Fast or Safe? How Performance Objectives Determine Modality Output Choices while Interacting on the Move

Duncan P. Brumby, Samantha C.E. Davies, Christian P. Janssen, and Justin J. Grace

UCL Interaction Centre
University College London
London WC1E 6BT UK
Brumby@cs.ucl.ac.uk

ABSTRACT

In-car devices that use audio output have been shown to be less distracting than traditional graphical user interfaces, but can be cumbersome and slow to use. In this paper, we report an experiment that demonstrates how these performance characteristics impact whether people will elect to use an audio interface in a multitasking situation. While steering a simulated vehicle, participants had to locate a source of information in a short passage of text. The text was presented either on a visual interface, or using a text-to-speech audio interface. The relative importance of each task was varied. A no-choice/choice paradigm was used in which participants first gained experience with each of the two interfaces, before being given a choice on which interface to use on later trials. The characteristics of the interaction with the interfaces, as measured in the no-choice phase, and the relative importance of each task, had an impact on which output modality was chosen in the choice phase. Participants that prioritized the secondary task tended to select the (faster yet more distracting) visual interface over the audio interface, and as a result had poorer lane keeping performance. This work demonstrates how a user's task objective will influence modality choices with multimodal devices in multitask environments.

Author Keywords

Multitasking, visual interface, audio interface, driving, strategy selection, performance trade-offs.

ACM Classification Keywords

H.1.2 User/Machine Systems: human factors; H.5.2 User Interfaces: evaluation/methodology, interaction styles.

General Terms

Experimentation, Human Factors.

INTRODUCTION

In recent years there has been a steady flow of reports on people's propensity to use their mobile devices while driving – some even feel the need to keep their Twitter stream up-to-date from behind the wheel [5]. There is a large body of work showing that this kind of behavior is reckless and can increase the risk of a crash. For example, Horrey and Wickens [11] offer a meta-analysis of 23 different studies investigating mobile phone use while driving, each of which demonstrates negative effects on driving performance. Because mobile devices support so many of our daily activities, efforts to discourage drivers from using their devices have often failed to have an impact [6]. Hence, the HCI community might consider alternative design solutions to lessen some of these problems.

Audio interfaces offer a promising alternative to the traditional Graphical User Interface (GUI) for alleviating some of the deleterious problems associated with using a mobile device while driving. This is because audio output does not demand the user's visual attention and so limits perceptual interference between tasks [27,33]. Audio is already used extensively in many systems to signal events to users (e.g., that a file has finished downloading or that a new message has been received in an ongoing chat). Recently, the use of audio has been taken a step further in the development of *eyes-free* interfaces [21,34,35]. Studies have shown that using audio in this way is a safer alternative to a visual interface [19,20,22,29,34], though the potential for cognitive distraction remains [11,26]. Despite these benefits, task times using audio interfaces can be slower than with traditional GUIs [29,34].

There has been little discussion of whether such trade-offs between speed and safety across different output modalities on a multimodal device might present a problem. We question whether drivers might forfeit the safety benefit offered by an audio interface when they can access the same information more quickly by glancing at the device's screen. Despite the importance of this question to the HCI community, research has not considered whether performance objectives have a strong influence on modality choices with multimodal devices. Instead, previous work has focused on evaluating whether one output modality interferes more or less with driving performance than

another [12,19,20,25,29,34], or in which environment particular output modalities work best [10].

In this paper, we investigate how performance objectives influence modality selection in a multitasking situation. In our experiment, participants were required to find the answer to a question in a short passage of text while controlling a simulated vehicle. This scenario resembles a situation in which someone is trying to read a text message while driving – a reckless activity, but alas one that is all too prevalent [6]. The text for the secondary task was presented either on a secondary display (visual interface) or was spoken by a text-to-speech program (audio interface). The driving task that we used was quite simple, in that, participants were only required to control the steering of the vehicle. Our steering task does therefore not allow us to investigate the full complexity of driver behavior [4,11], but lane keeping performance in a simulator can serve as a proxy for driver safety in the read-world [24].

We were primarily concerned with which output modality people will *choose* to interact with to complete a secondary task. To address this question we used a no-choice/choice paradigm [15,23,28,32]. This paradigm can be very useful for understanding factors that influence strategy selection. Alas its has been largely neglected in HCI research, though it was used in early studies to understand how users decide between different input techniques, namely, the mouse and the keyboard when on a desktop computer [23]. In the paradigm, participants first complete a series of trials using each of the available methods to complete the task at hand before being given the choice over which they would like to use. The benefits of this approach for investigating strategy selection are three-fold: (1) it gives an objective measure of the performance trade-offs at stake between different methods for achieving the same goal, (2) the user gains equal amounts of experience with each method (controlling for experience), and (3) it allows us to find out which option people choose and whether this choice varies under different conditions. This approach was useful here because it allowed us to move beyond the question that has occupied previous work, namely, whether in-car systems that use audio output are less distracting than interacting with a traditional GUI [12,19,20,25,29,34]. Instead, we focus on which output modality users will elect to use and whether this decision is sensitive to changes in the relative priority given to one task over another.

In the no-choice phase of the experiment, we expect to find that when the visual interface is used for the secondary task lane keeping performance will decline. This is because participants will have to shift their visual attention away from the road ahead to glance at the secondary display, and this will impact their ability to steer the vehicle [31]. In contrast, we expect the audio interface to be far less disruptive to steering performance because there is no competition for visual attention between the tasks [27,33].

In terms of time to complete the secondary task in the no-choice phase, we expect the audio interface to be slower because the audio output rate will be slower than a typical reading speed and participants will not be able to vary the rate at which the audio stream is outputted. With the visual interface, not only will participants be able to rapidly skim the content of the text to locate the target rapidly [7], but they will also be free to vary their strategy for interleaving attention between the two tasks [2,3,14,16-18]. In other words, participants can make longer or shorter glances to the secondary display before returning their attention to the road ahead. We do not investigate these fine-grained moment-to-moment interleaving patterns in detail here, as this has been studied extensively in previous work [2,3,14,16-18]. Instead, we use the no-choice phase to get a reliable measure of the performance trade-offs at stake between each option before giving participants a choice of output modality in the choice phase.

How might participants decide between these different modalities in the choice phase? It is well known that in single-task situations people tend to opt for the fastest method, even when it is only milliseconds faster [8]. Therefore we might expect participants to show a strong preference for the interface that allows them to complete the secondary task the fastest. Such a preference would mean that all participants might forfeit the safety benefits of using the audio output in order to complete the secondary task quickly. In contrast, the decision over which output modality to use might also be moderated by the users' *performance objective* (i.e., which of the two tasks is subjectively more important to them at that time).

To address the question of whether varying the relative value (or priority) of each task affects which modality is chosen, different instructions were given to participants. For half of the participants, the instructions emphasized that they should prioritize rapid completion of the secondary task, for the other half the instructions emphasized that they should prioritize safer steering performance. This manipulation reflects the idea that in many real world multitasking scenarios, a user might value one task more than another depending on what their goal is. For example, even a careful driver who usually avoids using their phone while driving might, on occasion, place a quick call if they are running late for a very important meeting. We therefore expect participants seeking to prioritize steering to forsake time and choose the audio interface. In contrast, participants seeking to prioritize the secondary task might forsake steering and choose the visual interface.

In summary, the aim of the study reported here is to investigate how people choose between different methods for completing a task in a dynamic multitask setting, and how these choices are affected by changes in the relative importance of one task relative to another. The value of this work is that it leads to a better understanding of how people interact with devices while multitasking, given their performance objectives and their experience with different

output modalities. The expected trade-offs in modality choice have not been studied in depth until now. However, a better understanding is useful, particularly given the recent interest in the HCI community for multimodal interfaces that give feedback across multiple modalities at the same time [10,34,35] – essentially giving the user a choice over which modality to attend to. We consider how people might resolve such choices while driving a simulated vehicle.

METHOD

Participants
Twenty-four participants (10 female) took part in the study. Participants were unpaid volunteers, aged between 20- and 38-years ($M = 27.0$ years). All had a valid driver's license and at least two years of driving experience ($M = 7.5$ years). All the participants were highly proficient at reading and speaking in English (i.e., to the standard set by the University's admissions criteria), and most (20 out of 24) were native English speakers.

Materials
A dual-task setup was used in which the participant had to perform a secondary information look-up task while steering a simulated vehicle. Figure 1 shows a schematic representation of how the task equipment was arranged relative to the participant. The driving simulation environment was displayed on a 30-inch monitor and controlled by a Logitech G25 Racing Wheel. The secondary task used two output modalities (audio and visual). Audio output was projected using standard desktop speakers positioned behind the participant, while visual output was presented on a 7-inch secondary display positioned to the left of the main driving task display.

The driving task that was used required participants to navigate the center lane of a three-way highway environment. Participants were only required to steer the vehicle and maintain a central lane position, while the vehicle's speed was held at a constant 55 mph (88.5 km/h). We refer to this as the steering task. To reinforce safe lane keeping, safety cones were placed at either side of the driver's central lane. A lead vehicle was placed at a fixed distance in front of the participant's vehicle. Noise was added to the vehicle dynamics, causing the vehicle to gradually drift about in the lane. This meant that the participant had to actively control and monitor the vehicle's lateral position and heading to maintain a central lane position. This driving set-up is identical to the one used in other driver distraction studies [3,17,26].

The secondary task was designed to resemble a scenario in which a user is trying to locate a piece of information in a short passage of text, such as is common in e-mail and web browsing activities. Thirty different excerpts of text were sampled from various websites (e.g., www.wikipedia.org, www.bbc.co.uk) covering topics ranging from major historical events and figures, to popular culture. Each excerpt was edited so that it was approximately 150-180 words in length. For each text, two questions were set that could be easily answered based on the information presented within them. Care was taken to ensure that the location of the target information was balanced across texts. In particular, for one of the questions the target information appeared in the top half of the text and for the other question the target appeared in the bottom half of the text.

The text for the secondary task could be presented using either visual or audio output. For the visual output, text was presented on a 7-inch secondary display positioned to the left of the main driving task display in font Courier bold, size 18. This meant that approximately eight words (or 45 characters) were displayed per line. Words were not broken between lines. For the audio output, auditory information was generated using SmartRead text-to-speech software (http://www.smartysoft.com/index.php). Audio output was projected using standard desktop speakers positioned behind the participant. SmartRead was left at its default reading speed of approximately 167 words per minute (2.8 words per second). This meant that listening to a complete passage of text took approximately one minute.

The NASA-TLX questionnaire [9] was used to assess the subjective workload associated with using the audio and the visual interface for the secondary task. The NASA-TLX is a multi-dimensional questionnaire that derives an overall workload score for the task at hand based on the sum of responses to six subscales (mental demand, physical demand, temporal demand, performance, effort, frustration). The higher the NASA-TLX score, the more demanding the person believed the task to be.

Design
The experiment used a no-choice/choice design [15,23,28,32]. In the no-choice phase of the experiment, a 2x2 (modality x task priority) mixed factorial design was used. For the manipulation of modality, all participants had to complete a series of dual-task trials using the audio interface and the visual interface for the secondary task. Task priority was manipulated as a between-subjects factor, in that participants were instructed to either focus on completing the secondary search task as quickly as possible (the searching-focus condition) or to concentrate on keeping the car as close as possible to lane center (the steering-focus condition). There were three purposes for the no-choice phase of the experiment. First, it allowed us to get an objective measure of dual-task performance in each condition (i.e., for each interface x priority combination). Second, it allows us to find out which option people choose in a particular setting and how this might vary as a function of the instructed performance objective. Third, it gave all participants equal practice and experience at completing the secondary task with both of the available output modalities.

In the choice phase of the experiment, participants completed a second series of dual-task trials in which they were free to choose which of the two output modalities they

wished to use. For this part of the experiment, a single between-subject factor design was used that manipulated task priority. Consistent with the no-choice phase, the same participants were again instructed to either prioritize completing the secondary task as quickly as possible or to prioritize keeping the car as close to the center of the middle lane as possible. The purpose of the choice phase of the experiment was to see whether changing the relative importance of each task had an impact on participants' preference for which output modality they used to access the information needed to complete the secondary task while steering.

For both phases of the experiment, the main dependent variables were the time taken to complete the secondary task and the impact that completing this task had on steering performance. Steering performance was indexed as the average distance that the vehicle drifted from lane center while the participant was working on the secondary task. The driving simulator logged the lateral distance of the vehicle at a rate of 200 Hz, and we report the root mean square error (RMSE) of these lateral deviation samples. Finally, we used the NASA-TLX questionnaire to get a measure of the subject workload associated with each condition in the no-choice phase of the experiment.

Procedure
Participants were informed that they would be required to find answers to questions in passages of text (as presented in audio or visual format) while steering a simulated vehicle. Participants were randomly assigned to one of the task priority conditions, though an effort was made to balance the distribution of gender and non-native English speakers across conditions. Participants in the steering-focus condition were told to prioritize the steering task and keep the vehicle as close to the lane center as possible while still completing the secondary task, whereas participants in the searching-focus condition were told to prioritize the secondary task and find the correct answer as quickly as possible while still maintaining lateral control over the vehicle. To emphasize these instructions, clear performance feedback on the relevant performance metric was given at the end of every dual-task trial. The whole experiment took about an hour to complete.

After receiving these instructions, participants were given an opportunity to practice each of the tasks separately. For the steering task, participants completed 10 practice trials in which they drove for a period of 30-seconds along a straight road. Participants were instructed to keep the vehicle as close to the center of the lane as possible. At the end of each trial, they received feedback on their steering performance (as RMSE lateral deviation). For the secondary task, participants completed two practice trials in which they had to find the answer to a question in a passage of text (as presented in audio or text format). Participants were instructed to inform the experimenter once they were confident that they had located the target information.
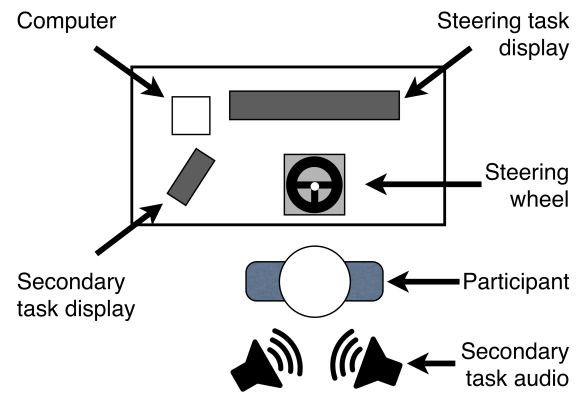

Figure 1. A schematic representation of how the displays for the steering and secondary task were arranged.

Participants were instructed to respond to the question as soon as they found the answer, either by reading it on the display, or by hearing it in the audio file. They indicated their intention to answer by saying, "got it" to the experimenter. At this time, the experimenter logged the event in the experimental software by pressing a key. After logging the event, the participant gave their answer and the experimenter coded this response as correct or incorrect. If an incorrect answer was given, then the trial was started over again until the correct answer was given. This was done to encourage participants not to give a fast (guessed) response for the secondary task, as repeating the trial took additional time. At the end of each trial, participants received feedback on task time.

Once familiar with the tasks, participants completed the no-choice phase of the experiment before completing the choice phase. For the no-choice phase, participants completed four dual-task trials using each of the two output modalities. The order in which each modality was experienced was randomized and counter-balanced across participants. For each dual-task trial, a pre-cue was shown on the driving display indicating which output modality was going to be used for the secondary task. The target question was then shown on the main driving task display, and participants were instructed to read the question out loud and inform the experimenter when they were ready to begin. Once ready, the driving environment was shown and when the vehicle had reached full-speed the participant was informed that they could start the secondary task. At this time, the audio output started playing or the passage of text was presented on the secondary display, depending on which modality was being used for the secondary task. A countdown on the main driving task indicated when the secondary task could be started. A trial ended when the participant told the experimenter the correct answer for the question. If an incorrect answer was given, then the trial was started over again until the correct answer was given. In addition, a trial was also repeated if no response was given within 100 seconds. When the correct answer was given, participants received feedback based on which condition they were in: participants in the steering-focus

condition received feedback about the vehicle's RMSE lateral deviation, whereas participants in the searching-focus condition received feedback on the time taken to complete the secondary task. After completing all four trials with a given modality, participants completed the NASA-TLX questionnaire.

In the choice phase of the experiment, participants completed a block of nine dual-task trials. The structure of each trial was the same as that outlined above, except that participants could choose which output modality they wished to use when the question for the next trial was presented to them. Participants made their choice by saying either "audio" or "visual" to the experimenter, who then made the appropriate input command for the program controlling the experiment. If a trial had to be repeated, participants were again free to choose which modality they wished to use on each subsequent attempt.

Throughout the experiment steering performance was assessed. In between each block of dual-task trials participants completed two single-task steering-only trials, which were similar to the practice session outlined above. This allowed us to determine whether there were any systematic differences in steering ability between participants in each of the two experimental conditions, and to get a measure of base-line steering performance.

Finally, it should be noted that across the experiment each passage of text was only used for a single question. The question that was used with each passage was randomized for each participant with the constraint that target location was balanced across trials (i.e., the target appeared in the top-half of the text for half of the trials and the bottom-half for the other half of the trials). Moreover, participants were not given any information at the start of a trial that might help them guess the location of the target information in the passage of text.

RESULTS

We only consider data from trials in which the correct answer for the secondary task was given on the first attempt. Participants gave on average very few incorrect responses ($M = 1.42$ trials, $SD = 1.32$ trials, *range* 0-4), and in total, data from only 34 trials (8.33%) were excluded. Moreover, a correct response was always given on the second attempt at a question, meaning that no participant made more than one consecutive error on a given trial. Although errors are interesting, no solid conclusions can be drawn here given the low frequency of errors. We therefore excluded error trials from further statistical analysis.

In the following, we report separate analyses for the no-choice and choice phases of the experiment. For all statistical analyses, effects were judged significant if they reached a .05 significance level. First though, we briefly consider single-task steering performance to determine whether participants in either of the different task priority conditions were better at the steering task. For statistical analysis we use an independent samples *t*-test. As can be seen in Figure 2, there was no difference in participants' single-task baseline steering performance across the two conditions, $t < 1$; in both cases, RMSE lateral deviation was approximately 0.27 m from the lane center. This value is consistent with previous driving simulator studies [3,17,26].

No-Choice Trials
We first consider participants' performance when they were forced to use either the audio or the visual interface for the secondary task. We wanted to know if any interface was less distracting, and whether instructions on how to prioritize the tasks would lead to different dual-task performance trade-offs. We also use this no-choice phase to get a sense of participants' subjective assessment of using each output modality for the secondary task. For statistical analysis of these data, a 2x2 (modality x task priority) mixed factorial Analysis of Variance (ANOVA) was used.

Steering performance
Figure 2 shows the RMSE lateral deviation for each of the different conditions. For the no-choice data, it can be seen that steering performance was better (i.e., there was lower RMSE lateral deviation) when audio output was used for the secondary task than when visual output was used. It can also be seen that when audio was used steering performance was unaffected by task priority condition, and that steering performance was more or less equivalent to that observed in the single-task baseline condition. In contrast, when participants were forced to use the visual output and had to read the text on the secondary display, steering performance declined noticeably, in that the vehicle drifted further from the lane center. However, the extent to which the vehicle drifted was clearly moderated by which task the participant was choosing to prioritize: participants that had been instructed to prioritize the steering task maintained better lateral control of the vehicle (though still notably worse than when the audio interface was used).

Statistical analyses support these observations. There were main effects on steering performance of both modality type, $F(1, 22) = 73.28$, $p < .001$, $\eta_p^2 = .77$, and priority condition, $F(1, 22) = 7.06$, $p < .01$, $\eta_p^2 = .24$. Critically though the interaction was significant, $F(1, 22) = 12.06$, $p < .005$, $\eta_p^2 = .35$. Follow up tests of simple main effects found that there was no effect of task priority when participants used the audio interface, $F < 1$. There was however an effect of task priority when participants used the visual interface, $F(1, 22) = 9.27$, $p < .01$, $\eta_p^2 = .30$.

Task time performance
Figure 3 shows the mean time to complete the secondary task for each of the different conditions. For the no-choice data, it can be seen that trial times were slower, on average, when participants were forced to use the audio interface than when the visual interface was used for the secondary task. Interestingly, task times using the visual interface were clearly moderated by the participant's task priority:
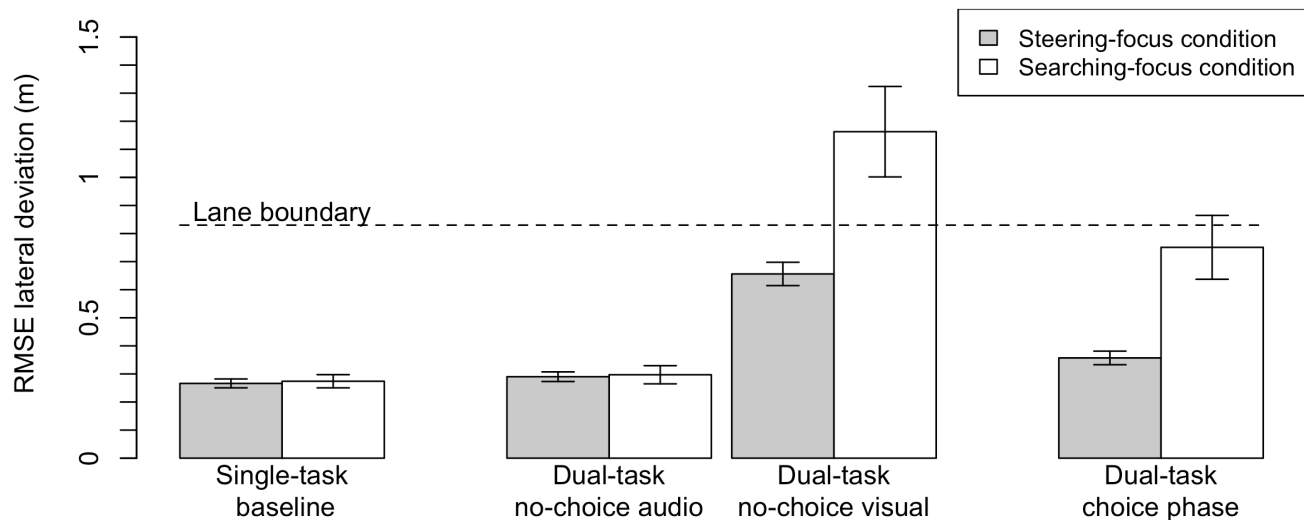
Figure 2. Steering performance for single-task baseline trials, dual-task no-choice trials with each output modality, and choice trials. Error bars depict standard errors.
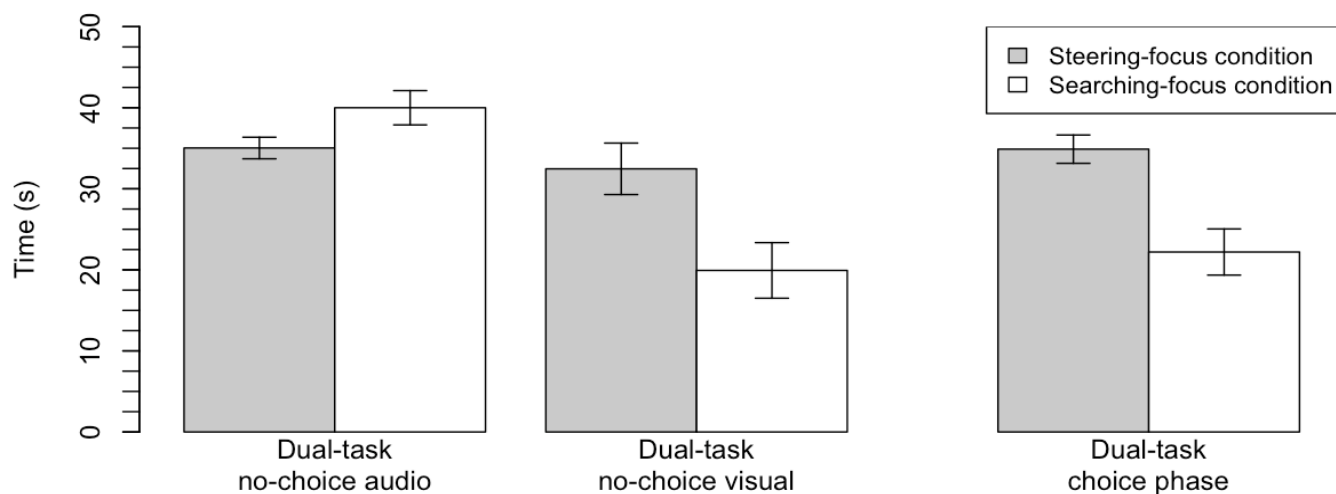

Figure 3. Task time for dual-task no-choice trials with each output modality, and choice trials. Error bars depict standard errors.

participants that had been instructed to prioritize the steering task were slower than those that had been instructed to prioritize the secondary task. There was no such influence of task priority when the audio interface was used – in fact, it is not really conceivable to think how there might have been given that the participant could not actively speed up or slow down the audio output stream.

Statistical analyses support these observations. Overall, there was a main effect of interface modality on task time, $F(1, 22) = 19.27$, $p < .001$, $\eta_p^2 = .47$. While there was no significant main effect of task priority, $F(1, 22) = 1.94$, $p = .18$, $\eta_p^2 = .08$, there was however a significant interaction between task priority and modality, $F(1, 22) = 11.51$, $p < .005$, $\eta_p^2 = .34$. Follow up tests of this interaction reflect what can patently be seen in the figure: there is a

clear effect of priority on task time when the visual interface was used, $F(1, 22) = 7.20$, $p < .01$, $\eta_p^2 = .25$. But there is only a marginal effect of priority on task time when the audio interface was used, $F(1, 22) = 3.95$, $p = .06$, $\eta_p^2 = .15$. The astute reader will have noticed from the figure that this trend actually points in the opposite directions to what is expected: participants in the steering-focus condition were marginally faster than participants that had been instructed to complete the secondary task as quickly as possible. This would appear to be a fortuitous result, and in our estimation it is unlikely to replicate.

Subjective workload
The above measures provide an objective measure of participants' performance. But how was their subjective experience? To this end, NASA-TLX scores were analyzed

to determine the subjective workload associated with both the audio and the visual interface for the secondary task. Results show that participants rated the audio interface as less demanding to use ($M = 45.38$, $SD = 15.99$) than the visual interface ($M = 73.04$, $SD = 15.32$), $F(1, 22) = 59.32$, $p < .001$, $\eta_p^2 = .73$. Moreover, this significant difference between conditions was reflected for each of the NASA-TLX subscales (i.e., mental demand, physical demand, temporal demand, performance, effort, frustration), p's $< .001$. There was no effect of task priority on workload ratings, nor was there a significant interaction, F's < 1.

Choice Trials

The above analysis demonstrates that when participants were forced to use the audio output for the secondary task, it took longer to complete a trial but steering performance was better. In contrast, when participants were forced to use the visual interface for the secondary task, they could complete the task faster but steering performance declined more because their eyes had to be taken off the road. However, the exact impact on steering performance of using the visual interfaces was moderated by the driver's priority. In addition, having to share visual attention between the steering and the search task was rated as being far more demanding than using the audio interface. Given these characteristics, which modality might participants prefer to use, and how might this choice be affected by differences in task priority instructions?

In the choice trials, participants were given a two-alternative forced-choice over which output modality they would like to use at the start of each trial. For all statistical analyses of the choice phase data, an independent samples t-test was used.

Figure 4 is a histogram showing the proportion of trials that each participant chose the audio interface for each priority condition. It can be seen in the figure that participants who were instructed to prioritize the steering task were more likely to select the audio interface ($M = .88$, $SD = .16$) than participants who were instructed to prioritize completing the secondary task quickly ($M = .33$, $SD = .23$), $t(22) = 6.77$, $p < .001$, $d = 2.77$, 95% CI [.38, .71]. The figure also shows that 7 out of the 12 participants in the steering-focus condition chose the audio interface for every trial. Whereas, participants in the searching-focus condition did not show such a strong preference for one interface over another, choosing instead to flip between the audio and the visual interface on different trials. We next consider the impact of these choices on the main task performance metrics.

The two right-most columns of Figure 2 and 3 show respectively steering task performance (RMSE lateral deviation) and task time performance for the choice trials. In terms of steering performance, it can be seen that participants in the steering-focus condition maintained an RMSE lateral deviation equivalent to that observed in both the no-choice audio condition and at baseline single-task. In contrast, participants in the searching-focus condition
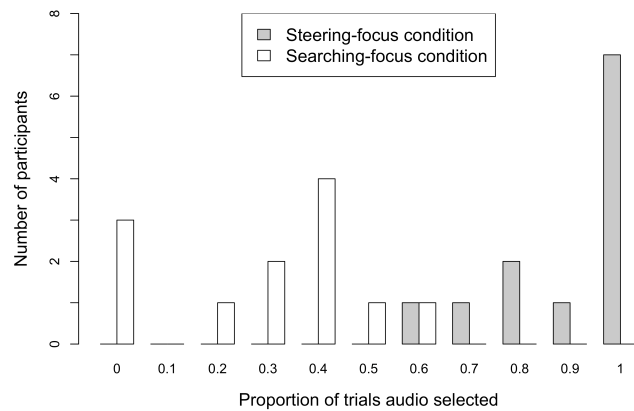

Figure 4. Histogram showing the proportion of trials that each participant selected the audio interface across choice trials

deviate more from the lane center, and allowed the vehicle to drift out to the lane boundary. The inverse pattern can be found for task time data, in that participants in the steering-focus condition completed the task in the time associated with using the audio interface, whereas participants in the searching-focus condition were at the lower task times associated with using the visual interface in the no-choice searching-focus condition. These observations on dual-task performance are consistent with the choices made by participants in each priority condition. Statistical analysis support these observations, and show that participants in the steering-focus condition took significantly longer to complete the secondary task than participants in the searching-focus condition, $t(22) = 3.80$, $p < .001$, $d = 1.55$, 95% CI [5.76, 19.62], but these participants had significantly better steering performance, $t(22) = 3.39$, $p < .005$, $d = 1.38$, 95% CI [.15, .64].

DISCUSSION

Audio output offers a promising interaction technique for alleviating some of the problems associated with using mobile devices while driving, such as visual distraction. However, the results of this study highlight that task completion time is a critical dimension for designing such audio interfaces. We found that participants that were encouraged to complete a secondary 'lookup' task quickly chose to forfeit the benefits of using audio output. Instead, they choose to glance over to a secondary display so that they could locate the target information quickly, resulting in a significant decline in lane keeping performance on the primary driving task. In contrast, participants that were encouraged to prioritize the driving task tended to select the slower audio output option so that they could maintain lateral control of the vehicle.

The results of this work have important implications for the design of multimodal interfaces that give feedback across multiple modalities at the same time [10,34,35], because such interfaces present the user with just the kind of choice over which modality to attend to as considered here. Our results show that the user's performance objective will have

a strong influence on which modality is attended to on a multimodal device, and that audio and visual output can serve time and safety goals differently.

The use of a no-choice/choice paradigm [15,23,28,32] was essential for this work. The no-choice phase allowed participants to gain experience with each output modality and for us to gather an objective measure of the performance characteristics of each, before participants were given a choice about which method they preferred to use to meet a specific task objective. It was found that decisions about which modality to use were moderated by the users' *performance objective* (i.e., which of the two tasks was subjectively more important to them). Decisions were not made on the basis of simply selecting the fastest strategy for completing the task at hand [8].

These results contribute to a growing body of work on understanding how people adapt their behavior while multitasking to meet specific performance objectives [2,3,14,16-18]. The results of the current work are distinct from this previous line of research however, in that we focus here on how performance objectives influence modality *selection* rather than fine-grained task interleaving behavior. Our results show that people can make different overt choices between methods for completing a secondary task based upon the performance characteristics of each method and whether they meet the desired performance objective. In the following section, we consider a number of important limitations of the current study and some questions that future research in this area might examine.

Limitations and future research
One concern with generalizing the findings of this research to real-world contexts is that we were limited to using a low-fidelity driving simulator. Our driving task required participants only to control the steering of a vehicle as it travelled at a constant speed down a straight highway environment. This lane keeping performance measure can serve as a useful proxy for assessing drivers' attention, since steering control is an open-loop process that requires constant visual input [31]. However, this setup clearly does not allow us to investigate the full complexity of driver behavior. In particular, there is a concern that our lane keeping measure is not sensitive to the effects of cognitive distraction associated with performing a non-visual secondary task. Previous studies have found that such effects tend to have a stronger impact on a driver's ability to maintain a constant speed when following another vehicle, and their ability to react to unexpected events [1,11]. The effects of cognitive interference do show up on lane keeping performance metrics, but the effects can be small [26]. Consequently, there is a concern that our steering measure was maybe not sensitive enough to detect any negative effects of using the audio output for the secondary task. Our objective here however was not to show that an audio interface is not distracting for drivers,

but instead to highlight the issue of how performance objectives will influence modality choices.

Future work might investigate people's modality choices in a more complex driving environment. It is reasonable to assume that people might have a stronger preference for using audio output in a more demanding environment, such as when there is heavy traffic or when driving at a high speed. In contrast, we might expect drivers to shift to using a visual interface when there is very little demand from the driving task, such as when waiting at a red light. The question would be whether such changes in modality preference would occur independent of the participant's objective (as manipulated here). There is some support for this view. A recent study by Iqbal et al. [16] showed that multitasking drivers give greater attention to the road when the driving environment is more challenging.

In contrast though, the results of a recent on-road driving study by Horrey and Lesch [13] showed that drivers show no preference for waiting until periods of decreased driving demand (i.e., when waiting at a red stop light) to engage in secondary in-car tasks. This latter finding would suggest that it is unlikely that drivers will spontaneously shift from one modality to another based on the changing demands of the road, or might they? A further possibility is that drivers might cope with the demands of performing a secondary in-car task in other ways, such as by reducing their speed [4,16]. Further research is required to address these questions.

One concern for generalizing from these results to how people might use multimodal devices in general is that we forced participants to make a choice between one of two output modalities. In contrast, we might have presented information via both output modalities simultaneously. Such a setup would allow the user to choose on the fly which modality to pay attention to, and to switch back and forth between different modalities while completing a task. The results of previous studies that have evaluated the use of multimodal interfaces while driving [22,34] are broadly consistent with those reported here: audio feedback is less distracting for the driver than visual feedback. Critically, it does not seem to matter whether multimodal audio+visual feedback is given or whether only audio feedback is given; the safety benefit is given by audio feedback.

Our results suggest that there might be considerable value in applications that disable visual output and present audio output in isolation. By having only a single output modality, the driver would not be tempted into glancing over to a device's display to complete a secondary task more quickly. This idea seems to have made its way into recent applications aimed for in-car use, such as DriveSafe.ly™ (http://www.drivesafe.ly/), which reads text (SMS) messages and emails aloud while not presenting the message on the phone's screen. Based upon the results of the current work, this would seem to be a promising approach as it removes the temptation to use an

inappropriate visual modality. However, the potential for cognitive distraction from engaging in a secondary task while driving might remain [11,26]. Further work is required to unpack these issues more thoroughly.

Our results hinge on the assumption that the audio interface was slower for completing the secondary task than the visual interface. However, this does not have to be the case. There would be great value in pursuing design solutions that make audio interfaces faster to use so that users are not forced to make the kinds of trade-offs considered here. Interesting work has been done to evaluate various methods for compressing audio information so that it can be listened to more quickly while preserving critical information content [30]. We believe that such developments might hold promise for audio interfaces.

A second consideration is to allow the user more active control over the presentation of audio information. In our study, the audio output was read at the default reading speed of the software, which meant that listening to a complete passage of text took a fixed amount of time. The participant could not actively speed up or slow down the audio output stream, nor fast-forward or rewind it. This could be done differently in future studies. Indeed, examples of interactive audio interfaces abound, such as [21,34,35]. Zhao's earPod [34,35] for instance uses interruptible audio that plays the aural information associated with the currently highlighted menu option regardless of whether or not the previous audio file has completed. This allows the user to rapidly skim through menu options without having to exhaustively listen in full to each option.

There are then at least two ways in which the audio interface used here might be improved. First, audio content might be presented at a faster pace. Second, an interactive model might be developed to allow the user to skim though audio content to locate the relevant section more quickly. Both of these alternatives would conceivably work to allow the secondary task to be completed more quickly using the audio interface. We would expect this to have an impact on people's willingness to use the audio interface in the kind of multitask setup used here. This is because the results of the current study suggest that people consider the time taken to complete a task using each output option when deciding between them.

CONCLUSION

To summarize, a study was conducted to investigate whether people will choose to use a safer audio interface over a traditional GUI to complete a secondary task while driving in situations with different priorities. A no-choice/choice paradigm was used in which participants first completed a series of trials using each of the two output interface modalities before being given a choice on which to use. The benefit of using this paradigm are that we got both a reliable measure of the performance characteristics associated with each interface, and a measure of which interface participants preferred to use. Results from the no-choice phase showed that the audio interface did not interfere with driving performance but that it was slower to use than a traditional GUI. Given the choice of which interface modality to use and the performance characteristics associated with each, participants made different choices based on which task they were told to give greater emphasis to. Those participants that had been instructed to prioritize safer steering tended to select the audio interface, while those that had been encouraged to prioritize the secondary task chose the visual interface. These decisions had a large effect on dual-task performance metrics. The results of this work suggest that multimodal interfaces that make use of audio feedback offer a promising solution to alleviate some of the problems associated with using mobile devices in multitask environments. However, attention should be given to the time required to complete basic interactive tasks using audio feedback because if time is of the essence, people will likely forsake whatever safety benefits there are in order to achieve their goal quickly.

REFERENCES
1. Brookhuis, K., De Vries, G., & De Waard, D. The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention, 23,* (1991), 309-316.

2. Brumby, D.P., Salvucci, D.D., & Howes, A. Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. In *Proc. CHI 2009*, ACM Press (2009), 1629-1638.

3. Brumby, D.P., Del Rosario, N., & Janssen, C.P. When to switch? Understanding how performance trade-offs shape dual-task strategy. In *Proc. ICCM 2010*, Drexel University (2010), 19-24.

4. Cnossen, F., Meijman, T., & Rothengatter, T. Adaptive strategy changes as a function of task demands: A study of car drivers. *Ergonomics, 47,* (2004), 218-236.

5. Crowd Science. Social media insights – twitter. http://www.crowdscience.com/blog/article/social_media _study_twitter_while_driving/, September 4th 2009.

6. Diels, C., Reed, N., and Weaver, L., *Drivers' attitudes to distraction and other motorists' behaviour: A focus group and observational study.* TRL Limited, Wokingham, UK, 2009.

7. Duggan, G.B. & Payne, S.J. Text skimming: The process and effectiveness of foraging through text under

time pressure. *Journal of Experimental Psychology: Applied, 15,* (2009), 228-242.

8. Gray, W.D. & Boehm-Davis, D.A. Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6,* (2000), 322-335.

9. Hart, S.G. & Staveland, L.E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human Mental Workload, 1,* (1988), 139–183.

10. Hoggan, E., Crossan, A., Brewster, S., & Kaaresoja, T. Audio or tactile feedback: Which modality when? In *Proc. CHI 2009*, ACM Press (2009), 2253-2256.

11. Horrey, W. & Wickens, C.D. Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48,* (2006), 196-205.

12. Horrey, W.J. & Wickens, C.D. Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors, 46,* (2004), 611-624.

13. Horrey, W.J. & Lesch, M.F. Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation. *Accident Analysis and Prevention, 41,* (2009), 115-122.

14. Horrey, W.J., Wickens, C.D., & Consalus, K.P. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12,* (2006), 67-78.

15. Howes, A., Duggan, G.B., Kalidindi, K., Tseng, Y.-C., & Lewis, R.L. Bounded optimality in a cognitive utility learning paradigm. (under review).

16. Iqbal, S.T., Ju, Y.-C., & Horvitz, E. Cars, calls, and cognition: Investigating driving and divided attention. In *Proc. CHI 2010*, ACM Press. (2010), 1281-1290.

17. Janssen, C.P. & Brumby, D.P. Strategic adaptation to performance objectives in a dual-task setting. *Cognitive Science, 34,* (2010), 1548-1560.

18. Janssen, C.P., Brumby, D.P., Dowell, J., Chater, N., & Howes, A. Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. *Topics in Cognitive Science, 3,* (2011), 123-139.

19. Lai, J., Cheng, K., Green, P., & Tsimhoni, O. On the road and on the web?: Comprehension of synthetic and human speech while driving. In *Proc. CHI 2001*, ACM Press (2001), 206-212.

20. Lee, J.D., Caven, B., Haake, S., & Brown, T.L. 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.

21. Li, K.A., Baudisch, P., & Hinckley, K. Blindsight: Eyes-free access to mobile phones. In *Proc. CHI 2008*, ACM Press (2008), 1389-1398.

22. Liu, Y. Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics, 44,* (2001), 425-442.

23. MacLean, A., Barnard, P., & Wilson, M. Evaluating the human interface of a data entry system: User choice and performance measures yield different trade-off functions. In P. Johnson & S. Cook (Eds.), *People and Computers: Designing the Interface*. Cambridge University Press, 1985, 172-185.

24. Reed, M. & Green, P. Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics, 42,* (1999), 1015-1037.

25. Salvucci, D.D. Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human Computer Studies, 55,* (2001), 85-107.

26. Salvucci, D.D. & Beltowska, J. Effects of memory rehearsal on driver performance: Experiment and theoretical account. *Human Factors, 50,* (2008), 834-44.

27. Salvucci, D.D. & Taatgen, N.A. Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review, 115,* (2008), 101-130.

28. Siegler, R.S. & Lemaire, P. Older and younger adults' strategy choices in multiplication: Testing predictions of ascm using the choice/no-choice method. *Journal of Experimental Psychology: General, 126,* (1997), 71-92.

29. Sodnik, J., Dicke, C., Tomazic, S., & Billinghurst, M. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies, 66,* (2008), 318-332.

30. Tucker, S. & Whittaker, S. Time is of the essence: An evaluation of temporal compression algorithms. In *Proc. CHI 2006*, ACM Press (2006), 329-338.

31. Wallis, G., Chatziastros, A., & Bülthoff, H. An unexpected role for visual feedback in vehicle steering control. *Current Biology, 12,* (2002), 295-299.

32. Walsh, M.M. & Anderson, J.R. The strategic nature of changing your mind. *Cognitive Psychology, 58,* (2009), 416-40.

33. Wickens, C.D. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3,* (2002), 159-177.

34. Zhao, S., *Earpod: Efficient hierarchical eyes-free menu selection*. Doctoral dissertation, Department of Computer Science, University of Toronto, 2009.

35. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., & Baudisch, P. Earpod: Eyes-free menu selection using touch input and reactive audio feedback. In *Proc. CHI 2007*, ACM Press (2007), 1395-1404.