*Human Factors* topic: Displays and Controls

# Speech Auditory Alerts Promote Memory for Alerted Events in a Video-simulated Self-driving Car Ride

Michael A. Nees, Benji Helbein, & Anna Porter

Lafayette College

Submitted for consideration for publication in *Human Factors.*

Main text word count: 4646

Corresponding author:
Michael A. Nees
Lafayette College
Department of Psychology
Oechsle Hall
350 Hamilton Street
Easton, PA 18042
Ph: (610) 330 5290
neesm@lafayette.edu
nees.michael@gmail.com

# Precis

An experiment examined memory for alerted events—a component of Level 1 situation awareness—with speech alerts, auditory icons, and visual text displays in a simulated self-driving car ride under routine circumstances with a visual secondary task. Results suggested that speech alerts were most effective, though both types of auditory displays were perceived to be more annoying.

# Abstract

**Objective:** Auditory displays could be essential to helping drivers maintain situation awareness in autonomous vehicles, but to date few or no studies have examined the effectiveness of different types of auditory displays for this application scenario.

**Background:** Recent advances in the development of autonomous vehicles (i.e., self-driving cars) have suggested that widespread automation of driving may be tenable in the near future. Drivers may be required to monitor the status of automation programs and vehicle conditions as they engage in secondary leisure or work tasks (entertainment, communication, etc.) in autonomous vehicles.

**Method:** An experiment compared memory for alerted events—a component of Level 1 situation awareness—using speech alerts, auditory icons, and a visual control condition during a video-simulated self-driving car ride with a visual secondary task. The alerts gave information about the vehicle's operating status and the driving scenario.

**Results:** Speech alerts resulted in better memory for alerted events. Both auditory display types resulted in less perceived effort devoted toward the study tasks, but also greater perceived annoyance with the alerts.

**Conclusion:** Speech auditory displays promoted Level 1 situation awareness during a simulation of a ride in a self-driving vehicle under routine conditions, but annoyance remains a concern with auditory displays.

**Application:** Speech auditory displays showed promise as a means of increasing Level 1 situation awareness of routine scenarios during an autonomous vehicle ride with an unrelated secondary task.

**Keywords**: auditory displays, autonomous vehicles, human-automation interaction, situation awareness, auditory icons, speech displays, dual-task performance, memory

# Speech Auditory Displays Promote Memory for Alerted Events in a Video-simulated Self-driving Car Ride

Self-driving vehicles have received considerable attention recently, as technological advancements have fostered optimism about their feasibility. Despite this enthusiasm, a number of formidable barriers to the deployment of autonomous vehicles remain. Perhaps most neglected among these obstacles are unresolved issues related to human factors, which have only just begun to be recognized as a high priority for research (Fagnant & Kockelman, 2013; National Highway Traffic Safety Administration, 2013) and may represent the most daunting challenge—greater than technological, legal, and security concerns—for self-driving cars (Hodson, 2015).

Both published guidelines (NHTSA, 2013) and current research (Merat et al., 2014) have suggested that the realization of fully automated vehicles will be preceded (or even indefinitely characterized) by a period during which drivers are expected to be able to resume control of vehicle functions in the event of automation failure (e.g., in bad weather when sensors are less reliable). The operator also likely will retain control of important decision-making and action tasks during routine, normal operation of automated programs. For the foreseeable future, even during periods of high or full vehicle automation the human operator will need to maintain awareness of the driving scenario to ensure safe and effective operation of automated vehicles,.

Initial reports (Merat, Jamson, Lai, & Carsten, 2012, 2014) have revealed potential problems when drivers must resume manual control of vehicles following a period of high or full automation.  The culprit for these difficulties seems to be failures of *situation awareness* (SA) during periods of high or full automation (Merat & Jamson, 2009).  SA is a construct—distinct from, but related to, task performance—that refers to the extent to which human operators perceive relevant information in a complex task (Level 1 SA), comprehend the information

(Level 2 SA), and use the information to successfully anticipate future system states (Level 3 SA) (Endsley, 1995). Achievement of Level 1 SA has been theorized as a prerequisite to achieving Levels 2 and 3 SA (Endsley, 1995), and studies have attributed the majority of SA failures to Level 1 factors such as perception, memory, or mistakes with monitoring the status of systems (Jones & Endsley, 1995, 1996).

Failures of Level 1 SA will be a concern during periods of automation in vehicles. Displays in self-driving vehicles will need to keep the driver in the loop by providing information about the vehicle state and driving scenario during routine operation. Although a variety of visual displays have been developed to inform drivers of automation status and vehicle conditions in automated vehicles (see Manca, De Winter, & Happee, 2015), research has shown that higher levels of automation in vehicles will be accompanied by greater driver willingness and desire to engage in visually-intensive secondary tasks (Kyriakidis, Happee, & De Winter, 2014; Llaneras, Salinger, & Green, 2013; Merat et al., 2014), including watching movies or televisions shows, reading, texting, or emailing. Displays should be designed to support monitoring and awareness of the vehicle status during concurrent visual secondary tasks.

Auditory displays (see Kramer, 1994) offer an appealing candidate solution that could keep the human operator of autonomous vehicles in the information loop while also permitting some degree of engagement with non-driving tasks. Research has suggested that auditory or multimodal displays may be superior to visual displays for monitoring tasks over relatively long (115 hours) task periods (see Colquhon, 1975; Doll & Hanna, 1989), and long commutes may involve prolonged monitoring of automated vehicle systems.  Further, in previous research auditory displays allowed for adequate monitoring performance while also freeing visual resources for secondary tasks (Seagull, Wickens, & Loeb, 2001), because auditory and visual information channels represent distinct pools of modality resources (Wickens, 2002).

There are several precedents for the use of auditory alerts for in-vehicle technologies (for a review, see Nees & Walker, 2011). In autonomous vehicle research, Merat and Jamson (2009) used an unspecified auditory alarm to indicate that manual resumption of vehicle control from automation was required during a critical event, and Merat et al. (2014) used speech messages in addition to visual displays to provide information about the status and availability of automated programs. Google's self-driving car uses an auditory chime to signal imminent resumption of manual control by the human driver (Bilger, 2013), and van den Beukel and van der Voort (2013) used a three tone alert to signal that the driver needed to respond to a critical event.

Few systematic investigations of the effectiveness of different types of auditory displays for status monitoring in autonomous vehicles have been conducted, however, as reported designs to date have defaulted mostly to basic tones and chimes. Auditory display research has suggested that abstract tones and chimes may be inferior to available alternatives (for a review of alternatives, see Nees & Walker, 2009), because the meanings of abstract sounds are generally harder to learn and remember (e.g., Bonebright & Nees, 2007; McKeown & Isherwood, 2007). Both speech auditory displays and auditory icons—nonspeech sounds that bear some ecological relationship to their referent processes—have fared better for in-vehicle alerts. McKeown and Isherwood (2007) showed that speech and auditory icons were better candidates for in-vehicle alerts than abstract sounds or unrelated environmental sounds. The speech alerts they used were spoken messages (e.g., "oil is low"), and their auditory icons were brief, naturalistic sounds that were chosen on the basis of their non-arbitrary, though indirect, relationship with their referent processes (e.g., the sound of water pouring represented low fuel). In a task that required participants to match an auditory alert with a visual pictorial referent, identification performance was nearly perfect for families of speech alerts and auditory icons, and both resulted in faster response times than the other candidate auditory displays, which suggested that both speech alerts and auditory icons are learnable and memorable (also see Belz, Robinson, & Casali, 1999; McKeown, Isherwood, & Conway, 2010).

Memory for information is a component of Level 1 SA, and past research has shown that recall of information is related to overall SA. For example, using simulated (animated) driving scenarios that were 18-35 s in duration as stimuli, Gugerty (1997) examined both recall and performance measures of situation awareness. The recall task required participants to view the animations with the car on "autopilot" (i.e., full automation). After the animation completed, participants recalled all vehicles in the vicinity of their car. For performance measures, participants had to intervene to avoid collisions with responses on computer keyboard arrows. Performance and recall measures were positively correlated, and Gugerty concluded that both relied upon explicit (rather than implicit) knowledge—information that can be accessed in memory recall tasks. Similarly, other researchers (Ma & Kaber, 2005; van den Beukel & van der Voort, 2013) have used recall tasks to gauge Level 1 SA in simulated driving tasks.

The present study examined memory for auditory alerts in a simulated self-driving car ride under routine conditions. A secondary task was present, but no critical driving response was required. These routine conditions likely will represent the majority of riding time during periods of full automation in a self-driving vehicle. Despite the routine nature of these scenarios, drivers will need to remain in the loop about the vehicle and driving conditions. Maintenance of SA will begin with Level 1 SA—in this case perceiving, remembering, and monitoring on-going information about the driving scenario during routine vehicle operations. Failures of Level 1 SA have been identified as the most common SA failures (Jones & Endsley, 1995, 1996), and Level 1 SA failures lead to breakdowns in Level 2 and Level 3 SA. Loss of SA is a precursor for eventual mistakes in response to critical incidents (see Endsley, 1995). As such, memory for alerted events during routine conditions offers one surrogate measure of the extent to which alerts are effective at helping establish Level 1 SA.

We investigated speech and auditory icons, as both have shown promise over other candidate auditory display types (e.g., abstract sounds) in past research. Videos depicted a real, routine driving scenario. During videos, alerts about the vehicle's status and the driving scenario were provided via visual text (a control condition), speech, or auditory icons in a between-subjects design. Participants also performed a visual secondary task—a word search puzzle—during the simulated rides. The word search task was chosen because it imposed visual demands on attention and presumably was a task with which participants were already familiar. Further, it was representative of one of the types of leisure activities that might be undertaken during a commute.

We compared memory for alerted events across these three display formats using a free recall query. Participants also rated their confidence in their ability to assume manual control of the car, the usefulness of the alerts, and their annoyance with the alerts. Subjective workload was also assessed. We expected that both auditory display conditions would result in better memory, greater confidence, greater perceived usefulness, and lower subjective workload than the visual control condition, because the visual alert condition presented an attentional conflict with the visual secondary task (see Wickens, 2002). Our comparison of speech and auditory icons was exploratory. We included an annoyance rating, because annoyance has been identified in previous research (Marshall, Lee, & Austria, 2007, also see Nees & Walker, 2011) as a potential problem for auditory displays for in-vehicle technologies.

# Method

## Participants

Participants (*N* = 90) were recruited from undergraduate psychology courses and were compensated with course credit. Self-reported status as a licensed driver was an inclusion criterion. Data from five participants were excluded from all analyses. Two completed the experiment for course credit but self-reported that they were not licensed drivers. One withdrew before completing all procedures. One appeared to fall asleep during study procedures. One found no words across all word search puzzles and was deemed to have failed to follow instructions. The final sample used in analyses included *N* = 85 participants (*n* = 30 in the visual condition, *n* = 28 in the auditory icons condition, and *n* = 27 in the speech condition). Fifty-two participants were females, and the mean age of participants was 19.35 years (*SD* = 1.03).

## Stimuli, Materials, and Apparatus

**Self-driving vehicle video simulations.** Videos with ambient vehicle cabin sounds were recorded using a dash-mounted iPhone 6 (30 frames per second, 1080p high definition). Three different videos of continuous driving scenarios provided variety in the stimuli and multiple measurements of dependent variables. The first and second videos featured scenes from a small town, while the third video involved merging onto a freeway and entering a construction zone. The videos were 170, 176, and 309 s in length, respectively, so the onset of the memory probe was not predictable from previous trials. Although the videos were selected to represent a variety of driving scenarios, they were nonequivalent and were not selected to evaluate different driving conditions per se. Thus, all analyses used averaged responses across videos. Videos were edited in iMovie for length, content, and to add the alerts for each of the three conditions described below. The videos presented 7, 9, and 11 alerts, respectively. The alerts and the videos in which they appeared are listed in Table 1. Maintenance and condition alerts (e.g., “low fuel” and “rain reported ahead”) were inserted arbitrarily. Hazard alerts (“pedestrian,” “side hazard,” and “front hazard”) corresponded to events that were present in the videos and thus appeared more frequently. Pedestrian alerts corresponded to pedestrians visible in upcoming crosswalks or approaching upcoming crosswalks. Side hazards corresponded to vehicles approaching intersections from sides streets or in the vehicle’s blind spot. Front hazards corresponded to vehicles crossing the center line (see Figure 1), entering the roadway directly in front of the vehicle from side streets, or stopping in front of the vehicle at a location that was not a controlled intersection (i.e., in the construction zone in the third video). The type of alert used was the only difference in the videos across conditions.

Table 1

*Alerts used in the study*

| Visual text/speech message | Auditory icon sound | Appeared in videos: |
|---|---|---|
| “Low fuel” | Water gurgle | 1, 3 |
| “Low tire pressure” | Air hiss | 1 |
| “Low wiper fluid” | Water squirt | 2, 3 |
| “Rain reported ahead” | Thunder | 3 |
| “Traffic reported ahead” | Multiple car horns | 1 |
| “Construction reported ahead” | Jackhammer | 2, 3 |
| “Pedestrian” | Crossing guard whistle | 1, 2 (3 times), 3 |
| “Side hazard” | Single car horn twice | 1, 2 (4 times), 3 (4 times) |
| “Front hazard” | Tires screeching | 1 (twice), 3 (twice) |

**Alert conditions.** The visual condition presented the text of the alert (see Table 1) in large (approximately 72 point), white font at the bottom of the video screen as a head-up display (see Figure 1). For the speech condition, the text of the visual alerts was converted to speech using the NeoSpeech (www.neospeech.com) text-to-speech (TTS) engine (US English voice "Kate" at the default speed setting). Speech alerts ranged in duration from 0.77 to 1.51 s (*M* = 1.11, *SD* = .28). For the auditory icons condition, sounds (listed in Table 1) were created from the environmental sounds used by Marcell, Borella, Greene, Kerr, and Rogers (2000) and from sound effects from the Edge Edition Volume 1 Sound Effects library (http://www.sound-ideas.com). Auditory icons ranged in duration from 0.27 to 2.20 s (*M* = 1.13, *SD* = .72). The auditory icon signal-referent mappings chosen for this experiment were comparable to those used in previous research on in-vehicle alerts (McKeown & Isherwood, 2007; McKeown et al., 2010).

*Figure 1. Screen shot from the visual alerts condition showing the "Front Hazard" alert as an oncoming vehicle crossed the center-dividing lane lines.*

**Word search materials.** Three word searches were obtained from a free word search website (http://www.puzzle-club.com/). All were described by the website as medium difficulty, and covered different themes (1990s bands, 1930s films, and comic book superhero names). Pairing of the word searches with the videos was counterbalanced across participants.

**Apparatus.** Participants were seated at desks in individual testing rooms in front of a 55 cm widescreen Dell LCD monitor, which subtended approximately 55 degrees of visual angle during the experimental procedure. All sounds were presented through Koss UR-20 headphones. By default, the computer was set to a level that was approximately the intensity of a normal conversation (~ 60 dB SPL); participants were allowed to adjust the volume to a level that they felt was comfortable for listening. Piloting by the experimenters determined that all auditory alerts were clearly audible above the background vehicle noise recording. A program written in Adobe Director presented all stimuli and collected all data except for the word search data, which were written by hand.

## Procedure

Participants were randomly assigned to one of the three display conditions. Participants in all conditions experienced a brief training phase, during which they saw or heard each of the alerts for their condition and were informed of the alerts' meanings. In the audio conditions, participants could listen to the sound as many times as needed to learn the message. A recognition test followed the training phase; participants had to correctly recognize every alert before they continued. For the main experiment, participants were told to pay attention to the video and try to remember the alerts. Before each video, participants received a new word search on paper, and the word search was placed on the desk directly in front of the computer monitor. Participants were instructed to find as many words as possible during the video while also maintaining awareness of the vehicle and its surroundings. Participants also were instructed to only write on the word search paper to circle answers; no participant wrote the names of alerts or other situational information about the video on the paper. Each video ended with an open-ended prompt (similar to a freeze probe) that asked participants to perform free recall of as much information as they could about the driving scenario just experienced. Participants were allowed to type as much as they wished into a text response box, and they could take as long as they wished to respond. Participants also responded to three questions on 7-point Likert rating scales following each video. The questions asked participants to rate the usefulness of the alerts, the annoyance caused by the alerts, and the extent to which they would feel confident resuming control of the vehicle at the moment of the video stoppage. The three videos were stopped, respectively, with: 1) the vehicle edging by a stopped delivery truck as it approached an active intersection with passing traffic; 2) the vehicle stopped at a stop sign at an active intersection with passing traffic; and 3) the vehicle being passed on the freeway by other cars with a car emerging from its blind spot. Verbatim instructions and questions are included in Appendix A. Participants completed a computer version of the NASA-TLX (Hart & Staveland, 1988) at the end of the experiment to assess the workload they encountered during the simulated rides.

## Results

Answers to the recall question were coded by two independent raters. The raters were not aware of the experimental conditions of participants. The raters counted the number of alerts correctly recalled in the response (e.g., "There was a message for low fuel."). Ratings for each participant across all three videos (255 total cases) showed that inter-rater reliability (*ICC*, absolute agreement for averaged ratings, see McGraw and Wong, 1996) between the two raters was .96. For final analyses, the average of discrepant ratings was used. The three occurrences of each variable (one from each video) from each participant were averaged into a mean percent of events recalled.

Visual inspection of histograms revealed approximately normal distributions of all variables except for the TLX physical workload scale, which exhibited moderate positive skew.

Given recent affirmations of the robustness of analysis of variance techniques against violations of normality (e.g., Schmider, Ziegler, Danay, Beyer, & Bühner, 2010), one-way ANOVAs were used in all analyses. Tukey's test was used for pairwise post hoc comparisons following significant omnibus tests.

## Memory Measure

Memory for the percent of alerted events recalled showed a significant effect of condition, $F(2,82) = 6.91$, $p = .002$, $\eta^2_p = .14$. Participants recalled more events in the speech condition ($M = 63\%$, $SD = 15\%$) than the auditory icons ($M = 52\%$, $SD = 15\%$), $p = .02$, or visual conditions ($M = 49\%$, $SD = 16\%$), $p = .002$. The difference between the auditory icons and visual conditions was not significant, $p = .75$.

## Likert Ratings

There was no effect of condition on participants' mean ratings of confidence in their ability to resume control of the vehicle at the moment the video ended (Grand $M = 4.30$, $SD = 1.60$), $F(2,82) = 0.89$, $p = .42$. There was also no effect of condition on mean ratings of the usefulness of the alerts (Grand $M = 4.38$, $SD = 1.33$), $F(2,82) = 2.76$, $p = .07$. There was a significant effect of condition on ratings of annoyance, $F(2,82) = 8.03$, $p = .001$, $\eta^2_p = .16$. Participants found the visual alerts ($M = 2.68$, $SD = 1.42$) less annoying than the auditory icons ($M = 4.12$, $SD = 1.55$), $p = .001$, and speech alerts ($M = 3.70$, $SD = 1.25$), $p = .019$. The difference between the auditory icons and speech alerts was not significant, $p = .51$.

## NASA TLX

The weighted scores for the six subscales of the NASA TLX were analyzed for differences by condition. There was a significant effect of condition on the perceived effort subscale, $F(2,82) = 5.73$, $p = .005$, $\eta^2_p = .12$. Participants felt the visual condition required greater effort (*M* = 51.00, *SD* = 23.24) than the auditory icons (*M* = 34.30, *SD* = 18.21), *p* = .007, or speech (*M* = 36.86, *SD* = 18.56), *p* = .025, conditions. The difference between the auditory icons and speech conditions was not significant, *p* = .886. There were no significant effects of condition on any other TLX subscales (see Table 2).

Table 2

*One-way ANOVA Results for the Effect of Alert Condition on NASA TLX subscales*

| TLX subscale | *F*(2,82) | *p* |
|---|---|---|
| Effort | 5.73 | .005* |
| Mental | 1.40 | .252 |
| Physical | 1.62 | .205 |
| Temporal | 1.74 | .183 |
| Performance | 1.51 | .227 |
| Frustration | 0.13 | .878 |
| Composite (overall) | 0.84 | .435 |

*Note* * = $p < .05$

## Word Search Task

There was no effect of condition on the mean number of words found in the word search task (Grand *M* = 4.13, *SD* = 1.56), *F*(2,82) = 0.34, *p* = .72. To examine the possibility of a tradeoff between performance of the word search puzzle task and attending to the driving scenario, we correlated the mean number of words found per word search with the mean number of events recalled per video. We found a significant positive correlation when scores were aggregated across all conditions of the experiment, *r*(83) = .23, *p* = .03, although none of the correlations within conditions were statistically significant when examined separately, *p*s = .10 - . 34.

# Discussion

The speech alerts resulted in greater memory for alerted events than either of the other conditions. The overall positive correlation between the number of words found in the secondary word search task and the number of events recalled suggested that people who found more words also had a tendency to remember slightly more about the driving scenario (i.e., the tests showed weak positive manifold). Thus, the memory results did not seem to be attributable to a performance trade-off with the secondary task. Our data suggested that participants were not attuned to the memory advantage for the speech display, as there were not differences across conditions in their subjective impressions of the usefulness of the alerts or their confidence in their preparedness to resume manual control of the vehicle. Both types of auditory displays resulted in lower perceived effort exerted toward the task as compared to the control condition, which is perhaps not surprising given that the word search secondary task also required visual resources (see Wickens, 2002). None of the other five TLX subscales, however, indicated differences in perceived workload across display conditions.

Recall performance with auditory icons was inferior to performance with speech alerts in this experiment. Auditory icons and speech alerts were comparable and both near ceiling for identification performance in previous research (McKeown & Isherwood, 2007). One possible explanation is that our training phase did not adequately familiarize participants with the meaning of the auditory icons. More research is needed to examine this result further, however, because our participants passed an auditory icons recognition test before the main experiment.

Our results showed that annoyance with auditory displays may be a concern in self-driving vehicle applications. Participants found the visual displays less annoying than both types of auditory displays. Annoyance remains a major potential problem with auditory displays in general (for a discussion, see Kramer, 1994; Nees & Walker, 2011), and especially with discrete alarms (see Edworthy & Hellier, 2000). Annoying auditory displays run the risk of being turned off or ignored. Further, Likert ratings suggested that participants did not perceive that either of the auditory displays were more useful, despite the recall data suggesting that the speech displays, in particular, were more useful for the memory task. If the end users perceive auditory displays to be annoying and to add no value, acceptance of the displays will present a challenge.

Our study sampled college undergraduates by convenience and used a video simulation of a fully automated self-driving car, so a discussion of the generalizability of these results to different samples (i.e., of middle-aged and older adults) and conditions (i.e., an actual automated vehicle) is warranted. Although driving experience is highly relevant for understanding manual driving performance, we are aware of no theory that would predict that sampling older participants with more manual driving experience would alter perception or memory for auditory alerts. Further, except for hearing loss that occurs as a normal part of the aging process, there is little evidence to suggest that age per se would affect perception of auditory displays for in-vehicle technologies (for a discussion, see Nees & Walker, 2011).

Regarding the representativeness of the video stimuli to an actual ride in a self-driving vehicle, the fidelity of the video was much lower than the fidelity of an actual vehicle or even a high fidelity simulation (with motion, a more complete field of view, a simulated vehicle cockpit, etc.). Representative design, however, requires that the cues provided in the experimental setting are representative of the cues relevant for performing the task in the scenarios to which results are generalized (see Araujo, Davids, & Passos, 2007). Our task was to maintain memory for alerted events during a routine, fully automated driving scenario for which no critical intervention (i.e., engagement with cockpit controls) was required. We simulated a scenario for which the videos should have provided the relevant cues to match the scenario of interest. The videos arguably were more representative of our application scenario than many laboratory simulations of driving tasks reported in the literature, some of which have used animated scenes and keyboard responses to simulate driving responses (e.g., Gugerty, 1997).

For example, the motion cues provided by naturalistic driving or high fidelity simulations seem to be most relevant to tasks related to manual control of the vehicle (Kemeny & Panerai, 2003)—tasks that were not applicable in our simulations. Further, we believe there is utility in our approach; low fidelity simulations to examine candidate auditory displays seem to be an economical way to inform future auditory alert design decisions in high-fidelity simulations or naturalistic tests of interfaces for automated vehicles. To date, high fidelity autonomous simulations seem to have made arbitrary decisions about auditory displays for alerting, so our approach may complement and inform, rather than replace, future high-fidelity simulations and naturalistic driving research.

Of course, future research should examine the utility of alerts for responding to critical events that require manual intervention, and this scenario will likely be more heavily influenced by sample characteristics (e.g., driving experience) and the fidelity of the simulation than the scenario examined here. We also designed our experiment to examine the auditory displays as families of speech alerts or auditory icons—an approach that obscured evaluation of our individual alerts. Past research, however, has reported data on individual alerts like the ones used here (McKeown & Isherwood, 2007). Those results suggested that within-category variability in performance with auditory alerts (i.e., within a set of auditory icons) was small compared to large between-category performance differences (i.e., comparing a set of auditory icons to a set of abstract sounds) (see McKeown & Isherwood, Figure 1). Still, more research is needed to optimize auditory alerts both within and across alert types. If discrete warnings and alerts are to be used in self-driving vehicles, other pressing research needs include, but are not limited to: (1) a better understanding of the optimal timing of delivery of auditory warnings in autonomous vehicles when a critical action is required; (2) quantification of the number of auditory alerts the human operator will tolerate over a period of time during routine operation before alarm fatigue and annoyance set in. Inappropriate overuse of warnings can impair other vital lines of communication and heighten negative affective responses to stressful situations (Edworthy & Hellier, 2000; Edworthy, 1998).

These limitations and qualifications notwithstanding, our results offered initial evidence in support of the use of speech alerts over visual alerts or auditory icons for promoting Level 1 SA during periods of full automation in which drivers are engaged in visual secondary tasks. Although speech alerts have been used to provide drivers with information about the status of automated programs in self-driving car research (e.g., Merat et al., 2014), other researchers have reported using unspecified auditory alarms (Merat & Jamson, 2009) or abstract tones (van den Beukel & van der Voort, 2013) to alert drivers to critical events during periods of automated driving. Studies have suggested that tones may be particularly problematic for in-vehicle technologies due to their abstract nature (for a review, see Nees & Walker, 2011); our findings suggested that speech alerts may be useful for practical applications involving auditory displays for self-driving cars.

## Conclusions

We have argued that the introduction of automation into vehicles will require a cooperative effort between humans and the automation technology whereby human drivers will need to maintain situation awareness of the driving scenario. Our experiment showed a general pattern such that speech alerts showed advantages over auditory icons and a visual display control condition for maintaining memory for events—an important component of Level 1 SA—during a video-simulated ride in a self-driving car with a visual secondary task. These findings suggested that speech displays should be considered to promote situation awareness in practical applications in autonomous vehicles. Although the auditory modality may offer a viable alternative or complement to visual displays in this application scenario, we also found relatively higher perceived annoyance with both types of auditory displays. Future research should quantify the relationship between annoyance and the number and types of auditory alerts that are presented.

**Key Points**

--Even highly-automated vehicles will likely require the driver to maintain situation awareness of the driving scenario during routine operation.

--Recall of alerted events was higher for speech alerts as compared to auditory icons and visual text alerts, which suggested that speech alerts were better for promoting Level 1 SA in a simulated self-driving car ride.

--Annoyance remains a concern with audio alerts, as both auditory displays were rated as more annoying than visual alerts.

# References

Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology*, *19*(1), 69–78.

Belz, S., Robinson, G., & Casali, J. (1999). A new class of auditory warning signals for complex systems: Auditory icons. *Human Factors*, *41*(4), 608.

Bilger, B. (2013). Auto correct: has the self-driving car at last arrived. *The New Yorker*, 96–117.

Bonebright, T. L., & Nees, M. A. (2007). Memory for auditory icons and earcons with localization cues. *Proceedings of the International Conference on Auditory Display (ICAD07).* Montreal, Canada. (pp. 419–422).

Colquhon, W. P. (1975). Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. *Human Factors*, *17*, 425–437.

Doll, T. J., & Hanna, T. E. (1989). Enhanced detection with bimodal sonar displays. *Human Factors*, *31*(5), 539–550.

Edworthy, J. (1998). Does sound help us to work better with machines? A commentary on Rautenberg's paper "About the importance of auditory alarms during the operation of a plant simulator." *Interacting with Computers*, *10*, 401–409.

Edworthy, J., & Hellier, E. (2000). Auditory warnings in noisy environments. *Noise and Health*, *2*(6), 27–39.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64.

Fagnant, D., & Kockelman, K. (2013). *Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations*. Washington, D.C.

Gugerty, L. J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, *3*(1), 42.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239–250). Amsterdam: North Holland Press.

Hodson, H. (2015). Driverless cars in gridlock. *New Scientist*, *225*(3008), 20–21. http://doi.org/10.1016/S0262-4079(15)60303-7

Jones, D.G., & Endsley, M.R. (1995). Investigating situation awareness errors. *Proceedings of the International Symposium on Aviation Psychology,* 746-751.

Jones, D.G., & Endsley, M.R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space, and Environmental Medicine, 67(6),* 507-512.

Kemeny, A., & Panerai, F. (2003). Evaluating perception in driving simulation experiments. *Trends in Cognitive Sciences*, *7*(1), 31–37.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces* (pp. 1–78). Reading, MA: Addison Wesley.

Kyriakidis, M., Happee, R., & De Winter, J. (2014). Public Opinion on Automated Driving: Results of an International Questionnaire Among 5,000 Respondents. *Available at SSRN 2506579*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2506579

Llaneras, R. E., Salinger, J., & Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. In *Proceedings of the Seventh International Driving Symposium on Human*

*Factors in Driver Assessment, Training, and Vehicle Design* (pp. 92–98). Bolton Landing, New York.

Manca, L., De Winter, J., & Happee, R. (2015). Visual Displays for Automated Driving: A Survey. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI 2015)* (pp. 1–5). Nottingham, UK.

Marcell, M. M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology*, *22*(6), 830–864.

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics*, *35*(10), 939–953.

Marshall, D. C., Lee, J. D., & Austria, P. A. (2007). Alerts for in-vehicle information systems: Annoyance, urgency, and appropriateness. *Human Factors*, *49*(1), 145–157.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30–46.

McKeown, D., & Isherwood, S. (2007). Mapping candidate within-vehicle auditory displays to their referents. *Human Factors*, *49*(3), 417–428.

McKeown, D., Isherwood, S., & Conway, G. (2010). Auditory displays as occasion setters. *Human Factors*, *52*(1), 54–62.

Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car. In *Proceedings of the 5th International Driving Symposium on Human Factors in Driver*

*Assessment, Training and Vehicle Design* (pp. 514–521). Retrieved from http://drivingassessment.uiowa.edu/DA2009/072_MeratJamson.pdf

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, *54*(5), 762–771.

Merat, N., Jamson, H. A., Lai, F., & Carsten, O. (2014). Human factors of highly automated driving: Results from the EASY and CityMobil Projects. In G. Meyer & S. Beiker (Eds.), *Road Vehicle Automation* (pp. 113–125). Springer International Publishing.

National Highway Traffic Safety Administration. (2013). *Preliminary Statement of Policy Concerning Automated Vehicles*. Retrieved from http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

Nees, M. A., & Walker, B. N. (2009). Auditory interfaces and sonification. In C. Stephanidis (Ed.), *The Universal Access Handbook* (pp. 32.1–32.15). Boca Raton, FL: CRC Press.

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. In P. Delucia (Ed.), *Reviews of Human Factors and Ergonomics* (pp. 58–99). Thousand Oaks, CA: Sage Publishing/Human Factors and Ergonomics Society.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*(4), 147.

Seagull, F. J., Wickens, C. D., & Loeb, R. G. (2001). When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, pp. 1395–1399).

van den Beukel, A. P., & van der Voort, M. C. (2013). Retrieving Human Control After Situations of Automated Driving: How to Measure Situation Awareness. In J. Fischer-Wolfarth & G. Meyer (Eds.), *Advanced Microsystems for Automotive Applications 2013* (pp. 43–53). Springer International Publishing.

Walker, B. N., & Nees, M. A. (2011). Theory of sonification. In T. Hermann, A. Hunt, & J. Neuhoff (Eds.), *Principles of Sonification: An Introduction to Auditory Display* (pp. 9–39). Berlin, Germany: Logos Publishing House.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177.

## Appendix A: Experimental Instructions

### Main experiment instructions

"In this experiment, you will go for a ride in a self-driving car. The ride is a simulation in the form of a video. You will see the view from the driver's perspective in the car. During the ride, you will receive messages about the status of the vehicle and the driving scenario. The car will take care of all of the driving. Because some self-driving cars may need you to take over driving from the car, you should pay attention and try to remember the status messages."

### Word search instructions (before each video)

"You will also do a word search during your ride in the vehicle. Please find the paper to your left labeled **"Word Search A."** Do not begin the word search until your car ride begins. The ride will begin after you click below to continue. You should find as many words as possible during your ride while also maintaining awareness of the vehicle and its surroundings. Your goal is to finish the word search during the ride. **Be sure you have "Word Search A" in front of you right now. Do not write on the paper except to do the word search. Find as many words as possible during your ride and maintain awareness of the vehicle and its surroundings. Click below to begin."**

### Memory prompt

"Please tell me everything you can remember about this last scenario riding in the self-driving car. Describe any hazards you encountered. Describe any maintenance that is needed. Describe any driving conditions that you expect to encounter. Please be as specific as possible in all responses. For example, if you experienced a hazard, describe what the hazard was, and also where it was in relation to the car (e.g., on the right or left; in front of or behind the vehicle)."

### Likert rating questions

1. "In some scenarios, drivers may need to resume manual control of a self-driving car. On a scale of 1 to 7, please rate how prepared you would feel if you had to take control of the vehicle at the end of the scenario you just experienced. A rating of **"1"** means you feel **completely unprepared** to resume control of the vehicle. A rating of **"7"** means that you feel **completely prepared** to resume control of the vehicle."
2. "How useful were the messages provided by the vehicle? A rating of **"1"** means you thought the messages were **completely useless**. A rating of "**7"** means that you thought the messages were **completely useful**."
3. "How annoying were the messages provided by the vehicle? A rating of **"1"** means you thought the messages were **not annoying at all**. A rating of "**7"** means that you thought the messages were **the most annoying messages you can imagine**."

**Short Biographies**

Michael A. Nees is an Assistant Professor and directs the Human Factors, Perception, and Cognition Lab in the Department of Psychology at Lafayette College. He was awarded a Ph.D. from the Georgia Institute of Technology in 2009. His research examines applied and theoretical aspects of auditory perception and cognition, auditory and multimodal displays, assistive technologies, and auditory displays for in-vehicle technologies.

Benji Helbein is a research assistant in the Human Factors, Perception, and Cognition Lab in the Department of Psychology at Lafayette College. He is a mechanical engineering major and a member of the class of 2016. His research interests include human factors and autonomous vehicles.

Anna L. Porter is an EXCEL scholar and a research assistant in the Human Factors, Perception, and Cognition Lab in the Department of Psychology at Lafayette College. She is a psychology major and a member of the class of 2016. Her research interests include human factors and health psychology.