"Watch Out!" Semi-Autonomous Vehicles Using Assertive Voices to Grab Distracted Drivers' Attention

Priscilla N. Y. Wong
Duncan P. Brumby
ngoi.wong13@ucl.ac.uk
d.brumby@ucl.ac.uk
UCL Interaction Centre
University College London
London, UK

Harsha Vardhan Ramesh Babu
Kota Kobayashi
harsha@ustwo.com
kota@ustwo.com
ustwo London
London, UK

ABSTRACT

Semi-autonomous vehicles are gradually appearing on our roads, and have already been involved in several high-profile accidents. These accidents usually occurred because the driver did not intervene in time when the automated system failed. An important issue for the design of future semi-autonomous vehicles is identifying effective methods for alerting drivers to critical events that require their intervention. To investigate this, we report the results of a lab-based simulator study in which participants had to respond to driving events while also playing an immersive mobile game on a phone. Results show that a more assertive voice alerting the driver to driving events resulted in faster reaction times and was perceived as more urgent than a less assertive voice, regardless of how immersed the driver was in the mobile game. These results suggest that the designers of future semi-autonomous vehicles should use assertive voice commends to alert drivers to critical events that require their intervention.

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312838

KEYWORDS
Autonomous vehicles; Voice assistant; Assertiveness

INTRODUCTION

It is hoped that autonomous vehicles (AVs) will make our roads safer by removing human error, which accounts for around 90% of traffic accidents [7]. Despite the optimistic view of the future of AVs [10], the public is still in doubt of the reliability of AVs [2]. This problem was accentuated by the fatal accident involving an Uber AV in 2018 that was overemphasised in the media where the footage of the accident was captured and showed that the driver was distracted by her smartphone at the moment of impact [9]. Later investigation showed that the emergency braking system was turned off and no feedback was given to warn the driver of it. This high-profile accident highlights that current-generation AVs may not adequately communicate the status of the automated system to the driver. To avoid drivers becoming too complacent about the still unreliable automated system, we investigated the use of voice commands in a semi-AV environment to help drivers stay focused on the road.

Semi-AVs are typically known nowadays as the ones that become autonomous on some proportion of the road (e.g., highways) and not on others e.g., the Tesla's Enhanced Autopilot. Pre-alerts for the transitions between the modes (i.e., handovers) have been extensively explored in the HCI literature to draw drivers' attention in a semi-AV environment. They are used to intervene drivers' ongoing task (e.g., playing a mobile game) and to prepare drivers for the vehicles' transition from automated driving to manual driving or vice versa [8, 11]. However, it should be noted that currently handovers only happen when the system decides to make the transition and therefore the system does not account for the potential risks that it might not be able to pick up. Therefore, we explore the novel idea of frequent updates about low level hazards on the road, which may potentially encourage drivers to be more aware of their surroundings.

Previous research has explored how varying the sense of urgency in a voice command affects drivers' behaviors in handover situations. By adopting words such as "Danger", "Warning" and "Caution" which were found different on their level of perceived urgency [1] and varying in how urgently spoken warnings were, Politis et al. [8] found that drivers responded quicker to the urgent pre-alerts. However, urgency alone does not convey an anthropomorphic trait which is more trusted by people [6].

Assertiveness may be a related humanized trait that would potentially exert trustworthiness and may be equally as effective in drawing driver's attention. Large and Burnett [4] investigated different navigation voices, which were readily available on the market, and studied their assertiveness. Participants were more likely to choose a voice that was rated more assertive to be their everyday navigation voice. They also found assertive voices more trustworthy. However, assertiveness has not been investigated in semi-AV environment before. We provide insights into how this humanized trait may impact the behavior of a distracted driver in a semi-AV. We expect that by eliciting urgency and more humanized and natural way of communication, assertive voices will result in faster reactions.

Table 1: Assertive and Non-Assertive Execution Commands

	Non-assertive	Assertive	t
Indicate Left(L)/ Right(R)	*Indicate L/R if possible.*	*Look up! Action to indicate L/R is needed.*	6.18*
Braking	*Please apply the brakes.*	*Watch out! Brake immediately.*	3.08*
Slow Down	*I suggest you slow down gradually.*	*You need to slow down immediately.*	5.04*

Note: * indicates $p < 0.05$, $df = 14$.

Table 2: Voice Rating Questionnaire (7-point Likert Scale)

- Do you think that this voice is...?
  - Clear
  - Distracting from the game
  - Trustworthy
  - Assertive
  - Friendly
  - Annoying
  - Entertaining
  - Urgent
- Does this voice make it feel like there is somebody with you?
- How likely would you be to use this as your everyday car assistant voice?
- How likely would you be to use this on a one-off occasion such as a day-out?
- What is your overall rating of this voice?

METHOD

Participants

Twenty drivers were recruited through opportunistic sampling (12 males and 8 females). The age range was from 21 to 48 years ($M = 26.3, SD = 7.3$). Six people usually drove in the UK and others mostly drove in their home countries (Canada, China, Poland, USA).

Design and Materials

A 2×2 (Game $\times$ Assertiveness) mixed factorial design was used. The between subject variable was how immersive the mobile game was that was used as a secondary task to distract the driver. Two games were selected: *Fruit Ninja* (high immersive game) and *Smart Shape* (low immersive game). The immersiveness of each game was established during a pre-study manipulation check in which 12 participants (different to those recruited for the main study) played each game on an Apple iPhone 7 Plus. After playing each game for three minutes, participants completed the Immersive Experience Questionnaire (IEQ) [3]. Participants reported significantly higher IEQ scores after playing Fruit Ninja ($M = 153.08, SD = 26.2$) than after playing Smart Shape ($M = 99.00, SD = 37.5$), $t(11) = 4.39, p = .001, d = 1.27$.

The within-subject variable was the level of assertiveness of the voice commands. Voice commands which were recorded with a British male voice were associated with driving scenarios, such as roundabouts, lane changes, T-junctions. Drivers were first given a warning at the approach of one of these driving scenarios (e.g., "Exiting roundabout ahead"), and these varied only in tone i.e., neutral or serious. This was then followed by one of three different execution commands (indicate left/right, braking, slow down), which varied both in tone and phrasing. A pre-study manipulation check was conducted in which 15 participants (different to those recruited for the main study) assessed the perceived assertiveness of these different voice commands, rating "Do you think that this voice is assertive?" on a 7-point Likert scale (1- Not at all, 7- Completely). Table 1 shows the execution commands that were used. It can be seen in the table that participants gave significantly higher ratings to voice commands used in the assertive condition than in the non-assertive condition.

Assertive and non-assertive voice commands were given in an identical set of driving videos separately. These videos ranged from two and a half minutes to five and a half minutes. Six voice commands were in each video. Videos were shown on a driving simulator (See Figure 1).

Participant's reaction time and accuracy of their response to the voice commands and their perceptions on the voices were measured using a Voice Rating Questionnaire (VRQ) [4]. Table 2 shows the complete VRQ; responses were given on a 7-point Likert scale.

Figure 1: Driving Simulator Set-up

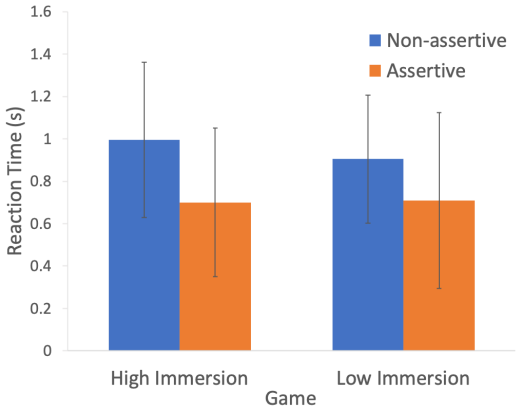

Figure 2: Reaction Time for Different Levels of Assertiveness and Immersion. The error bars represent the standard deviation of the means.

Procedure

After giving their consent and filling out their demographics, participants were positioned in front of the driving simulator in a lab room. They were told that though they were in a semi-AV environment, actions would still be required to be executed on the simulator to stay safe. They were instructed to follow the instructions given by voice commands. The to-be-executed actions are illustrated in Table 1. They were also asked to play a mobile game concurrently, which was either a lower immersion game (i.e., Smart Shape) or a high immersion game (i.e., Fruit Ninja). Before the main experiment, they were given a round of practice trials.

In the main study, participants were given a set of four videos. Assertive voice commands were randomly allocated to half of the videos and non-assertive to the other half. The order of the videos was randomized and counterbalanced. Reaction times and accuracies in response to the voice commands were recorded. Participants filled in a VRQ after each video.

RESULTS

Participants' first game pad response to each voice command was recorded as their reaction time. If participants' responses was consistent with the instruction, then it was coded as accurate. For statistical analysis a 2×2 (Immersion $\times$ Assertiveness) mixed factorial ANOVA was used on participants' reaction time and accuracy data. Repeated measures ANOVAs were used to evaluate survey ratings based on the Assertiveness conditions. Effects with a p value $\leq .05$ were deemed as significant.

Assertiveness

Figure 2 shows reaction time for the different conditions in the study. Results show that participants were significantly faster when an assertive voice command was used compared to when a non-assertive voice command was used, $F(1, 18) = 13.95, p = .002, \eta_p^2 = .437$. Response accuracy to these voice commands was generally very high: Mean accuracy for the non-assertive condition was between 75-100% ($M = 95\%, SD = 8.32\%$) and for assertive condition was between 92-100% ($M = 98\%, SD = 3.4\%$). There was however no significant effect of voice command on accuracy, $F(1, 18) = 3.06, p = .098, \eta_p^2 = .145$.

Game Immersion

Participants tended to have similar response times and accuracy levels regardless of whether they were playing the high or the low immersion game. Results show that there were no significant effects of game immersion on either reaction times or accuracy; all interaction effects were also not significant.

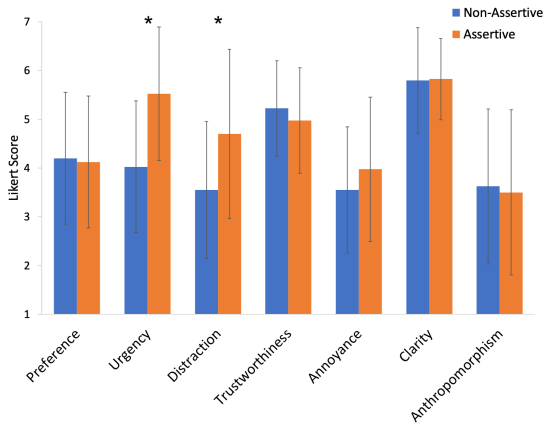

Figure 3: Likert Scale Ratings of Voice Command Related Questions. The error bars represent the standard deviations of the means.
Note: * indicates $p < 0.05$

Voice Rating Questionnaire

Figure 3 shows all comparisons of participants' voice ratings. There was a significant main effect of assertiveness for urgency, $F(1, 19) = 11.18, p = .003, \eta_p^2 = .370$, and for distraction, $F(1, 19) = 10.35, p = .005, \eta_p^2 = .353$; all other differences were not significant.

DISCUSSION

The findings in this study show that participants responded quicker to assertive voice commands than non-assertive voice commands as hypothesized despite how immersive the secondary task was. Though Large and Burnett [5] investigated the assertiveness of navigation voices, they did not take quantitative measures, such as reaction time. This study therefore provides a novel finding that assertiveness affects drivers' reaction time to low level hazard warnings. This finding suggests that assertive voices have the potential to warn drivers effectively and prevent complacency.

One explanation for our results is that assertive voice commands sounded more urgent than non-assertive voices despite conveying in a serious tone. This validates Politis et al.'s results [8] where more urgently sounded voices would result in faster reactions in driving situations. Also, wording were found effective. They were able to draw attention from drivers who were distracted to different extent. The effect of assertiveness shows commonality to that of urgency as we found that different nature of tones and wordings was able to effectively grab drivers' attention.

Due to the faster reaction time, we infer that drivers became less complacent about the automated system as it was more effective in grabbing drivers' attention. However, our findings did not include gaze durations or fixations which are more relevant measures in reflecting drivers' level of complacency. Further study which uses an eye-tracker might be needed to confirm our inference.

Despite the effectiveness of the assertive voice commands, they were not found to be more trustworthy nor more human-like than the non-assertive ones. According to Lee et al. [6], a more anthropomorphic robotic agent was able to elicit higher trust in people. Therefore, the present drivers might not have trusted assertive voice commands more because it did not convey a stronger sense of anthropomorphism. One explanation might be because the communication was not dynamic enough. Large et al. [5] showed that it was the language properties such as conversational exchanges that made interactions between drivers and interfaces natural.

As conversations involve formulating appropriate answers to other's initial response, no one conversation can be entirely duplicated. Therefore, due to this reason, conversations would have been almost impossible to control if it was manipulated in this study, highlighting one of the limitations in this study. In fact, Large et al. [5] was able to study conversations between drivers and conversational agent because they took a qualitative approach. Another concern is that the observed effect might reflect a difference in the choice of wording used (e.g., immediately vs. gradually), or the tone that

was used by the actor when voicing these different command words. Future research is needed to clarify and determine whether the results were significant due to the tone, phrasing or both.

CONCLUSION

This study successfully demonstrates the effectiveness of assertive voice commands in influencing people's speed in executing actions on a semi-AV regardless of how cognitively occupied the driver is with secondary task, e.g. playing a mobile game. The findings show how people's perceptions and behaviors change responsively to stronger and more urgent attitudes. This study offers a novel, simple and effective way for developers to influence semi-autonomous drivers' attention on the road.

REFERENCES

[1] Carryl L Baldwin and Colleen Moore. 2002. Perceived urgency, alerting effectiveness and annoyance of verbal collision avoidance system messages. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 46. SAGE Publications Sage CA: Los Angeles, CA, 1848–1852.

[2] Craig A. Giffi, Joseph Jr. Vitale, Thomas Schiller, and Ryan Robinson. 2018. A reality check on advanced vehicle technologies. *Deloitte Insights* (2018). https://www2.deloitte.com/insights/us/en/industry/automotive/advanced-vehicle-technologies-autonomous-electric-vehicles.html

[3] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies* 66, 9 (2008), 641–661.

[4] David R Large and Gary E Burnett. 2013. Drivers' preferences and emotional responses to satellite navigation voices. *International Journal of Vehicle Noise and Vibration* 9, 1-2 (2013), 28–46.

[5] David R Large, Leigh Clark, Annie Quandt, Gary Burnett, and Lee Skrypchuk. 2017. Steering the conversation: a linguistic exploration of natural language interactions with a digital assistant during simulated driving. *Applied ergonomics* 63 (2017), 53–61.

[6] Jae-Gil Lee, Ki Joon Kim, Sangwon Lee, and Dong-Hee Shin. 2015. Can autonomous vehicles be safe and trustworthy? Effects of appearance and autonomy of unmanned driving systems. *International Journal of Human-Computer Interaction* 31, 10 (2015), 682–691.

[7] National Highway Traffic Safety Administration (NHTSA). 2015. *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*. Technical Report. U.S. Department of Transport. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

[8] Ioannis Politis, Stephen Brewster, and Frank Pollick. 2015. Language-based multimodal displays for the handover of control in autonomous cars. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 3–10.

[9] Katyanna Quach. 2018. Uber robo-ride's deadly crash: Self-driving car had emergency braking switched off by design. *The Register* (2018). https://www.theregister.co.uk/2018/05/24/uber_self_driving_software_crash_report/

[10] Thomson Reuters. 2015. Google is much more optimistic than automakers about self-driving cars. *Business Insider* (2015). https://www.businessinsider.com/r-automakers-google-take-different-roads-to-automated-cars-2015-9

[11] Remo van der Heiden, Shamsi T Iqbal, and Christian P Janssen. 2017. Priming Drivers before Handover in Semi-Autonomous Cars. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 392–404.