Hang on a Sec! Effects of Proactive Mediation of Phone Conversations while Driving

Shamsi T. Iqbal[1], Eric Horvitz[1], Yun-Cheng Ju[1] and Ella Mathews[2]

[1]Microsoft Research
One Microsoft Way, Redmond, WA 98052
{shamsi, horvitz, yuncj}@microsoft.com

[2]California Institute of Technology
1200 E California Boulevard, Pasadena, CA 91125
emathews@caltech.edu

ABSTRACT

Conversing on cell phones while driving is a risky, yet commonplace activity. State legislatures in the U.S. have enacted rules that limit hand-held phone conversations while driving but that allow for hands-free conversations. However, studies have demonstrated that the cognitive load of conversation is a significant source of distraction that increases the likelihood of accidents. We explore in a controlled study with a driving simulator the effectiveness of proactive alerting and mediation of communications during phone conversations while driving. We study the use of auditory messages indicating upcoming critical road conditions and placing calls on hold. We found that such actions reduce driving errors and that alerts sharing details about situations were more effective than general alerts. Drivers found such a system valuable in most situations for maintaining driving safety. These results provide evidence that context-sensitive mediation systems could play a valuable role in focusing drivers' attention on the road during phone conversations.

Author Keywords

Driving, attention, context-sensitive systems, cell phones.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g. HCI): Miscellaneous.

General Terms

Experimentation, human factors, measurement.

INTRODUCTION

Engaging in cell phone conversations while driving is a dangerous yet common practice. Recognizing the threat to road safety that simultaneous driving and conversing poses to drivers and others, several U.S. states have enacted laws that limit calls to hands-free configurations [1]. Yet studies have shown that using hands-free setups may not reduce the distraction [22, 28], and may lull drivers into a false sense of safety. More than half of all U.S. states forbid novice drivers and school bus operators from any use of cell phones while driving [1]. However, general bans on phone conversations while driving are unlikely given expectations of modern mobile communication. Thus, we seek solutions that enable safer driving while communicating, focusing on opportunities for automated mediation of communications.

Automated mediation of human attention in multitasking situations has been a focus of the research community, especially in the desktop domain [11, 12]. Our broader goal is to explore the effectiveness of similar mediation in driving scenarios, where a context-sensitive system can recognize when a phone call is in progress and intervene by interrupting the call to deliver alerts about road conditions, and even suspend conversations if deemed beneficial. Such systems might jointly share interventions with drivers and the people they are speaking with to provide the remote participants with useful grounding about the driving situation [27]. Establishing such mutual understanding may help with joint regulation of conversations, enabling drivers to better focus on maintaining driving safety.

As a first step, we conducted a controlled study using a medium fidelity driving simulator. Different interventions were tested as drivers piloted a car within the simulator and engaged in phone conversations with remote users. We investigated the effectiveness and acceptability of short and long intervention messages, putting the call on hold, and the timing of an intervention, exploring outcomes of actions that interrupt conversations at natural conversational breakpoints versus at other times.

Our results showed that interventions positively affected driving, reducing driving errors compared to the baseline. The positive effect on driving safety was associated with perceived degradations in the quality of the conversation experience. Subjective feedback showed that drivers found interventions to be useful for most situations. Their general preference was towards having a shorter shared message at a conversational break followed by the call being put on hold. Non-driving participants in conversations were more neutral about the use of interventions, despite the fact that all were experienced drivers themselves. We believe the results highlight the promise of proactive mediation of phone conversations during driving and motivate further explorations. The results also make salient the need to educate the general population about the dangers of driving and talking on a cell phone whether the phone is balanced manually on the ear or embedded in a hands-free system.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

RELATED WORK

We shall first describe related work on how secondary tasks distract driving, strategies to help better manage secondary tasks, and the role of intervention in overall attention management, with a focus on the driving domain.

Driving and Distraction

While piloting a vehicle, core aspects of driving are “automated” and easily backgrounded as other tasks come to the fore [16]. Multiple studies have shown the detrimental effects of cell phone conversations, texting, or interacting with music players on driving safety [3, 4, 8, 23, 25]. Drivers who engage in phone conversations while driving have been found to have slower braking reaction time [2, 15], degraded steering performance [4], and to have accidents with higher likelihood [22] than drivers piloting cars without phone conversations. The common solution of mandating that calls be made in a hands-free manner while driving, has not been found to improve safety [22, 28]. A constellation of findings suggests that it is not the motor action of holding the phone, but primarily the cognitive demands of dual-task scenarios that degrade performance on one or both tasks [21]. A recent study demonstrated that the difficulty of the driving situation and the content of the conversation influence both driving and conversation performance, where conversations that include information recall challenges have the most detrimental effects on driving [14]; multiple recall tasks of varying difficulty certainly occur in the course of typical conversations.

Strategies to manage secondary tasks during driving

Humans have the ability to perform multiple concurrent activities, given availability of cognitive resources and absence of conflicts in resource demands from different tasks [29]. Researchers have explored the abilities of people to interleave driving with other activities. Brumby et al. studied how dialing a phone number can be interleaved with driving by chunking the dialing task to be accommodated within driving [5]. In another study, it was found that searching for music in a music player without decomposing it into subtasks resulted in faster performance, but it was safer to prioritize the driving task at the cost of increasing performance time on the search task [6]. Iqbal et al. found that the formulation and relay of directions to a remote caller while piloting a vehicle has less impact on driving than answering questions that involve challenging information recall tasks [14]. The authors propose that the task of generating directions may be decomposed into smaller subtasks and interleaved with driving demands while retrieval tasks may employ a less decomposable process of engaging long-term memory.

We explore here the effects of interventions during phone conversations while driving. Although the main goal is to direct driver attention back to road conditions through informative alerts, we also look to minimize the additional distraction associated with alerting by identifying more opportune moments and least disruptive interrupting mechanisms.

Intervention or Mediation to Manage Attention

The notion of using mediation in helping users manage attention has been explored in many domains. McFarlane’s seminal work proposed mediation as one of four interruption management methods [19]. The concept of attention manager systems has been studied in the desktop computing domain [10-12]. Some of these systems infer the cost of interrupting a user, and use this cost in decisions about if and when to pass on incoming information to the user. One particularly effective approach for mediation is deferring interruptions using a bounded deferral technique [9], which means waiting until a task or perceptual breakpoint within a given deadline [13].

In the domain of driving, researchers have explored the effectiveness of systems for aiding driving, e.g. providing local danger alerts [7], mediating communications among car passengers [17], or persuading people to drive in a more economical manner [20]. A challenge that such systems face is finding the best moment and modality for delivering information to the driver. Such decisions are complicated in multitasking scenarios, where cognitive resources of the driver are already engaged in driving and secondary tasks.

Findings that conversations with collocated passengers have less negative impact on driving compared to those with remote callers motivate further research on sharing driving context with remote callers [27]. Manalavan et al. explored how signaling a remote caller can indirectly reduce driver distraction [18]. They first investigated whether providing cues about the driving situation to a remote caller caused the caller to suspend the conversation (no actual driving occurred). Using findings from the first phase of analysis as ground truth, they performed a separate study to see if driving was safer when a remote caller stopped talking. While these disjoint studies provide initial intuition on how cues may affect conversations and consequently driving safety, further understanding is needed of cues and interventions in more realistic scenarios where driving and conversing occur together. In related work, Schneider and Kiesler found that providing remote callers with driving context via a visual display of the driver’s view positively influenced driving performance during conversations [26].

We shall focus here on the influence on driving performance of delivering auditory alerting messages of different types and timings. We face the challenge of decisions about intervening with information coming in the midst of a preexisting dual-task scenario of driving and conversing, and seek to balance the value of the intervention with the goals of maintaining driving safety, while conforming to the social expectations of parties involved in the conversation.

STUDY

We pursue an understanding of the effects of different actions taken to direct drivers’ attention to upcoming road conditions while the drivers are engaged in a phone conversation. The research questions include:

1) How effective are the interventions in maintaining driving safety compared to no intervention?
2) How do the interventions influence the conversation?
3) How do the drivers and the other participants in the conversations feel about such interventions?

To answer these questions, we conducted a controlled study using a driving simulator where we explored different types of interventions during a phone conversation between a driver and a remote participant. The study was set up so that the remote participant (situated in a separate room) talked to the driver over a speaker phone while the driver was driving through a predetermined route in the simulator, mimicking a hands-free phone call setting. The auditory interventions interrupted the conversation when some critical event or situation was soon to occur on the route.

Experimental design
The study was designed as a 2 (Intervention Mode) X 2 (Call Hold) X 2 (Intervention Timing) repeated measures within subjects study. Each condition was repeated twice. Additionally, there were two baseline trials with only driving (no conversation), and two baseline trials with only answering questions (no driving).

Users
Users participated in pairs, with one playing the role of the *driver* and the other in the role of the remote participant (the *caller*). Drivers were recruited from a mailing list for people who had been previously trained and calibrated for using the driving simulator. Recruited drivers were asked to bring their own partner for the study; otherwise the experimenters matched them up with partners. A total of 18 pairs were recruited. 15 of the 18 pairs were prior acquaintances, including friends, colleagues or family members. Prior acquaintance did not have any effects on the results. Users were compensated with a free software or hardware gratuity. To incentivize appropriate attention on performing the task, an extra $50 was offered to the pair who had the best overall performance in the study.

Task
For each trial, the driver was given a specific destination as the goal and was instructed to drive a 3-4 minute route where road signs would direct them towards the desired destination. Routes were composed of a combination of road segments (*regular)* with light or heavy traffic and had a mixture of signals and stop signs. Additionally, each route contained a number of *critical segments*, where the driver had to pay extra attention to avoid driving misdemeanors. Such segments consisted of road signs indicating turns to the destination, construction sites, pedestrian crossings, residential areas, police cars, and accident scenes.

While piloting the vehicle, the driver answered questions asked by the caller over a speaker phone system (fig. 1). A predefined set of no more than 25 questions were asked in each trial. Questions were designed to resemble daily chit-chat, requiring recall of recent activities and background,

Figure 1: Participant driving the STISIM simulator. The simulator provides a console with speedometer, steering wheel with turn signals, and traditional brake and accelerator. Three 47” screens generate an impression of driving a vehicle. The driver used a speaker phone system (center left) to converse with the remote caller (inset).

e.g. “*When did you last get gas for your car?”* and *“Name the last movie you saw.”* We chose these retrieval-centric questions in light of findings from a previous study showing that drivers have difficulty with driving while being challenged with retrieval tasks [14]. The caller was instructed to ask questions in a predefined order and to write down each answer before moving on to the next one. Repetitions of questions for clarification were allowed.

To provide incentive for allocating attention to both driving and answering questions, participants were informed that their performance scores depended on both driving and the conversations. They were told that they would be scored on the number of questions that they completed within a given amount of time. They were further told that they would be docked points for driving misdemeanors such as accidents, missing stop signs and turn signals, and unnecessarily slowing down or speeding. The caller was placed in a separate room and did not have access to the visual rendering of the driving scene, but the driver could provide relevant information to the caller over the phone to negotiate safe driving while answering questions.

Intervention
Before a critical segment (consisting of a critical event requiring the driver’s attention) was about to start in the route, the conversation would be interrupted with one of the auditory interventions being studied, and heard by both the driver and remote caller. Interventions happened only for critical segments. The intervention was one of two modes (*Mode)*: a short message simply stating “*Focus needed*,” and a longer, more descriptive message stating exactly what to expect, e.g. “*Turn ahead”* or “*Construction ahead.”* For each intervention mode, the call was put on hold (*Hold*) for half of the trials, where, the caller also received a message

stating that their call was put on hold temporarily. The caller was then muted until the critical event passed, followed by a message to resume conversation right after. The driver did not hear the hold message that the caller heard, but both were told at the beginning of the trial whether calls will be muted after receiving an intervention message, so the silence from the caller's end after delivery of the intervention message was expected for the driver. For the other half of cases, the message was delivered, but the conversation was not put on hold. In these cases, it was left to the driver and the caller to negotiate safe driving.

In terms of the timing of the interventions (*Timing*), for half of the trials, the intervention message was delivered immediately, as soon as the beginning of the critical segment could be seen in the driving horizon, typically a few seconds before the vehicle reached the start of the critical segment. This resulted in the intervention message often interrupting the conversation in the middle of asking a question or providing an answer. For the other half, a *bounded deferral* [9] technique was used, where the message would be withheld until the driver had finished answering the question from the caller, and delivered at a conversational breakpoint (before the start of the next question). In cases where the answer did not end until the critical event was about to happen, the message was delivered right at the beginning of the critical segment. The intervention was conducted in a Wizard of Oz setting where one of the experimenters who had full view of the driving scene and could hear the conversation could determine when to deliver the intervention message.

Before each trial, participants were told about the mode (receiving short or long message) and whether or not the call would be put on hold. We did not tell them whether the call would be delivered at a conversational breakpoint to test whether participants would notice this shift in timing, and whether interrupting in the middle of a conversation was considered to be more 'rude' or costly in another way compared to waiting until an answer was completed.

Methodology

Participants arriving at the lab were guided through an informed consent process, followed by an overview of the study. The caller was provided a set of questionnaires that s/he would ask over the phone (one for each trial) and then taken to a separate room where one of the experimenters remained with her throughout the duration of the trials. The communication between the simulator room and the caller room was maintained through an audio conferencing system, also used to simulate a phone conversation.

The participants were then taken through a series of practice trials, the first involving only questions over the phone in the absence of driving, and the second involving driving with no questions. These were followed by two practice trials where driving and question answering were combined, one demonstrating the short message, call on Hold combination, and the other demonstrating the longer message, call not on hold condition. This provided the participants a sense of what to expect during the actual trials. The practice trials could be repeated if needed.

The participants then moved on to the experimental trials. Each pair had a total of 20 trials, though not everyone could complete all trials due to time limitations. Two of the trials were a baseline of driving with no conversations, and two were a baseline of conversations in the absence of driving. The remaining 16 trials occurred according to various intervention conditions. Combinations of factors (Mode, Timing, and Hold) were presented according to a Latin square design. Each combination was presented twice.

At the end of each trial, both the driver and caller answered a short questionnaire on the difficulty of the trial and about the effectiveness and disruptiveness of the intervention. The entire study lasted approximately 2 hours per pair.

Measures

Quantitative data on performance on both driving and conversation was collected. For driving, we collected the number of errors with turning and collisions (for both of which intervention messages provided alerts ahead of time), as a measure of how the intervention messages affected these metrics. To indirectly assess whether an intervention added to the cognitive load of drivers in a manner that may distract drivers and force other driving errors, we collected the number of missed red lights and stop signs, and roadside excursions. These measures were automatically recorded by the simulator. For all collected measures we computed incidence rates (per minute) separately over the span of the critical segments (when critical events occurred), and the span of regular segments (part of the route excluding the critical segments). This was done for both trials with phone calls and the baseline driving in the absence of phone calls.

For performance on the phone conversations, we computed the number of questions answered per minute, again for both trials with phone conversations and driving, and the baseline of phone conversations with no driving.

We also collected qualitative data on participants' perceptions on the intervention experience. We asked them to assess the overall difficulty of each trial, and to rate how they felt about the effectiveness of the intervention in maintaining driving safety, the disruptiveness of the intervention for the task of driving and answering questions, and how rude they felt the system was.

At the end of all trials, participants in both roles answered a final questionnaire on the promise of using the interventions in real life, where they were asked to rank order the nine situations that they had experienced during the study.

Effects of Intervention on Driving Performance

We now review the metrics for driving performance to understand the effects of different types of interventions. Unless otherwise stated, our analysis for each metric conformed to the following steps:

1. Baseline comparison of the metric between regular and critical segments to establish relative difficulty of the critical segments and validation that these segments required extra attention from the driver.
2. For critical segments only, comparison of the metric across three contexts: trials with driving only (baseline control), trials with driving and conversation (control to assess effects on performance due to talking), and trials with driving, conversation, and intervention (to understand intervention effects).
3. For critical segments only on the trials that had interventions, comparison of the metric among different types of intervention along the dimensions of Mode, Hold, and Timing.

Effects of Interventions in Reducing Turning errors

A turning error is characterized as not making a turn when required (as instructed by road signs), making a wrong turn (e.g. making a left turn instead of a right), or turning when not needed. Making turns were required only in the critical segments and an intervention message was delivered either directing the driver's attention to the road (shorter message) or explicitly announcing an upcoming turn (longer message). Turning errors could still occur in regular segments if drivers made a turn at an intersection where they were not supposed to.

For baseline comparison, we ran an independent samples *t*-test for turning errors per minute between critical (M=0.23, S.D=0.48) and regular (M=0.15, S.D.=0.3) segments across all trials. Results showed that drivers had more errors in turning during the critical segments ($t(572)=2.78$, $p<0.006$), which was not unexpected as specific instructions for turning was provided only during the critical segments. This also indicates the need for the driver to pay more attention to the turn instructions during the critical segments. Interestingly, turning errors were nonzero for the regular segments, happening when drivers made turns when not needed, indicating lack of attention on the driver's part.

Looking at only the critical segments (where turns were specifically required), a one-way ANOVA on turning errors /minute showed a significant difference ($F(2, 337)= 35.6$) across driving (M=0.54, S.D.=0.75), driving+ conversation (M=0.73, S.D.=0.83), and driving+conversation +intervention (M=0.14, S.D.=0.3). Post hoc Bonferroni tests showed that turning errors/minute during driving+ conversation+intervention was significantly lower than both driving+conversation ($p<0.0001$) and driving ($p<0.0001$).

To explore the influence of different types of intervention, we conducted a 3-way repeated measures ANOVA analysis on the trials with driving, conversation, and intervention, where Mode, Hold and Timing were the within subject variables, and Instance (1 or 2) was the between subject variable (for each repetition for each combination). Results showed a main effect of Mode ($F(1,22)=8.3$, $p<0.009$) - drivers missed less turns/minute when they received the longer message (M=0.09, S.D.=0.24) compared to the shorter message (M=0.18, S.D.=0.34), see Table 1. The more explicit intervention seemed to be more effective in directing the driver's attention to upcoming turns.

Effects of Interventions in Reducing Collisions

A collision was recorded when the driver hit another vehicle, a pedestrian, or lane markers in construction zones. An intervention message was delivered when there was a possibility of a collision ahead, but collisions could also occur in regular segments. An independent samples *t*-test for number of collisions /minute between critical (M=0.192, S.D.=0.15) and regular (M=0.038, S.D.=0.43) segments across all trials showed that drivers had more collisions during critical segments ($t(420)=6.2$, $p<0.0001$). This again validates the need to pay more attention while traversing a critical segment compared to a regular segment.

Looking only at the critical segments, a one-way ANOVA showed no significant differences in collisions per minute across driving (M=0.13, S.D.=0.34), driving+conversation (M=0.29, S.D.=0.59), and driving+conversation+ intervention (M=0.21, S.D.=0.46), but rate of collisions while conversing were slightly lower with interventions compared to no intervention. We explored whether there were any significant differences for the various intervention conditions. A three-way repeated measures ANOVA with Mode, Hold, and Timing as the within subject variables also did not show significant differences in collisions/minute. As seen in Table 1, the longer message caused less collisions/minute compared to the shorter message, though the differences did not reach significance.

Effects of Interventions in Forcing Roadside excursions

Roadside excursions were recorded when the vehicle deviated from its regular driving path for any reason, often due to lack of control. Unlike for turns and possibility of collisions where interventions provided prior alerts, roadside excursions were measured to understand potential side effects of an intervention, e.g. the prospect that the overhead of messaging would cause roadside excursions.

Driving, Conversation and Intervention		**Turning errors/minute**	**Collisions/ minute**
Factors	**Levels**	Driving Only: 0.54(0.75) Driving+Conv: 0.73(0.83)	Driving Only: 0.13(0.34) Driving+Conv: 0.29(0.59)
Mode	Shorter message	0.181(0.34)	0.263 (0.525)
	Longer message	0.088 (0.24)	0.16(0.366)
Hold	No hold	0.125(0.28)	0.156(0.355)
	Hold	0.149(0.32)	0.272(0.539)
Timing	Immediate	0.136 (0.31)	0.195(0.456)
	Breakpoint	0.138 (0.29)	0.236(0.463)

Table 1. Means (standard deviations in parenthesis) of rate of turning errors and collisions (per minute) across various levels of Mode, Hold, and Timing. Baselines are provided for comparison.

An independent samples *t*-test showed a higher incidence of roadside excursions per minute during critical (M=0.13, S.D.=0.37) compared to regular (M=0.07, S.D.=0.29) segments, $t(643)=2.33$, $p<0.03$, providing further evidence that managing driving safety during critical segments was more difficult than during regular segments. A one way ANOVA among driving, driving+conversation, and driving+ conversation+intervention showed no significant differences in roadside excursions/min, indicating that the intervention did not affect this measure negatively compared to only driving, or driving and conversing.

We explored whether one type of intervention had a more negative effect than others in the critical segments. A three-way repeated measures ANOVA with Mode, Hold, and Timing as within subjects variables showed an interaction effect between Mode and Hold ($F(1,22)=5.319$, $p<0.031$). Following up on the interaction effect between Mode and Hold, post hoc tests showed that for calls put on hold, longer messages caused more roadside excursions (M=0.27, S.D.=0.61) compared to shorter messages (M=0.08, S.D.=0.29). This suggests that the additional load of processing the longer message may cause drivers to swerve even after putting the conversation on hold.

Effects of Interventions on Red Light/Stop Sign Misses

As with roadside excursions, we measured the rate of red light and stop sign misses to assess if the intervention caused additional load inhibiting drivers' visual processing. An independent-samples *t*-test between critical (M=0.06, S.D.=0.23) and regular (M=0.12, S.D.=0.26) segments across all trials showed that drivers tend to miss red lights and stop signs less during critical segments ($t(675)=2.8$, $p<0.006$), which may be because drivers tend to drive slower during these segments.

Looking only at critical segments, a one-way ANOVA on Missed Red lights or Stop signs per minute did not show a significant difference across driving only, driving and conversation, and driving, conversation and intervention. A three-way repeated measures ANOVA also did not find differences in the effects of the different types of intervention. This may suggest that processing the intervention does not interfere with the perceptual task of recognizing and responding to red lights and stop signs.

Effects of Intervention on Conversation

To measure performance on the conversation, we examined the number of questions that the pair answered in a given amount of time. As with driving performance, we first compared answers per minute across regular and critical segments, then compared critical segments across conversation, conversation+driving, and conversation+ driving+ intervention. Finally, we studied effects of Mode, Hold, and Timing on the number of questions answered.

An independent samples *t*-test showed that users answered more questions per min during regular segments (M=7.53, S.D.=2.01) compared to critical segments (M=3.14, S.D.=3.57), $t(534)=19.7$, $p<0.0001$. A one-way ANOVA across only the critical segments showed that user-pairs answered significantly fewer questions per minute ($F(1,304)=58.8$, $p<0.00001$) during driving+conversation+ intervention (M=2.98, S.D.=3.3) compared to driving+ conversation (M=7.7, S.D.=3.29). This is to be expected, as for some of the trials with intervention, the call was put on hold, substantially affecting the number of questions that could be answered for that trial.

To understand the effect of the factors Mode, Hold, and Timing on the number of questions answered per minute, we ran a 3-way repeated measures ANOVA. Results showed a significant main effect of the Hold condition ($F(1,22)=431.6$ $p<0.0001$). Post Hoc tests showed that pairs answered more questions during the NoHold trials (M=5.9, S.D.=2.2) compared to the 0 during the Hold trials, which again is not unexpected.

Prior work has shown speech to slow down during dual-task situations in driving [14]. We explored whether the third task of intervention further slowed down answering questions, indicating additional load. We ran a one-way ANOVA on the rate of answering questions/minute comparing only the no hold-trials during the critical segments within driving+conversation+intervention, to the no-hold trials during driving +conversation (hold trials were excluded as no conversation took place for those). Results showed that the pairs answered significantly fewer ($F(1, 167)=13.06$, $p<0.0001$) questions during the driving+ conversation+ intervention (M=5.9, S.D.=2.11) compared to the driving+ conversation (M=7.6, S.D.=3.3) trials. This shows that the intervention does indeed slow down the conversation (while it did not necessarily affect driving performance), but with the added benefit of reducing driving errors, as shown in the previous section.

Subjective Feedback on Interventions

To get a sense of how users felt about the intervention experience, we asked them to rate their experiences on a variety of dimensions (see below) after each trial on a Likert scale of 1 to 5. Feedback was collected from both drivers and callers using role-specific questionnaires. However, some questions on the questionnaires were the same for both drivers and callers, allowing us to compare how the role affected perception. For analysis of the questions common to both roles, we added a fourth factor, Role, in addition to Mode, Hold and Timing.

Difficulty of trials

Both drivers and callers were asked to rate the difficulty of the trials on a Likert scale of 1 (very easy) to 5 (very difficult). A univariate ANOVA with Role (driver, caller) and Trial (driving only, conversation only, driving+ conversation and driving+conversation+intervention) as factors showed no differences in the ratings across Role; both driver ratings (M=2.06, S.D.=0.86) and caller ratings (M=2.04, S.D.=1.01) being on the easier side (3=neutral). A main effect of Trial ($F(3, 679)=25.74$, $p<0.0001$) was found. Post hoc Bonferroni tests showed that trials with driving, conversations, and interventions were rated to be significantly more difficult (M=2.21, S.D.=0.91), compared to driving only (M=1.46, S.D.=0.74), conversation only (M=1.39, S.D.=0.82) and driving+conversation (M=1.73, S.D.=0.84), p<0.0001 for all. This suggests that both drivers and callers perceived the intervention to add to the difficulty of the trials in terms of the shared goal (drive safely and answer questions), even though performance results suggested tradeoffs only in the conversations.

For only trials with driving, conversation, and intervention, we ran a repeated measures ANOVA with 4 within subjects factors (Mode, Hold, Timing and Role) on the difficulty ratings to understand effects of the various intervention conditions. Only a main effect of Hold (F(1,17)=11.4, p<0.004) was found. Trials where the conversation was put on hold were rated to be more difficult (M=2.29, S.D.=0.9) compared to trials where the conversation was not put on hold (M=2.02, S.D.=0.9). This suggests that difficulty ratings were being impacted by whether or not the conversation was being inhibited, which would impact the broader goal of overall performance on the trial.

Feedback on the intervention

There were three questions about the intervention that we asked both drivers and callers: whether they felt the intervention made driving safer, how disruptive the intervention was to the conversation and perceived 'rudeness' of the system (Figure 3). A Likert scale of 1 to 5 was used, 1 being strongly disagree, 5 being strongly agree.

On the question of whether the particular type of intervention for the trial made driving safer, we ran a four way repeated measures ANOVA with Mode, Hold, Timing, and Role as within subjects factors and Instance (1 or 2) as the between subjects factor. Results showed a main effect of Role (F(1,13)=18.8, p<0.001) and an interaction effect between Mode and Timing (F(1,13)=10.156, p<0.007). For the main effect of Role, drivers more strongly agreed (M=4.462, S.D.=0.84) that the intervention made driving safer, compared to the closer to neutral ratings of the callers (M=3.42, S.D.=0.97), see Figure 3. This difference may be explained by answers to a related question asked to callers only on whether they felt that they had enough information about the driver's state. Callers were again close to neutral on this point (M=3.52, S.D.=1.1), and a number of callers stated that since they did not have a view of the driver's environment (even though they could hear what was going on) it was difficult for them to fully evaluate safety.

On the question of whether the intervention disrupted conversation, a four-way repeated measures ANOVA showed a main effect of both Role (F(1,14)=12,19,p<0.004) and Hold (F(1,14)=34.713, p<0.0001), as well as an interaction effect between Role and Hold (F(1,14)=13.87,p<0.002), and between Mode and Timing (F(1,14)=7.76, p<0.015). Overall, callers more strongly agreed that the intervention disrupted the conversation (M=4.01), compared to the more neutral ratings of the drivers (M=2.98), see figure 3. Looking at the interaction

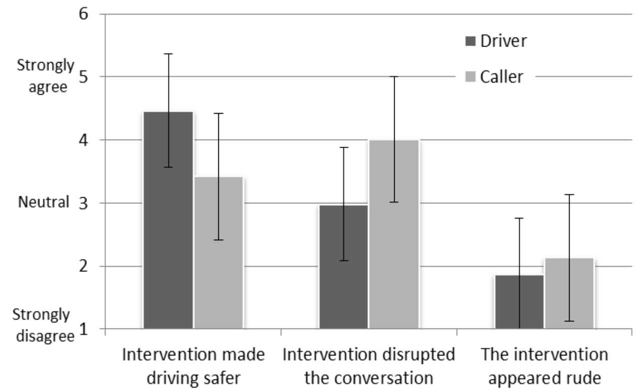

Figure 3: Comparison of driver and caller rating on driving safety, conversation disruption, and social attribution towards the intervention.

effect between Role and Hold, for trials where conversations were put on hold, callers had a higher rating of disruption (M=4.53, S.D.=0.62) compared to trials where conversations were not put on hold (M=3.48, S.D.=1.3, p<0.0001). The drivers' ratings of disruption of conversation did not vary significantly across the Hold condition. This again supports the observation that even in a cooperative environment with a broader common goal, individual roles of participants in a team can affect their judgments. For the interaction effect between Mode and Timing, the shorter message delivered immediately were considered to be more disruptive to the conversation (M=3.58, S.D.=1.23) compared to the longer messages delivered immediately (M=3.22, S.D.=1.28, p<0.049).

On the question of the perceived 'rudeness' of the intervention, the overall rating was low (M=1.988, S.D.=0.99) with drivers having a mean rating of 1.86 and callers having a mean rating of 2.11), see Figure 3. This suggests that the interventions were not deemed as unacceptable in terms of social norms. A repeated measures ANOVA showed no effect of Mode, Hold, Timing, or Role.

Drivers were also asked additional questions on specific effects of the intervention on the driving experience. On the question of whether they felt the intervention disrupted driving, they generally disagreed (M=1.9, S.D.=1.1) with no differences across Mode, Hold, or Timing. Drivers also agreed that the system provided the right amount of information (M=4.1, S.D.=0.9), but again were not biased towards any particular combination of the three factors. On a question of whether putting the conversation on hold helped, drivers leaned towards agreeing that it was useful (M=3.6, S.D.=1.2), but did not again show preference for any particular combination of factors.

Overall preferences

At the end of the study, both drivers and callers were asked about their overall opinion on the effectiveness of an intervention system that can help manage phone calls based on driving situations (Figure 4). Drivers were generally positive about the usefulness of such a system (17/18

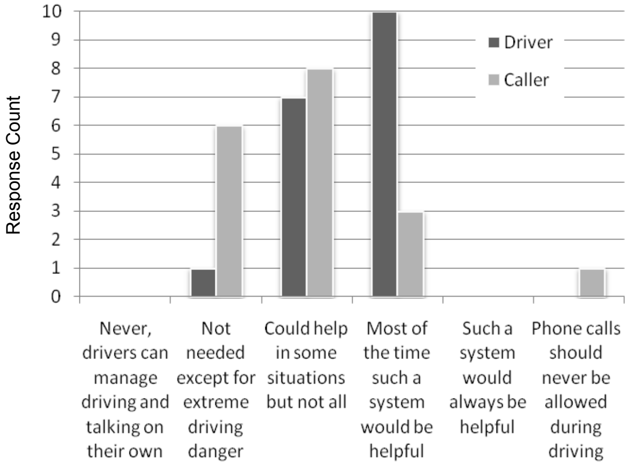

Figure 4: Histogram of preferences about intervention system for drivers and callers in the experiment. Majority of drivers were in favor of such a system, while majority of remote callers felt that such systems should be used occasionally at the best.

preferring such a system occasionally or most of the time). Interestingly, callers were less enthusiastic (14/18 preferring such a system occasionally, or only for extreme cases). This again suggests the need for callers to be more aware of effects of conversations on driving.

We also asked both drivers and callers to rank order their preferences for the various combinations of factors (Figure 5). Drivers mostly prefered the short message delivered at a breakpoint with the call put on hold, and remote callers mostly prefered the shorter message delivered at a breakpoint with the call not put on hold. Both drivers and callers least preferred the longer message that interrupted in the middle with the call on hold. No intervention was ranked 8 out of 9 for drivers, but was ranked 3 out of 9 for callers, indicating that drivers, through their more direct experience with the intervention during the study, perceived the benefits on driving more than their remote partners, who only experienced the effects on the conversation.

Drivers and callers also provided free form feedback on how they perceived the value of intervention. One driver commented: "*The prompts to pay attention were surprisingly useful. Wish I had them in my current auto system.*" Another driver commented: "*Announcement of turns was helpful.*" Of course, drivers did express their dislikes and preferences for the various forms of intervention. One driver had an initial comment of: "*Being put on hold doesn't help me, makes me wonder what happened…*" and after a few trials he remarked: "*After using it I seem to be getting used to the interruptions!*" A couple of drivers remarked that a visual indicator of when the call was on hold would have been helpful. Note that drivers and callers were informed right before the trial whether the call would be put on hold, but apart from silence from the caller's end after receiving the intervention message, there was no cue to the driver.

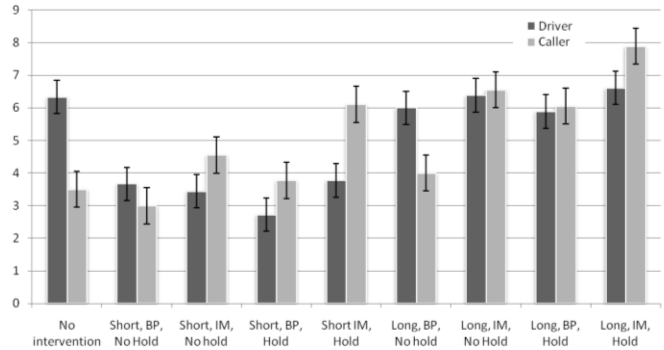

Figure 5. Average ranking for each experimental condition separated by drivers and callers. Lower scores indicate higher preferences (1= most preferred, 9= least preferred, shown on the *y-axis*). BP=breakpoint, IM=immediate.

The remote callers were less enthusiastic about the interventions. One caller commented: "*Not sure if I like my behavior dictated by a computer,*" while another remarked: "*hate being put on hold.*" One person found the hold being "*an annoyance,*" but also commented that "*but it gave me a few seconds to collect my thoughts.*" Another caller also voiced her preference towards the no-hold condition: "*Not putting on hold made the interventions ok to deal with as it is just a short break and the train of thought is not lost.*"

Callers generally prefered the shorter message as exemplified in the following feedback: "*I really like the shorter notifications – focus needed is plenty of info…*" and "*It was nice to have the detail about the driver was focusing – but shorter message is better.*" Overall, both drivers and callers seemed to see value in the interventions, but showed a preference towards being alerted and not necessarily to relinquish deeper control of the conversation to the system.

DISCUSSION

Our study on the effectiveness of various combinations of intervention factors revealed a positive effect of intervention on driving, resulting in lower incidence of turning errors and collisions as compared to trials without interventions. The mode of the intervention appeared to positively influence driving performance, with the more descriptive message being more effective in reducing driving errors. However, the intervention also affected performance on the conversation, and subsequently affected subjective preferences. Qualitatively, drivers showed strong preference towards interventions while callers, who did not directly benefit from interventions, were more neutral about the need for interventions while driving and conversing.

We also examined the effects of the additional cognitive load that interventions might place on a driver. We explored whether processing interventions for the critical events caused higher incidence of other driving errors such as roadside excursion and missing lights and signs. Results did not find interventions to affect these metrics compared to trials with only driving, or with driving and conversations, suggesting that the mediation did not significantly distract drivers. Pairing this with the benefits of improved performance on metrics for which interventions were provided, the positive ratings of drivers are understandable.

We found that interventions influenced conversations. For trials where calls were put on hold, the fluency of the conversation was understandably disrupted. Conversations slowed down during the critical events, even for cases where the call was not put on hold, compared to conversation during regular driving segments. This suggests that while facing a critical event, drivers were trading off performance on the conversation with driving safety, supporting findings in [14]. The shared intervention message provided callers with awareness of upcoming difficulty in driving, enabling them to assist with modulating conversations. In trials with no intervention, callers were unaware of the driving situation unless they were specifically told by the driver.

An overall preference was found towards short messages delivered at breakpoints during conversations, with drivers preferring the conversations to be also put on hold, even though longer messages were found be to be more effective. Callers, who generally found the call being put on hold annoying, could perhaps benefit from more information about the driving situation in addition to the sound bites, e.g. a visual abstraction of the driving scene [26].

Our study used a Wizard of Oz setup to perform an initial investigation of the effectiveness and acceptability of intervention during driving and conversing. To implement such a system in practice, an automated system could harness GPS, accelerometers, and proximity sensors along with information on the planned route, real-time traffic, accidents, detours, weather, etc., as well as statistics of events on upcoming segments to predict future criticality. The cognitive load of drivers could be estimated by considering the current driving situation (e.g. speed, braking, turning, and shifts in the overall complexity of driving) as well as physiological measures (e.g. heart rate [24]). The cognition required to perform the secondary task on top of driving (e.g. certain types of phone conversations may induce more cognitive engagement) might also be estimated. An alerting cost factor, representing a potential deficit of cognitive resources for driving, might be determined by considering the demands of the driving task and the demands of the secondary task. Such a cost factor could be applied to compute the overall expected cost of engaging in both tasks and the cost measure could be used to determine if, when, and how to best intervene.

An intervention system could be realized as an extension of in-vehicle systems equipped with the ability to interrupt and put ongoing calls on hold, or intercept incoming calls and direct them to a voice message system. While some of the technology exists today, further research is needed to determine the most effective form of such a system.

We note that such intervention systems are not a panacea for safe driving; people will likely continue to make driving mistakes regardless of whether they are talking on the phone or not. We expect that an intervention system would be able to help drivers better manage their attention while in a divided-attention scenario (e.g. taking on the role of a vigilant passenger), by directing their attention towards commonly backgrounded tasks of driving (e.g. details of the visual scene). The results on driving performance in our study appear to support this hypothesis, but further research is required to determine how interventions would work in the real world.

Many challenges remain with the design of such intervention systems. What are the most appropriate granularities of events that system should represent and direct the driver's attention to? What events are feasible to detect and which ones are not, and how should events be prioritized? More research is needed to understand how to determine critical upcoming events and tolerance for alerts about these events. Inappropriately low thresholds to alerting about events that would be recognized without notification can lead to situations where users ignore many of the interventions, including ones that may be more important. On the other hand, settings that are too conservative may lead to the suppression of alerts about potentially dangerous events. There are also risks associated with becoming dependent on an intervention system. Drivers might become less attentive towards road conditions expecting the system to alert them whenever a difficult situation is arising. Might such a growing dependence encourage multitasking while driving, with a new reliance on the cushion of being warned when attention needs to be directed towards the road? Future research is needed to address these questions.

Finally, despite having driving experience themselves, the callers in our study appeared overall to be agnostic about the usefulness of interventions, while the drivers in the study were more positive towards it. These differences are likely based in the roles that each of the participants had in the study which may have affected their perception. The results overall highlight the broader need to educate the population about the dangers of driving and multitasking, and awareness about the compromises that participating parties may have to make to ensure the safety of driving.

CONCLUSION AND FUTURE WORK

We conducted a study investigating the effectiveness and acceptability of informative alerting and communication mediation during simultaneous driving and phone conversation. Results showed that intervention with more explicit messages about upcoming critical conditions is effective in reducing turning errors and collisions, compared to situations with no intervention. No effects of additional load on the driver were found. However, the benefits on driving were traded off with a corresponding slowing down of the conversation. Overall, drivers showed preference towards the benefits of interventions, whereas callers tended to be more neutral. Our results provide insights on how different types of intervention during driving can help people drive more safely during conversations. Future work includes development of models of risk for combinations of driving situations and secondary tasks and the study of the engineering of context- and cognition-aware communication mediation systems.

ACKNOWLEGMENTS

We thank Ivan Tashev for his help with the driving simulator and also thank all the participants in our study.

REFERENCES

1. *Cell Phone and Texting Laws*. Dec. 2010, Retrieved from http://ghsa.org/html/stateinfo/laws/cellphone_laws.html
2. Alm, H. and Nilsson, L. The effects of a mobile telephone task on driver behavior in a car following situation. *Accident Analysis and Prevention*, *27* (5). 707-715.
3. Briem, V. and Hedman, L.R. Behavioral effects of mobile telephone use during simulated driving. *Ergonomics*, *38*. 2536-2562.
4. Brookhuis, K.A., De Vries, G. and De Waard, D. The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention*, *23,* 309-316.
5. Brumby, D.P., Salvucci, D.D. and Howes, A. Focus on driving: How cognitive constraints shape the adaptation of strategy while dialing while driving. *Proc CHI*, 2009, 1629-1638.
6. Brumby, D.P., Salvucci, D.D., Mankowski, W. and Howes, A. A cognitive constraint model of the effects of portable music-player use on driver performance. *Proc HFES,* 2007, 1531-1535.
7. Cao, Y., Castronovo, S., Mahr, A. and Muller, C., On Timing and Modality Choice with Local Danger Warnings for Drivers. *Proc. AutomotiveUI*, 2009, 75-78.
8. Horrey, W.J. and Wickens, C.D. Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. *Human Factors*, *48* (1). 196-205.
9. Horvitz, E., Apacible, J. and Subramani, M. Balancing Awareness and Interruption: Investigation of Notification Deferral Policies. Ardissono, L., Brna, P. and Mitrovic, A. eds. *Proc. User Modeling*, 2005, 433-437.
10. Horvitz, E., Kadie, C.M., Paek, T. and Hovel, D. Models of Attention in Computing and Communications: From Principles to Applications. *CACM*, 2003, 52-59.
11. Horvitz, E., Koch, P., Kadie, C.M. and Jacobs, A. Coordinate: Probabilistic Forecasting of Presence and Availability. *Proc UAI*, 2002, 224-233.
12. Iqbal, S.T. and Bailey, B.P. Oasis: A Framework for Linking Notification Delivery to the Perceptual Structure of Goal-Directed Tasks. *ACM Transactions on Computer-Human Interaction*, 2010, *17* (4).
13. Iqbal, S.T. and Bailey, B.P., Understanding and Developing Models for Detecting and Differentiating Breakpoints during Interactive Tasks. *Proc. CHI*, 2007, 697-706.
14. Iqbal, S.T., Ju, Y.-C. and Horvitz, E. Cars, Calls, and Cognition: Investigating Driving and Divided Attention. *Proc. CHI*, 2010, 1281-1290.
15. Lee, J.D., Caven, B., Haake, S. and Brown, T. Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Driver's attention to the Roadway. *Human Factors*, *43*. 631-640.
16. Levy, J. and Pashler, H. Task Prioritisation in Multitasking during Driving: Opportunity to Abort a Concurrent Task Does not Insulate Braking Responses from Dual Task Slowing. *Applied Cognitive Psychology*, *22*. 507-525.
17. Mahr, A., Pentcheva, M. and Muller, C., Towards System-Mediated Car Passenger Communication. *Proc. AutomotiveUI*, 2009, 79-80.
18. Manalavan, P., Samar, A., Schneider, M., Kiesler, S.B. and Siewiorek, D.P. In-car Cell Phone Use: Mitigating Risk by Signaling Remote Callers *CHI Extended Abstracts*, 2002, 790-791.
19. McFarlane, D.C. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human Computer Interaction*, *17* (1). 63-139.
20. Meschtscherjakov, A., Wilfinger, D., Scherndl, T. and Tscheligi, M., Acceptance of Future Persuasive In-Car Interfaces Towards a more Economic Driving Behavior. *Proc AutomotiveUI*, 2009, 81-88.
21. Nunes, L.M. and Recarte, M.A. Cognitive Demands of hands-free-phone conversation while driving. *Transportation research Part F*, *5*. 133-144.
22. Redelmeier, D.A. and Tibshirani, R.J. Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, *336* (7). 453-458.
23. Reed, N. and Robbins, R. The Effect of Text Messaging on Driver Behavior: A Simulator Study *Transport Research Lab Report*, 2008, PPR 367.
24. Riener, A., Ferscha, A. and Aly, M., Heart on the road: HRV analysis for monitoring a driver's affective state. *Proc. AutomotiveUI* 2009, 99-106.
25. Salvucci, D.D., Markley, D., Zuber, M. and Brumby, D.P. iPod distraction: effects of portable music-player use on driver performance *Proc. CHI* , 2007, 243-250.
26. Schneider, M. and Kiesler, S. Calling while driving: effects of providing remote traffic context. *Proc CHI*, 2005 561-569
27. Strayer, D.L. and Drews, F.A. Cell-Phone Induced Driver Distraction. *Current Directions in Psycological Science*, *16* (3). 128-131.
28. Strayer, D.L. and Johnston, W.A. Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. *Psychological Science*, *12*. 462-466.
29. Wickens, C.D. Processing Resources and Attention. in Parsuraman, R. and Davies, R. eds. *Varieties of Attention*, Academic Press, New York, 1984.