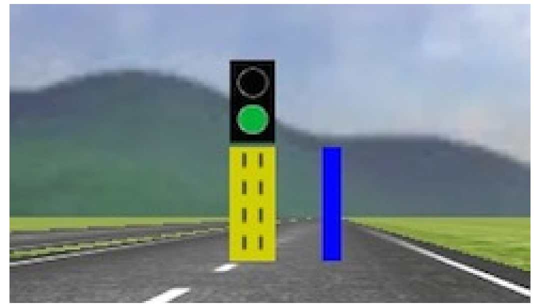

Figure 1: Screenshot of the ConTRe (Continuous Tracking and Reaction) Task that displays the yellow reference cylinder with the traffic light on top, and the blue tracking cylinder beside it.

Experimental Design

The study was designed as a 2 (Audio/Visual modes) X 2 (Mediated/Non-mediated conditions) repeated measures within subjects study. This was done to mitigate individual variance in performance for the primary and secondary tasks. To control for possible effects of order the study was double counterbalanced for mode and condition factors. Additionally, two baseline conditions included performing the ConTRe task in low and high workload settings without notifications.

Users

20 people participated in the study, recruited through a call sent out to students selected randomly from a graduate engineering school population. It included 10 males and 10 females. The mean age of the participants was 26.4 years, with a standard deviation of 2.7 years. Participants were rewarded with a $40 gift cards for completing the study.

Primary Task: ConTRe

The ConTRe task is an add-on for OpenDS, an open-source driving simulator [21]. It is an abstracted and simplified task that comprises of actions required for normal driving, i.e. operating the brake and acceleration pedals, as well as using the steering wheel. This focuses the user's task and simplifies the recording of tracking behavior. Fine grained measures of performance on the primary task relative to the secondary task requests can be obtained, which is necessary for our investigation.

Here the car moves with a constant speed on a unidirectional straight road consisting of two lanes. The simulator shows two cylinders at a constant distance in front of the car: a yellow reference cylinder, and a blue tracking cylinder. The yellow reference cylinder moves autonomously and unpredictably. The lateral position of the blue tracking cylinder is controlled by the user through the use of the steering wheel. The cylinder moves left or right depending on the direction and angular velocity of the steering wheel, i.e the steering wheel controls the cylinder's lateral acceleration. Their goal is to track the yellow reference cylinder, by overlapping it with the user-controlled blue cylinder, as closely as possible. Effectively, this corresponds to a task where the user has to follow a curvy road. For the low and high task load conditions, the lateral speed of the reference cylinder was set to values that were empirically determined to create low and high workloads for the user, respectively.

Furthermore, there is a signal light that illuminates an upper circle red or lower circle green. The signal light is atop a yellow reference cylinder. At any time, neither of the lights or only one is turned on. The red light requires that the user respond by depressing a logitech driving simulator brake pedal, while the green light corresponds to the driving simulator's accelerator pedal. This operates independently of the steering function. As soon as the user reacts to the light by depressing the correct pedal, the light turns off.

Secondary Task: Notifications

The secondary notification task is based on widely used measures of working memory capacity, which include operation span and reading span tasks [7]. Working memory has been purported to be involved in a wide range, of complex cognitive behaviors, such as comprehension, reasoning, and problem solving as it is thought to reflect primarily domain-general, executive attention demands of the task [10]. In this work we do not aim to measure working memory, but instead want to measure the effect of processing a notification across the four experimental conditions. Thus, we modify the span tasks for our purposes as described below.

In each condition, drivers were presented with a series of twenty items, which included ten math equations and ten sentences taken from widely used span tasks [7] (see Table 1). The math equations and sentences are representative of the symbolic and verbal types of notifications, respectively, that users typically receive. Using standardized stimuli allows for consistency and replicability. Both types of notifications were randomly interspersed, so as to prevent the person from getting into a rhythm of expecting either one. After the subject had read or listened to each item, they verbally indicated if the notification was *true* or *false*. Sentences are true when they are semantically and syntactically correct, while the math equations are true when the equality holds.

After each item, the participant was presented with an isolated letter, which represents something they had to remember from the notification. After two, three or four items, the ConTRe simulator was paused, and they were asked to recall the letters in sequence. This is like *responding* to a text message or similar notification while the car is stopped at a light. Recall tasks are already known to have the most detrimental effects on primary task performance [18]. Pausing the simulator separates the recall effort from the recorded ConTRe task performance (even today's drivers are encouraged to stop their car before interacting with any request from their phone). This focuses the experiment solely on *attending* to notifications and its resulting effect on task performance.

Table 1: Examples of the two types of notifications

Type	**Notification**
Math	2/2 + 1 = 1
Sentence	After yelling at the game, I knew I would have a tall voice

Mediation

Mediation was done relative to task load. In the non-mediated (control) condition, notifications were presented randomly in both the low and high workloads. In the mediated (test) condition, notifications were presented only during low workload. The bounded deferral technique [14] was used, where the notifications would be delayed while the driver was in a high workload setting. The notification would then be delivered a few seconds into the low workload setting. The mediation was conducted by one of the experimenters who had full view of the simulator and could determine when to deliver the notification. Modality apprpriate changes were made if a notification had been delivered, and the workload changed from low to high before the driver responded. For the audio mode, the notification could be paused and continued at the next low workload period, or simply repeated. In the visual mode, the notification could be hidden till the next low workload period, when it would become visible again.

Before each condition, participants were told whether they would be receiving audio or visual notifications. However, they did not receive any indication as to whether the notifications would be mediated by task load.

Methodology

Participants arriving at the lab were guided through an informed consent process, followed by an overview of the study. They were aided through the process of setting up a number of sensors attached to their body to record their physiological responses, including heart rate, electrodermal activity, skin temperature, etc. The participant was then seated in the simulator and shown how notifications would be delivered on Glass, and through speakers.

The participant was then taken through a series of practice runs to get them comfortable with the primary ConTRe task. When done with the practice, the low benchmark was recorded using the low workload setting on the simulator. After one minute, they were asked to repeat a series of ten sentences that were read out to them, one-by-one, while they were still performing the ConTRe task. The same routine was performed to record the high benchmark using the high workload setting on the simulator.

This was followed by another set of practice rounds, where the secondary notification task was described and demonstrated to the participant. Next, a practice trial combined both the ConTRe task (with the randomly alternating workloads) and the notifications task. The notification task included a set of five items, three of which were math equations, and two were sentences. This provided the participants with a sense of what to expect during the actual trials. The practice trials could be repeated until the participant felt comfortable that they understood the scenario.

The participants then moved on to the experimental trials. Each participant was presented with four trials, one for each condition. At the end of the four trials, the participant was interviewed about the disruptivity and effectiveness of audio and visual notifications. The entire study lasted approximately 2 hours per participant.

Measures

Quantitative performance data on both primary and secondary tasks were collected. From the ConTRe task, collected performance data included: steering deviation, i.e. the difference in distance between the reference cylinder and the tracking cylinder; reaction times to respond to the red and green lights, i.e. the amount of time from when the light went off to when the correct pedal was depressed; and the error rate of depressing the wrong pedal. These measures were automatically recorded by the simulator.

In the mediated condition, notifications were presented in the low workload section. In the non-mediated condition, notifications were presented in the low and high workload sections. The expectation was that the ConTRe performance in the low workload sections would be identical for both conditions. Hence, the analysis focused on the performance data from the high workload sections of the mediated and non-mediated conditions. Steering deviation was continuously sampled at 570 Hz. A rolling median was used to filter out noise and infrequent occurrences of sudden deviations from the trend. The average steering deviation of each user in each condition was then recorded. Accelerator and brake reaction tasks produced an average of 31.5 brake reaction and 31.3 accelerator reaction data points per user for each condition. Like the steering deviation, the reaction times for both the brake and accelerator tasks were low-pass filtered using a rolling median. The mean reaction times were then calculated and recorded for each user in each of the four conditions.

For performance on the secondary notification task, the response times was the time from when the notification was presented to the driver, to when they respond to indicate true or false. As before, the data is filtered using a rolling median. The mean response times for math and sentences are then recorded for each user in every condition. The errors in the responses were also calculated, as well as the error in recalling the sequence of letters that were presented to the driver after each notification. The sequence could be two, three or four letters long.

At the end of all the trials, participants were interviewed about their preferences regarding the modality of the notification, and the effect of its disruptivity on their primary task performance. They were also asked if they perceived any difference between the two audio or the two visual conditions, i.e between the mediated and non-mediated conditions.

RESULTS

Processing the data as described in the previous section produced 10 data points per user (one for each measure) for each of the 2 (Modality) X 2 (Mediation) conditions: Audio Mediated (AM), Audio Non-mediated (AN), Visual Mediated (VM), Visual Non-mediated (VN). This totals to 40 data points per user, and a total of 800 data points. Described below are the results from the analysis of these data points, starting with the primary ConTRe task.

Effects on Primary ConTRe Task

Analysis of the ConTRe performance measures was done to evaluate the effects of mediation and modality on the primary task. These measures include Steering Deviation, Reaction Time for Acceleration and Braking, and Errors in Acceleration and Braking. A multivariate ANOVA (MANOVA) was performed using all five driving performance measures as dependent variables. As opposed to running multiple univariate *F* tests for each dependent variable, MANOVA has the advantage of reducing the likelihood of a Type I error, and revealing differences not discovered by ANOVA tests [27].

A two-factor repeated measures MANOVA with within-subject factors (Mediation, Modality) showed a significant effect from Mediation, $F(1,19) = 25.46$, $p < .001$, and no significant effect from Modality $F(1,19) = 1.16$, $p = .29$. There was no significant interaction between the two main effects $F(1,19) = 1.20$, $p = .28$, which validates the main effect analysis. The notifications were distracting and negatively impacted user performance on the ConTRe task. It did not matter if the notifications were audio or visual.

Given the omnibus multivariate *F*-test revealed a significant effect from Mediation, we further analyze its effect on the different metrics separately. The effect from Modality was also analysed. Thus for each measure, we present four planned comparisons using paired t-tests: a) Mediated and Non-mediated *Audio* (AM-AN), b) Mediated and Non-mediated *Visual* (VM-VN), c) *Mediated* Audio and Visual (AM-VM), and d) *Non-mediated* Audio and Visual conditions (AN-VN) (see Table 2). To control for Type I errors we use the Bonferroni adjusted alpha levels of .0125 per test (.05/4).

The first two planned comparisons (AM-AN & VM-VN) emphasizes the impact of mediation in the audio and visual modalities, separately. The next two planned comparisons (AM-VM & AN-VN) contrasts the audio with the visual modes. As there were no notifications presented to the participants in the mediated condition, the ConTRe performance should be identical for the audio and visual modes in the mediated condition, i.e. VM and AM. Any difference in the effect from modality should be borne out on the ConTRe task performance in the nonmedaited condition, i.e. VN and AN.

Steering Deviation

Comparing the steering deviation for Mediated ($M = 18.02$ %, $SD = 3.27$ pp) and Non-mediated conditions ($M = 17.75$ %, $SD = 3.08$ pp) in the *Audio* mode does not reveal any significance, $t(19) = 0.71$, $p = .48$, which matches results from previous work that found cognitive load costs are minimally borne out on steering deviation [6, 12]. In the *Visual* mode, there was a

Table 2: Average pair-wise difference for each primary task measure, with p-values from paired t-tests in parenthesis.

Primary Task Measures	**AM-AN**	**VM-VN**	**AM-VM**	**AN-VN**
Steering Deviation (pp)	0.27 (.48)	-1.66 (**.001**)	-0.13 (.67)	-2.06 (< **.001**)
Acceleration Reaction Time (ms)	-109.75 (**.002**)	-156.13 (< **.001**)	13.84 (.52)	-32.53 (.45)
Brake Reaction Time (ms)	-90.71 (.017)	-142.05 (.015)	-18.49 (.59)	-69.83 (.17)
Acceleration Response Error (pp)	-3.05 (.06)	-1.57 (.34)	-0.91 (.59)	0.57(.76)
Brake Response Error (pp)	-3.26 (.08)	-0.87 (.65)	-1.92 (.35)	0.47 (.78)

significant difference between Mediated (M = 18.15 %, SD = 2.95 pp) and Non-mediated conditions (M = 19.81 %, SD = 2.75 pp), $t(19) = -3.84$, $p = .001$. This indicates that the Visual mode effects the visual requirements of the primary task, i.e. tracking the lateral movement of the system-controlled yellow cylinder. This is consistent with findings in the literature which indicate that tracking will be negatively impacted by glances away from the road [13].

In the *Mediated* condition, comparing the Audio and Visual modes showed no significant differences. As explained before, the *Mediated* condition is equivalent to driving without any notifications. In the *Non-mediated* condition, Audio was significantly less disruptive than the Visual mode $t(19) = -0.43$, $p < 0.001$. Again, this can be attributed to adding a visual source of distraction to a primary task that depends on visual input.

Reaction Time for Acceleration and Braking

In the *Audio* mode, the mean reaction time for acceleration was significantly reduced in the Mediated condition (M = 892.2 ms, SD = 176.4 ms) as compared to the Non-mediated condition (M = 1001.9 ms, SD = 214.6 ms), $t(19) = -3.52$, $p = .002$. The same effect carried on into the *Visual* mode with the Mediated condition (M = 878.4 ms, SD = 152.4 ms) being significantly less than the Non-mediated condition (M = 1034.5 ms, SD = 214.6 ms), $t(19) = -4.81$, $p < 0.001$. This is again consistent with findings which indicate that diverting focal attention from the road will result in longer reaction times [12]. The difference between Audio and Visual was not significant in the *Mediated* or *Non-mediated* condition.

Performing the same analysis for the braking reaction times, the *Audio* mode showed a near significant difference between the Mediated (M = 995.6 ms, SD = 216.8 ms) and Non-Mediated means (M = 1086.3 ms, SD = 238.8 ms), $t(19) = -2.59$, $p = .017$. In the *Visual* mode, the difference was near significant as well with the Mediated condition (M = 1014.2 ms, SD = 189.1 ms) being lower than the Non-mediated condition (M = 1156.2 ms, SD = 254.8 ms), $t(19) = -2.67$, $p = .015$. No difference was found when comparing the Audio and Visual modalities.

While acceleration and braking test for similar things, the slight increase in reaction times for braking compared to acceleration might be attributed to the extra time it takes the user to move their foot from the accelerator pedal (over which it used to hover by default for most users) to the braking pedal. It is plausible that the act of braking itself introduces a larger manual source of distraction compared to acceleration.

Errors in Acceleration and Braking

As with the reaction time analysis, we begin by analyzing the acceleration results. In the *Audio* mode, there were fewer errors in the Mediated condition (M = 10.64 %, SD = 8.94 pp) as compared to the Non-mediated condition (M = 13.69 %, SD = 7.24 pp), $t(19) = -1.96$, $p = 0.06$, but this did not reach significance. In the *Visual* mode there was no difference between the Mediated (M = 11.54 %, SD = 6.27 pp) and Non-mediated conditions (M = 13.11 %, SD = 5.56 pp), $t(19) = -0.97$, $p = 0.34$. Again for both the *Mediated* and *Non-mediated* conditions, no difference was found between the Audio and Visual modalities.

For the errors in braking responses, there was a slight differences between the Mediated (M = 10.02 %, SD = 7.37 pp) and Non-mediated conditions (M = 13.26 %, SD = 9.50 pp), $t(19) = -0.46$, $p = 0.08$, in the *Audio* mode, but this did not reach significance. In the *Visual* mode, there was no significant difference between the Mediated (M = 11.91, SD = 7.10 pp) and Non-mediated conditions (M = 12.78, SD = 6.59 pp), $t(19) = -0.46$, $p = 0.65$. No difference was found in the *Mediated* and *Non-mediated* conditions across both modalities. Due to the similarity in the acceleration and braking response, only the acceleration response errors were plotted.

Effects on Secondary Notification Task

The main effects of Mediation and Modality on the notification task are analyzed through the following measures: Response Times for Math and Sentences, Response Errors for Math and Sentences, and Recall. Similar to the primary driving task analysis, this was done with a multivariate ANOVA (MANOVA) using all five notification task measures as dependent variables.

A two-factor repeated measures MANOVA using within-subject factors (Mediation, Modality) showed that all effects were significant at the .05 significance level. The main effect of Mediation yielded an F ratio of $F(1,19) = 5.49$, $p = .03$. The main effect of Modality yielded an F ratio of $F(1,19) = 12.81$, $p = .002$. There was also a significant interaction effect, $F(1,19) = 6.90$, $p = .017$. An analysis of the simple effects for each of the two levels in the independent variables (Mediation, Modality) explains this interaction.

The simple effects of Mediation is analysed by setting the independent Modality variable to *Audio*. A one-way MANOVA

Table 3: Average pair-wise difference for each secondary task measure, with p-values from paired t-tests in parenthesis.

Secondary Task Measures	**AM-AN**	**VM-VN**	**AM-VM**	**AN-VN**
Math Response Time (ms)	0.03 (.80)	-0.57 (.02)	2.52 (< **.001**)	1.92 (< **.001**)
Sentence Response Time (ms)	-0.008 (.95)	-0.52 (.04)	-0.09 (.75)	-0.61 (.07)
Math Response Error (pp)	1.92 (.48)	2.11 (.41)	-2.57 (.29)	-2.38 (.35)
Sentence Response Error (pp)	-4.02 (.27)	0.34 (.93)	12.99 (**.002**)	17.35 (**.001**)
Recall Error (pp)	1.05 (.74)	-5.76 (.07)	-1.29 (.68)	-8.10 (**.01**)

with Mediation as the within-subject variable showed no significant effect between the Mediated and Non-mediated conditions, $F(1,19) = 0.03$, $p = .85$. Whereas, setting the independent Modality variable to *Visual*, revealed a significant effect, $F(1,19) = 7.52$, $p = .01$. This implies that audio notifications are comprehended equally well under low and high workloads. Visual notifications, on the other hand, are comprehended differently under low and high workloads.

The simple effects of Modality is analysed by setting the independent Mediation variable to *Mediation*. A one-way MANOVA with Modality as the within-subject variable showed a highly significant effect of mediation between Audio and Visual modes $F(1,19) = 28.98$, $p < .001$. Setting the independent variable to *Non-mediation* did not show a significant effect $F(1,19) = 3.84$, $p = .06$. This analysis infers that under low workloads, users comprehend audio and visual notifications differently. Under high workloads, modality of notifications does not effect comprehension ability.

To understand the direction of the differences, and the impact on the different dependent variables, four planned comparisons using paired t-tests were performed, similar to the primary task analysis. For each dependent variable, we describe below comparisons between: a) effect of mediation in the *Audio* mode (AM-AN), b) effect of mediation in the *Visual* mode (VM-VN), c) effect of modality in the *Mediated* conditions (AM-VM), d) and the effect of modality in the *Non-mediated* conditions (AN-VN) (see Table 3). To control for Type I errors we use the Bonferroni adjusted alpha levels of .0125 per test (.05/4).

Response Times for Math and Sentences

Analysis of the reaction times for math in *Audio* mode, showed no difference between the Mediated (M = 5.69 s, SD = 0.43 s) and Non-mediated conditions (M = 5.66 s, SD = 0.47 s), $t(19) = 0.25$, $p = .8$. In the *Visual* mode there was a difference between the Mediated (M = 3.17 s, SD = 0.9 s) and Non-mediated conditions (M = 3.74 s, SD = 1.70 s), $t(19) = 0.25$, $p = .02$, but it did not reach significance. There was a highly significant difference in the reaction times between the Audio and Visual modes in both the *Mediated* and *Non-mediated* conditions, $p < .001$.

The differences in reaction times for sentences were less dramatic. In the *Audio* mode there was no difference between the Mediated (M = 5.10 s, SD = 0.63 s) and Non-mediated cases (M = 5.11 s, SD = 0.84 s), $t(19) = -0.056$, $p = .95$. In the *Visual* mode, there was a slight difference in the Mediated (M = 5.20 s, SD = 1.41) and Non-mediated conditions (M = 5.72 s, SD = 1.94 s) , $t(19) = -2.16$, $p = .04$. In the *Mediated* case, there was no difference between the Audio and Visual conditions, $t(19) = -0.31$, $p = .75$. In the *Non-mediated* case, there was a slight difference, but it did not reach significance, $t(19) = -1.89$, $p = .07$.

Errors in Math and Sentences

For errors in math responses, there were no differences between all four of the planned comparisons, for which reason they were not plotted. (Audio: M = 8.14 %, SD = 8.29 pp; Visual: M = 6.22 %, SD = 8.62 pp; Mediated: M = 10.72 %, SD = 7.59 pp; Non-mediated: M = 8.61 %, SD = 10.42 pp).

For the errors in sentence responses, there was no significant difference in the *Audio* and *Visual* modes between the Mediated and Non-mediated conditions. In the *Mediated* condition there was a significant difference in the Audio (M = 28.69 %, SD = 15.00 pp) and Visual conditions (M = 15.69 %, SD = 10.78 pp), $t(19) = 3.51$, $p = .002$. Similarly, there was a significant difference in the *Non-mediated* case between the Audio (M = 32.71 %, SD = 14.59 pp) and Visual conditions (M = 15.35 %, SD = 13.36 pp), $t(19) = 3.81$, $p = .001$. From this analysis, we might infer that modality has a significant effect on users' ability to comprehend sentences accurately, regardless of mediation. They made fewer errors when sentences were presented visually, as opposed to aurally.

Errors in Recall

Analyzing the errors in recall revealed no significant difference in the *Audio* mode (Mediated: M = 10.45 %, SD = 10.99 pp; Non-mediated: M = 9.40 %, SD = 9.32 pp), $t(19) = 0.33$, $p = .74$. There was a slight difference in the *Visual* mode between the Mediated (M = 11.74 %, SD = 12.65 pp) and Non-mediated conditions (M = 17.51 %, SD = 13.34 pp), $t(19) = -1.92$, $p = .06$, which did not reach significance. Comparing the *Mediated* Audio and Visual conditions did not show any difference, $t(19) = -0.41$, $p = 0.68$. There was a significant difference between the *Non-mediated* Audio and Visual conditions, $t(19) = -2.86$, $p = .01$.

DISCUSSION

Our study was focused on investigating the effects of attending to symbolic and verbal notifications while performing a complex sensorimotor task. By mediating the notifications based on the task load, we wanted to understand its effects on both the primary sensorimotor task, and the secondary notification task. The experimental results revealed that notifications are indeed distracting and impacts primary task performance. This effect was significant for both aurally and visually presented notifications. Furthermore, the visual modality did allow users to perform better on the secondary notification task when the notifications were mediated to be shown only during low workload.

The analysis of the different measures for ConTRe task performance (Table 2) showed that: 1) The added visual source of distraction from visual notifications was borne out on the steering deviation measure during high task workloads. This effect was absent when the notifications were presented in audio; 2) Compared to steering deviation, acceleration and braking response times were the most effected by the increased cognitive load of having to process notifications; 3) The braking response times appeared to be longer than the acceleration response times possibly due to the added manual source of distraction in moving the foot to the brake pedal (we noticed that users hovered their foot over the accelerator pedal by default); 4) With regards to the acceleration and braking response errors, mediation had a insignificant positive effect in the audio mode that might turn significant with more participants. Mediation appeared to have no effect on braking and acceleration in the visual mode. It might be the case that even with mediation, there is a competition for visual resources in the visual mode.

From the analysis of the performance measures in the secondary notification task (Table 3), it is clear that: 1) users respond faster to visual notifications when they involved the symbolic cognitive task of considering an equation. This effect did not hold for comprehending sentences, which might be more representative of communication notifications. Perhaps reading sentences while driving required more glances than reading an equation. If so, the result indicates that people were self-mediating by making trade-offs between performing the ConTRe task and comprehending sentences; 2) With regards to accuracies in responses, users were making more errors in comprehending sentences when they were presented aurally compared to visually. It could be that iconic visual memory is less susceptible to disruption than serial audio memory. There was no effect on math errors, however; 3) While recalling a sequence of letters, the condition with non-mediated visual notifications presented the most difficulty to drivers.

CONCLUSION AND FUTURE WORK

Technology has made computing devices ubiquitous and proactive, and we have come to depend on these to receive information and communicate with the world. Notifications, due to their efficiency and perceived safety, have become the primary mode of interaction with the user. Prior research has shown that interacting with these devices, i.e. noticing, attending and responding to messages can be disruptive, and can even have safety implications for tasks such as driving. The current work investigates whether even attending to such notifications, while engaged in a complex sensorimotor task, can effect performance. We experimented with presenting notifications aurally and visually (using a heads-up display). We also experimented with mediating notifications relative to task load. Results showed that attending to notifications during high workloads negatively impacts primary task performance. This impact was independent of the modality used to present the notification. Our results also showed that users attend to notifications best when they are visually presented during low workloads. These results provide insights on how even attending to notifications can impact primary task performance. The choice of the notification being presented aurally or visually, symbolically or verbally, can impact its value, and must be taken into account when designing proactive systems.

REFERENCES

1. Håkan Alm and Lena Nilsson. 1995. The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis & Prevention* 27, 5 (1995), 707–715.

2. Ernesto Arroyo and Ted Selker. 2011. Attention and intention goals can mediate disruption in human-computer interaction. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 6947 LNCS. 454–470. DOI: http://dx.doi.org/10.1007/978-3-642-23771-3_34

3. Ernesto Arroyo, Shawn Sullivan, and Ted Selker. 2006. CarCoach: a polite and effective driving coach. In *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems (CHI EA '06)*, Vol. 2. ACM, New York, NY, USA, 357–362. DOI: http://dx.doi.org/10.1145/1125451.1125529

4. B.P. Bailey, J.A. Konstan, and J.V. Carlis. 2000. Measuring the effects of interruptions on task performance in the user interface. In *IEEE International Conference on Systems, Man, and Cybernetics, 2000*, Vol. 2. IEEE, 757–762. DOI: http://dx.doi.org/10.1109/ICSMC.2000.885940

5. Jeff K. Caird, Kate A. Johnston, Chelsea R. Willness, Mark Asbridge, and Piers Steel. 2014. A meta-analysis of the effects of texting on driving. *Accident Analysis and Prevention* 71 (2014), 311–318. DOI: http://dx.doi.org/10.1016/j.aap.2014.06.005

6. Jeff K. Caird, Chelsea R. Willness, Piers Steel, and Chip Scialfa. 2008. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis and Prevention* 40, 4 (jul 2008), 1282–1293. DOI: http://dx.doi.org/10.1016/j.aap.2008.01.009

7. Andrew R. A. Conway, Michael J. Kane, Michael F. Bunting, D. Zach Hambrick, Oliver Wilhelm, and Randall W. Engle. 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12, 5 (oct 2005), 769–786. DOI: http://dx.doi.org/10.3758/BF03196772

8. Edward Cutrell, Mary Czerwinski, and Eric Horvitz. 2001. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. (2001). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.26.418

9. Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A Diary Study of Task Switching and Interruptions. In *CHI '04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vol. 6. ACM Press, New York, New York, USA, 175–182. DOI: http://dx.doi.org/10.1145/985692.985715

10. Randall W. Engle. 2002. Working memory capacity as executive attention. *Current Directions in Psychological Science* 11, 1 (feb 2002), 19–23. DOI: http://dx.doi.org/10.1111/1467-8721.00160

11. Johan Engström, Emma Johansson, and Joakim Östlund. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 2 (2005), 97–120.

12. William J. Horrey and Christopher D. Wickens. 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors* 48 (2006), 196–205. DOI: http://dx.doi.org/10.1518/001872006776412135

13. William J. Horrey, Christopher D. Wickens, and Kyle P. Consalus. 2006. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied* 12, 2 (2006), 67.

14. Eric Horvitz, Johnson Apacible, and Muru Subramani. 2005. Balancing Awareness and Interruption : Investigation of Notification Deferral Policies. *User Modeling 2005* (2005), 433–437. DOI: http://dx.doi.org/10.1007/11527886_59

15. Eric Horvitz, Carl Kadie, Tim Paek, and David Hovel. 2003. Models of attention in computing and communication: from principles to applications. *Communnications of the ACM* 46, 3 (2003), 52–59.

16. Shamsi T. Iqbal and Brian P. Bailey. 2010. Oasis. *ACM Transactions on Computer-Human Interaction* 17, 4 (dec 2010), 1–28. DOI: http://dx.doi.org/10.1145/1879831.1879833

17. Shamsi T. Iqbal, Eric Horvitz, Yun-cheng Ju, and Ella Mathews. 2011. Hang on a Sec ! Effects of Proactive Mediation of Phone Conversations While Driving. *Work* (2011), 463–472. DOI: http://dx.doi.org/10.1145/1978942.1979008

18. Shamsi T. Iqbal, Yun-Cheng Ju, and Eric Horvitz. 2010. Cars, calls, and cognition: Investigating driving and divided attention. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)* (2010), 1281–1290. DOI: http://dx.doi.org/10.1145/1753326.1753518

19. J. D. Lee, B. Caven, S. Haake, and T. L. Brown. 2001. Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors* 43 (2001), 631–640. DOI: http://dx.doi.org/10.1518/001872001775870340

20. Luis Leiva, Matthias Böhmer, Sven Gehring, and Antonio Krüger. 2012. Back to the app: The Costs of Mobile Appication Interruptions. *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '12* (2012), 291–294. DOI: http://dx.doi.org/10.1145/2371574.2371617

21. Angela Mahr, Michael Feld, Mohammad Mehdi Moniri, and Rafael Math. 2012. The ConTRe (Continuous Tracking and Reaction) task: A flexible approach for assessing driver cognitive workload with high sensitivity. *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* (2012), 88–91. http://www.dfki.de/web/forschung/publikationen/renameFileForDownload?filename=ConTRe

22. Daniel McFarlane. 2002. Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human-Computer Interaction* 17, 1 (mar 2002), 63–139. DOI: http://dx.doi.org/10.1207/S15327051HCI1701_2

23. Donald A. Redelmeier and Robert J. Tibshirani. 1997. Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine* 336, 7 (1997), 453–458.

24. Dario D. Salvucci, Daniel Markley, Mark Zuber, and Duncan P. Brumby. 2007. iPod distraction: Effects of portable music-player use on driver performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 243–250.

25. David L. Strayer and William A. Johnston. 2001. Driven to distraction: dual-Task studies of simulated driving and conversing on a cellular telephone. *Psychological science : a journal of the American Psychological Society / APS* 12 (2001), 462–466. DOI: http://dx.doi.org/10.1111/1467-9280.00386

26. T. Taylor, A. K. Pradhan, G. Divekar, M. Romoser, J. Muttart, R. Gomez, A. Pollatsek, and D. L. Fisher. 2013. The view from the road: The contribution of on-road glance-monitoring technologies to understanding driver behavior. *Accident Analysis and Prevention* 58 (2013), 175–186. DOI: http://dx.doi.org/10.1016/j.aap.2013.02.008

27. Russell T. Warne. 2014. A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists. *Practical Assessment, Research & Evaluation* 19, 17 (2014), 2.

28. Christopher D. Wickens. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3 (2002), 159–177. DOI: http://dx.doi.org/10.1080/14639220210123806