International Journal of Mobile Human Computer Interaction
Volume 9 • Issue 3 • July-September 2017

Using Multimodal Displays to Signify Critical Handovers of Control to Distracted Autonomous Car Drivers

Ioannis Politis, School of Computing Science, University of Glasgow, Glasgow, UK
Stephen Brewster, School of Computing Science, University of Glasgow, Glasgow, UK
Frank Pollick, School of Psychology, University of Glasgow, Glasgow, UK

ABSTRACT

Until full autonomy is achieved in cars, drivers will still be expected to take over control of driving, and critical warnings will be essential. This paper presents a comparison of abstract versus language-based multimodal warnings signifying handovers of control in autonomous cars. While using an autonomous car simulator, participants were distracted from the road by playing a game on a tablet. An automation failure together with a car in front braking was then simulated; a rare but very critical situation for a non-attentive driver to be in. Multimodal abstract or language-based warnings signifying this situation were then delivered, either from the simulator or from the tablet, in order to discover the most effective location. Results showed that abstract cues, including audio, and cues delivered from the tablet improved handovers. This indicates the potential of moving simple but salient autonomous car warnings to where a gaming side task takes place.

KEYWORDS

Audio, Autonomous Cars, Games, Handover, Multimodal Feedback, Speech, Tactile, Tactons, Urgency, Visual, Warnings

INTRODUCTION

Autonomous cars are becoming a more and more popular topic of research, although not without concerns from the public over the safety of this new technology (Kyriakidis, Happee, & Winter, 2014). To address such worries, there is careful examination of road accidents involving autonomous vehicles from technology providers (Google, 2015b). This shows the importance of safety while automation is becoming more robust. Car autonomy is a staged rather than binary process, with levels of autonomy increasing as driver involvement decreases (National Highway Traffic Safety Administration, 2013; SAE J3016 & J3016, 2014). Therefore, user interfaces are required that improve safety when driver involvement is reduced but still necessary. The handover, the point of transition of control from the car to the driver, and vice versa, is a critical part of this interaction. An effective warning mechanism for such a critical case is essential, as lack of clarity over who has control of the vehicle at a given moment can be catastrophic, e.g. (Politis, Brewster, & Pollick, 2015a).

DOI: 10.4018/ijmhci.2017070101

In parallel, as vehicle automation increases, drivers are more likely to engage in tasks other than driving. Gaming is a popular activity that drivers are expected to engage in while the car is in autonomous mode, and is a topic of ongoing research, e.g. (Krome, Goddard, Greuter, Walz, & Gerlicher, 2015; Neubauer, Matthews, & Saxby, 2014). Due to the high level of concentration required by a game, a particularly demanding scenario would be attending to a critical handover while gaming. A critical handover often examined is an automation failure, since it happens unexpectedly, leaving little time to react (Gold, Damböck, Lorenz, & Bengler, 2013; Mok et al., 2015; Pfromm, Khan, Oppelt, Abendroth, & Bruder, 2015). Signifying handovers with multimodal warnings (Naujoks, Mai, & Neukum, 2014; Politis et al., 2015a), using varying message contents (Koo et al., 2014) and evaluating transition times (Gold et al., 2013; Christian Gold & Bengler, 2014) are important aspects of this critical case. However, there is no work on how critical handovers can be facilitated by multimodal warnings originating from the game area. In this study, we use an engaging tablet gaming task and test the time required to resume driving during an automation failure. Handover notifications are moved to the tablet and abstract versus language-based multimodal warnings are compared as alerts for this scenario, both being novel interventions.

Multimodal Displays Varying in Urgency

Multimodal displays have consistently shown advantages in alerting drivers to various road events. Simple spatial vibrotactile cues coming from the direction of a threat improved reaction times of drivers (Ho, Tan, & Spence, 2005). Ho, Reed & Spence (Ho, Reed, & Spence, 2007) showed added benefit when a vibrotactile cue and a car horn sound were delivered in combination. The benefit of directionality in the cues was also observed by (Serrano, Di Stasi, Megías, & Catena, 2011), who found improved recognition performance of whether a road scene was hazardous or not when cues were presented form the direction of the hazard. In our study, we present the cues either from the front, where a threat is approaching, or from the tablet on which the participants are playing a game. In this way, we evaluate the effect of warnings from the area which the participants are focused on compared to the direction of an approaching threat which they are not attentive to.

In terms of design, reflecting the urgency of the event in the warning has repeatedly shown benefits. Politis, Brewster, & Pollick (2014a, 2013, 2014b) used all unimodal, bimodal and trimodal combinations of cues varying in urgency to achieve lower response times and higher perceived urgency for more urgent cues. In this paper, we use these results to design multimodal warnings that convey the increased urgency needed for a critical situation. Further, Politis, Brewster & Pollick (2015b) extended the comparison in all unimodal, bimodal and trimodal combinations of abstract and language-based warnings and found an advantage of abstract cues in terms of recognition times in a non-critical situation, but no difference in terms of reaction times in the presence of a critical event. In this work, we compare highly urgent abstract and language-based multimodal warnings presented from different locations in the context of autonomous cars, a comparison that has never been attempted.

Handovers of Control in Autonomous Cars

The case of an automation failure has been studied in the past as a likely reason for a handover of control. Gold et al. (2013) investigated automation failures when drivers were distracted through a tablet side task. A pure tone and a visual icon called the drivers back to the wheel during an unexpected event. These warnings were either delivered 5 sec or 7 sec before the required handover. It was found that 5 sec was a better time to prepare drivers to resume control. Gold & Bengler (2014) extended this discussion, reporting that during a handover of control, both time (how long it takes) as well as quality (driving performance during and after resumption of control) are important issues to be considered. In our work, we use the above ideas, by simulating an automation failure that coincides with a critical event, which makes driver intervention essential. We also measure both time and quality of the car to driver handovers in this scenario and present similar or quicker times of transitions achieved with the used warnings.

In terms of handover warning design, Naujoks et al. (2014) investigated a handover from the car to the driver due to an automation failure. A pure tone and a flashing icon on the dashboard created shorter handover times and better driving behaviour when combined compared to the icon alone. Telpaz, Rhindress, Zelman, & Tsimhoni (2015) used a haptic seat displaying the position of an approaching vehicle from the back using a mapping with tactile alerts. The use of the seat along with a simple audio and visual indication improved handover times and satisfaction compared to the absence of the seat. Walch Lange, Baumann, & Weber (2015) used speech and text to alert drivers about handovers during a sharp curve, when a vehicle was blocking the road or when there was no hazard. In all cases except the sharp curve, 3 sec was an adequate time for a safe handover in terms of response times and comfortable in terms of participant responses. Politis et al. (2015a) made a detailed evaluation of audio, visual and tactile language-based warnings to signify handovers between the car and a distracted driver who was playing a game. They found that more urgent handover warnings were identified as such in urgency ratings and created shorter handover times. Visual warnings presented from the driving simulator caused slower reaction times, since visual attention was directed to a tablet game and participants often missed these cues. This led to a dangerous situation where the drivers were not aware that they had control. We address the issue discovered in Politis et al. (2015a) by moving the warnings to the focus of interaction. All warnings are presented from either the simulator or a tablet. Further, we compare the effectiveness of cues varying in message content, by using abstract or language-based warnings signifying a critical handover. This is a novel approach both in terms of warning location as well as message content.

Driver Distraction in Autonomous Cars

When a car is partially or fully autonomous, the absence of a driving task allows the driver to engage in other activities. Neubauer et al. (2014) found that the engagement with a trivia game or a phone conversation during a drive that involved automated and manual parts reduced fatigue and improved driving metrics when participants had control of the vehicle. However, engagement with a secondary task created higher reaction times to an unexpected event. Miller et al. (2015) also found that media consumption on a mobile device reduced fatigue of drivers and did not slow handovers when invited back to driving shortly before entering an area of increased hazard. Their handover warnings were visual and auditory, coming either from the tablet or the dashboard, but no comparison between locations was made.

Other than the studies described above, interaction with games is a little explored topic, with available studies mainly in cases where the car is fully autonomous and no intervention is expected (see for example (Krome et al., 2015; Terken et al. 2013)). Further, resuming control with the help of warnings originating from the area of the gaming interaction as opposed to the car has not been studied. We address this gap in our experiment by investigating a set of urgent multimodal warnings designed for an automation failure, requiring immediate driver attention. We deliver the warnings either from the simulator, which is the most common approach in the literature or from a tablet where the user is playing a game as a secondary task. Different warning designs are used, utilising abstract and language-based cues never before compared in this setting.

WARNING DESIGN

The warnings designed addressed a highly urgent situation, where a car would hand over control to the driver during a critical event, due to an automation failure. The abstract warnings consisted of pure tones, colours or vibrations delivered as repeated pulses, as in Politis et al. (2013). In line with this study, the warnings had an increased pulse rate to convey high urgency. They consisted of 8 pulses having 0.1 sec single pulse duration and interpulse interval and had 1.5 sec duration. The auditory warning varied additionally in base frequency (1000 Hz) in line with Edworthy, Loxley, & Dennis (1991). As in Politis et al. (2013), the visual warning also varied in colour and was Red

(RGB(255,0,0)). The tactile warning had a frequency of 150 Hz, the nominal centre frequency of the ELV-1411A Tactor[1], used to deliver vibrational messages. In line with Politis et al. (2015b), the abstract audio and tactile cues had the same intensity as the speech cues. Simultaneous delivery of unimodal signals was used in the multimodal cues, creating a synchronous effect of sound, vibration and visuals.

For the language-based warnings, the speech message used was taken from Politis et al. (2015a). It was a high priority message according to Lee, Bricker, & Hoffman (2008) and SAE (2002), with the word "Danger!" added in the beginning to increase perceived urgency, in line with Baldwin (2011) Edworthy, Hellier, Walters, Clift-Mathews, & Crowther (2003) and Politis et al. (2014b). At the end of the message an explanation that the driver had vehicle control was added, as in Politis et al. (2015a). The resulting message was "Danger! Collision Imminent. You have control!". The message was spoken urgently by a female actor, as if a loved one was in danger, in line with Edworthy et al. (2003), and Politis et al. (2014b, 2015a). It was modified to remove pauses and decrease duration. The resulting duration of the message was 2.7 sec, with a peak of -0.0 dBFS and an average frequency of 371 Hz. The tactile equivalent of the audio warning was a Speech Tacton delivered with the ELV-1411A Tactor, which was constructed as described in Politis et al. (2014b). The duration of the tactile warning was also 2.7 sec, the peak -0.0 dBFS and the average frequency 370 Hz. The visual warning was the text of the warning displayed for the duration of the utterance in Red (RGB(255,0,0)), as in Politis et al. (2015b).

All warnings were delivered either from the driving simulator in front of the participant or from a Windows tablet to the right of the driver, as will be described below. We presented the abstract and language-based warnings in all combinations of the audio, visual and tactile modalities: Audio (A), Visual (V), Tactile (T), Audio + Visual (AV), Audio + Tactile (AT), Tactile + Visual (TV), Audio + Tactile + Visual (ATV). As a result 28 different cues were created, 7 cues with all modalities (A, T, V, AT, AV, TV, ATV) × 2 types of Information (Abstract, Language-based) × 2 Locations (Simulator, Tablet). These warnings were evaluated in an experiment looking at reaction times and driving metrics of participants when exposed to the cues.

EXPERIMENT

An experiment was conducted to investigate how quickly and effectively participants would be able to resume control of an autonomous car, while distracted by a game on a tablet. We used a task similar to Politis et al. (2015a), where a periodical transition back to driving would be enforced due to an unexpected critical event. In line with Politis et al. (2015a), we investigated how quickly and accurately such a transition would happen and how it would affect driving metrics. However, we used only critical warnings varying in design, and delivered from different locations. The reason was that the focus of this study was critical handovers as a result of an automation failure, in which Politis et al. (2015a) did not primarily focus, and, as described before, there is very little research on how to design such warnings. Investigating how delivering the cues from the tablet would improve results was not addressed in Politis et al. (2015a) or in any other study on the topic. As a result, a 7×2×2 within subjects design was used, with Modality, Information and Location as the independent variables and Response Time (RT), Response Accuracy (RA) and Lateral Deviation after Handover (LDaH) as the dependent ones. As in Politis et al. (2015a), RT would be a measure of alertness when resuming driving, RA would indicate any missed responses and LDaH would show the level of distraction when resuming driving (lower LDaH would indicate lower distraction, see also Lindgren, Angelelli, Mendoza, & Chen (2009), and Liu (2001).

The expectations forming the hypotheses of this study were firstly that the modalities used in the warnings would affect responses. As in Politis et al. (2015a), multimodal warnings were expected to be more effective that unimodal ones, while the visual displays on the simulator were expected to be problematic. In terms of Information, in line with Politis et al. (2015b), it was expected that abstract cues would create quick responses, while language-based ones would affect driving less. Finally, the location of the tablet was expected to affect responses positively, since it would be located in the participants' field of view, in line with Miller et al. (2015). As a result, there were the following hypotheses:

- RT will be influenced by Modality (H_{1a}), Information (H_{1b}) and Location (H_{1c});
- RA will be influenced by Modality (H_{2a}), Information (H_{2b}) and Location (H_{2c});
- LDaH will be influenced by Modality (H_{3a}), Information (H_{3b}) and Location (H_{3c}).

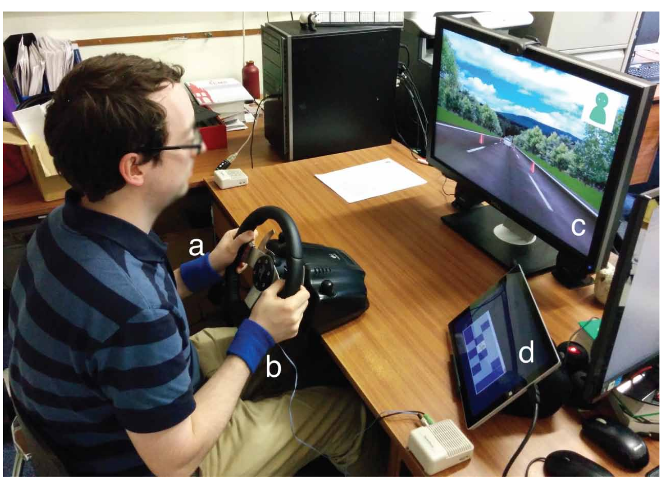
Participants and Equipment

Twenty participants (7 female) aged between 20 and 45 years ($M = 25.25$, $SD = 5.67$) took part in the experiment. There were 17 University students and 3 private employees. They had a valid driving license and between 1 and 24 years of driving experience ($M = 6.18$, $SD = 5.50$). All were right handed and reported normal vision and hearing. The experiment took place in a University room, where participants sat in front of 27-inch Dell 2709W monitor, a PC running the driving simulator, a Microsoft Surface Pro 3[2] tablet PC running a game (placed to the right of the driver) and a Logitech G27 gaming wheel and pedals. The driving simulator software depicted a rural road scene with a curvy road and a car in front, which has been used in many studies, e.g. Zhao, Brumby, Chignell, Salvucci, & Goyal (2013). See Figure 1 for the setup of the experiment.

The tablet was running the Concentration memory game, used also in Politis et al. (2015a) and based on Warnock, McGhee-Lennon, & Brewster (2011) (see Figure 2b). It was a simple card matching game on a 3×8 grid. This game has a well-defined set of performance metrics and requires high levels of concentration. As it is likely that drivers will occupy themselves with other activities while an autonomous vehicle is driving itself, this task was chosen so as to decrease their engagement with driving and create a more challenging handover.

Three sounds were added to the game in order to increase auditory distractions. The first sound was a 100 msec long 440 Hz tone (note A_4) that sounded every time the participant touched the tablet screen. The second sound was a 100 msec 330 Hz tone (note E_4) that sounded every time a pair of pictures revealed was not a match. The third one was an Earcon with three tones (100 msec of 262 Hz followed by 100 msec of 330 Hz followed by 100 msec of 392 Hz – notes C_4, E_4, G_4). This sounded every time a pair of cards was matched. In this way, an additional sensory load was created

Figure 1. The setup of the experiment with the tactile wristbands (a, b), the driving simulator (c) and the tablet (d)



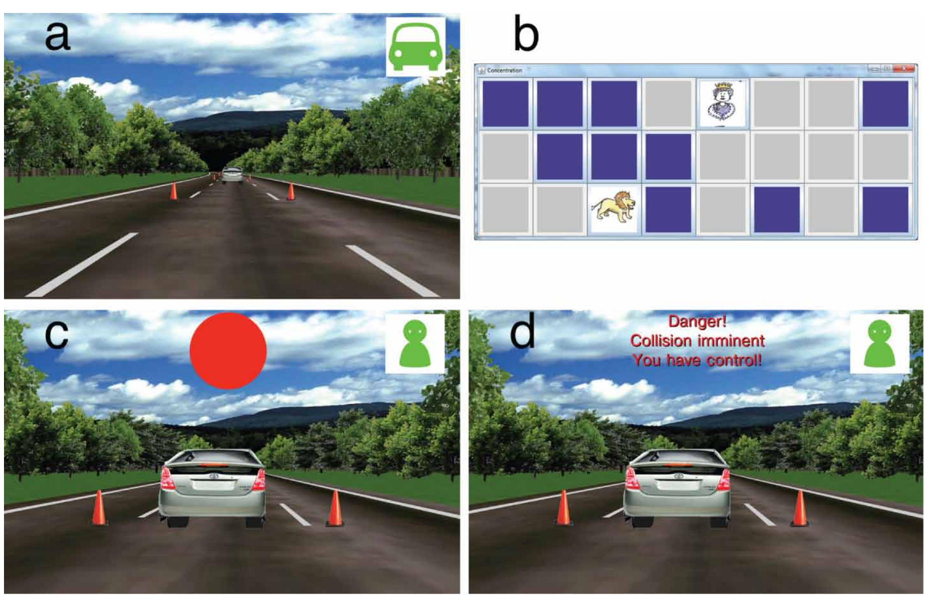
in the audio modality, which was not present in Politis et al. (2015a). Also, ecological validity was increased, since sound effects are frequently found in games.

Auditory cues and game sounds were displayed through three Betron portable speakers[3], one located behind the screen (for the Simulator warning location) two behind the tablet (for the Tablet warning location and one for the game sounds). Tactile cues were displayed through a wristband on both of the participants' hands. The right hand was used for the Tablet location, since it was the hand interacting with the tablet and the left hand for the simulator location, being the hand remaining on the steering wheel. Pilot studies showed that this mapping was clear to participants and they were also familiarized with it during training with the cues (see below). Visual abstract cues were displayed through Red circles that flashed in the top central area of the monitor (for the Simulator location, see Figure 2c) or the tablet (for the Tablet location, see Figure 2b), and were sized 400×400 pixels (about 12×12 cm for the monitor and 5×5 cm for the tablet). Visual language-based cues used Red text displaying the words from the speech warning, which appeared once and for as long as the warning was uttered in the top central area of the screen, and was sized 228×700 pixels (about 17×7 cm for the monitor and 7×3 cm for the tablet, see Figure 2d, 2b). The visual cues did not obstruct the lead car on the monitor or the game on the tablet.

Procedure

After being welcomed and explained the experimental procedure, the 28 cues were displayed in a random order to participants for familiarization. For each cue, they could either repeat it or go to the next when they were familiar with it. Afterwards, they were presented with the driving simulator software and the game to familiarize themselves. In the main experiment, in line with Politis et al. (2015a), participants were asked to focus on the game, unless interrupted by a warning. They were able to use their right hand to play the game on the tablet, which was placed on a stand to the right of the simulator. This would be a standard setup for left-hand drive car. If all cards in a grid were matched,

Figure 2. The driving simulator with the participant's car in autonomous mode, as indicated visually on the top right of the screen, and the car in front driving at a safe distance (a). The tablet game with some pairs already matched, indicated in grey (b). The handover situation, where the car in front brakes suddenly and the automation fails on the same time. In this case control is handed to the driver, as indicated visually on the top right of the screen (b,c). This handover is signified through an abstract warning (the visual warning is depicted in c) or a language-based warning (the visual warning is depicted in d).



the game would reload with a new set of cards chosen randomly. While playing the game, the car was in autonomous mode in the centre of the lane at a speed of around 60 mph. The car simulated Level 3 Automation (see NHTSA (2013)) not requiring continuous intervention, but expecting availability for occasional control (see Figure 2a for a screenshot of the simulator in autonomous mode).

At random intervals of any integral value between (and including) 27–32 sec (in line with Politis et al. (2015a)) a warning was presented. In this case, control was passed to the driver (see Figures 2c, 2d). This simulated automation failures the vehicle could not correct and therefore a switch to manual mode was needed. To create a more critical situation, the car in front started braking at the same time as the presentation of the warning, as in Politis et al. (2015a). Participants were then handed control and were instructed to brake immediately with their right foot and return to safe driving. Once the participant braked, the car in front would advance away from the participant's car. It should be noted that the interval of 27–32 sec had limited ecological validity, since critical events are expected to occur less frequently. However, it was found necessary in order to be able to evaluate all the different cues designed. Other studies have used similar or shorter intervals, e.g. Ho et al. (2005).

To manage experimental length, all abstract warnings were presented in one block of the experiment and all language-based ones in another, with the order of blocks counterbalanced across participants and with a small break between them. Each warning was presented twice in each block, resulting in a total of 56 presentations for both parts (7 Modalities × 2 types of Information × 2 Locations × 2 presentations). When back to driving, participants were able to steer using the wheel for 10 sec (there was no need to use the accelerator pedal).

During this period, they were asked to stay in the centre of the lane. After 10 sec, the car automatically took over from the participant, initiating the next trial. On the top right of the screen, a car icon would be displayed when the car was in autonomous mode or a person icon for manual mode (see Figure 2a, 2c, 2d).

Response time (RT) was calculated from the onset of a stimulus until the participant pressed the brake pedal. If participants did not respond to a cue, their response accuracy (RA) was 0, otherwise it was 1. Their Lateral Deviation after Handover (LDaH) was the RMSE of their lane position values, logged for 10 sec after the onset of a stimulus and start of the braking event of the lead car. The value of 10 sec was chosen since it has shown to be an adequate time to come back to driving in handover situations (Merat, Jamson, Lai, Daly, & Carsten, 2014; Politis et al., 2015a). The experiment lasted about 45 minutes and participants were then debriefed and paid £6.

RESULTS

Response Time

The data of one participant were excluded due to software issues. For the rest of the participants there were 1064 trials in total. If participants did not respond to a cue (which was the case in 14 trials – 1.3%), their RT was adjusted to the maximum available time to respond, 10 sec, to allow a three factor ANOVA analysis.

Data for RT were analysed using a three-way repeated measures ANOVA, with Modality, Information and Location as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(2.11,78.19) = 34.95$, $p < 0.001$). Contrasts revealed that V caused slower responses compared to all other modalities, see Figure 3a ($F(1,37) = 27.42$, $r = 0.65$, $p < 0.001$). Further, AV, AT, ATV and A created quicker responses compared to T and V ($F(1,37) = 17.45$, $r = 0.57$, $p < 0.001$), but not compared to TV. As a result, H_{1a} was accepted. See Table 1 for pairwise comparisons between modalities for Response Time. There was a significant main effect of Information, indicating that abstract cues caused faster responses than language based-ones ($F(1,37) = 20.50$, $r = 0.60$, $p < 0.001$). As a result, H_{1b} was accepted. Finally, there was a significant main effect of Location, indicating that warnings from the tablet caused faster reaction times than simulator ($F(1,37) =$

Table 1. Pairwise comparisons between modalities for Response Time. The significance (p) values are reported after Bonferroni corrections.

	AV	AT	ATV	A	TV	T	V
AV		.845	.576	.689	**.004**	**.001**	**.000**
AT	.845		.721	.700	**.002**	**.001**	**.000**
ATV	.576	.721		.822	**.009**	**.002**	**.000**
A	.689	.700	.822		.064	**.000**	**.000**
TV	**.004**	**.002**	**.009**	.064		.134	**.000**
T	**.001**	**.001**	**.002**	**.000**	.134		**.000**
V	**.000**	**.000**	**.000**	**.000**	**.000**	**.000**	

12.62, $r = 0.50$, $p < 0.01$). As a result, H_{1c} was accepted. See Table 2 for values of RT across all factors and Figure 3a for values across modalities.

There was a significant interaction between Modality and Information ($F(1.87,69.13) = 21.04$, $p < 0.001$), indicating that the disadvantage of the V modality was stronger in language-based warnings ($F(1,37) = 22.11$, $r = 0.61$, $p < 0.001$). There was a significant interaction between Modality and Location ($F(2.19,81.06) = 23.14$, $p < 0.001$), indicating that T warnings created quicker responses when coming from the simulator ($F(1,37) = 7.59$, $r = 0.41$, $p < 0.01$), while the observed disadvantage of V warnings was stronger when coming from the simulator ($F(1,37) = 33.25$, $r = 0.69$, $p < 0.001$). There was a significant interaction between Information and Location, indicating that the observed disadvantage of language-based cues was stronger when coming from the simulator compared to the tablet ($F(1,37) = 28.30$, $r = 0.66$, $p < 0.001$). Finally, there was an interaction between Modality, Information and Location ($F(2.35, 87.03) = 19.99$, $p < 0.001$), indicating that when coming from the simulator, the language-based V cues showed a disadvantage compared to TV cues, while when coming from the tablet the abstract V cues showed an advantage compared to TV ones ($F(1,37) = 30.54$, $r = 0.67$, $p < 0.001$). See Figure 3c for the interaction between Modality and Information and 3.e for the interaction between Modality and Location for RT.

Response Accuracy

Data for RA were treated as dichotomous and analysed with Cochran's Q tests. It was found that modality V was less accurate compared to AT, AV, TV and ATV ($Q(1) = 11.00$, $p < 0.01$) and also compared to T ($Q(1) = 9.00$, $p < 0.01$) and A ($Q(1) = 10.00$, $p < 0.01$). Abstract cues were more accurate than language-based ones ($Q(1) = 8.00$, $p < 0.01$). Finally, cues delivered through the tablet were more accurate than the simulator ($Q(1) = 8.00$, $p < 0.01$). As a result, H_{2a}, H_{2b} and H_{2c} were accepted. See Table 2 for values of RA across all factors.

Lateral Deviation

There were 1120 trials for LDaH, since no data were excluded for this metric. Data for LDaH were analysed using a three-way repeated measures ANOVA, with Modality, Information and Location as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(1.78,69.37) = 13.83$, $p < 0.001$). Contrasts revealed that V warnings created higher LDaH values compared to all other modalities ($F(1,39) = 16.76$, $r = 0.55$, $p < 0.001$). As a result, H_{3a} was accepted. There was a significant main effect of Information, revealing that language-based warnings created higher LDaH than abstract ($F(1,39) = 7.03$, $r = 0.39$, $p < 0.05$). As a result, H_{3b} was accepted. See Table 3 for pairwise comparisons between modalities for Lateral Deviation after Handover. There was a significant main effect of

Table 2. The mean values, of Response Time (RT), Lateral Deviation after Handover (LDaH) and Response Accuracy (RA). For RT and LDaH values in brackets indicate lower – upper 95% confidence intervals.

Measure	Factor		Mean
RT (sec)	**Modality**	A	**1.03** (0.88 – 1.18)
		T	**1.36** (1.13 – 1.59)
		V	**2.24** (1.89 – 2.60)
		AT	**1.01** (0.93 – 1.09)
		AV	**1.00** (0.92 – 1.08)
		TV	**1.19** (1.04 – 1.33)
		ATV	**1.01** (0.94 – 1.09)
	Information	Abstract	**1.11** (0.98 – 1.23)
		Language-based	**1.42** (1.27 – 1.56)
	Location	Simulator	**1.35** (1.21 – 1.49)
		Tablet	**1.18** (1.07 – 1.27)
LDaH (m)	**Modality**	A	**0.96** (0.67 – 1.24)
		T	**1.10** (0.71 – 1.47)
		V	**2.10** (1.43 – 2.78)
		AT	**0.87** (0.66 – 1.09)
		AV	**0.81** (0.63 – 0.99)
		TV	**0.93** (0.73 – 1.13)
		ATV	**0.84** (0.64 – 1.03)
	Information	Abstract	**0.94** (0.68 – 1.21)
		Language-based	**1.23** (0.93 – 1.53)
	Location	Simulator	**1.20** (0.90 – 1.50)
		Tablet	**0.97** (0.74 – 1.20)
RA	**Modality**	A	**99.3%**
		T	**98.7%**
		V	**92.8%**
		AT	**100%**
		AV	**100%**
		TV	**100%**
		ATV	**100%**
	Information	Abstract	**99.4%**
		Language-based	**97.9%**
	Location	Simulator	**97.9%**
		Tablet	**99.4%**

Location revealing that, when warnings were coming from the tablet, LDaH was lower compared to when coming from the simulator ($F(1,39) = 10.18$, $r = 0.45$, $p < 0.01$). As a result, H_{3c} was accepted. See Table 2 for values of LDaH across all factors and Figure 3b for values across modalities.

There was a significant interaction between Modality and Information ($F(1.63,63.73) = 12.01$, $p < 0.001$), revealing that the observed disadvantage of V warnings was mainly present in language-

Table 3. Pairwise comparisons between modalities for Lateral Deviation after Handover. The significance (p) values are reported after Bonferroni corrections.

	AV	ATV	AT	TV	A	T	V
AV		.512	.387	**.029**	.151	.059	**.000**
ATV	.512		.641	.075	.233	.058	**.000**
AT	.387	.641		.416	.481	.180	**.000**
TV	**.029**	.075	.416		.805	.236	**.000**
A	.151	.233	.481	.805		.265	**.000**
T	.059	.058	.180	.236	.265		**.000**
V	**.000**	**.000**	**.000**	**.000**	**.000**	**.000**	

based warnings ($F(1,39) = 12.11$, $r = 0.49$, $p < 0.01$). There was a significant interaction between Modality and Location ($F(1.53,59.60) = 11.89$, $p < 0.001$), revealing that the observed disadvantage of V warnings was mainly present when they were coming from the simulator ($F(1,39) = 13.02$, $r = 0.50$, $p < 0.01$). There was a significant interaction between Information and Location, revealing that the observed disadvantage of language-based warnings was mainly present when they were coming from the simulator ($F(1,39) = 16.25$, $r = 0.54$, $p < 0.001$). Finally, there was a significant interaction between Modality, Information and Location ($F(1.76,68.80) = 11.57$, $p < 0.001$), revealing that while for warnings coming from the simulator, modality V created higher LDaH for language-based warnings, when warnings were coming from the tablet the disadvantage was mainly present in language-based T cues ($F(1,39) = 14.92$, $r = 0.53$, $p < 0.001$). See Figure 3d for the interaction between Modality and Information and 3.f for the interaction between Modality and Location for LDaH.

DISCUSSION

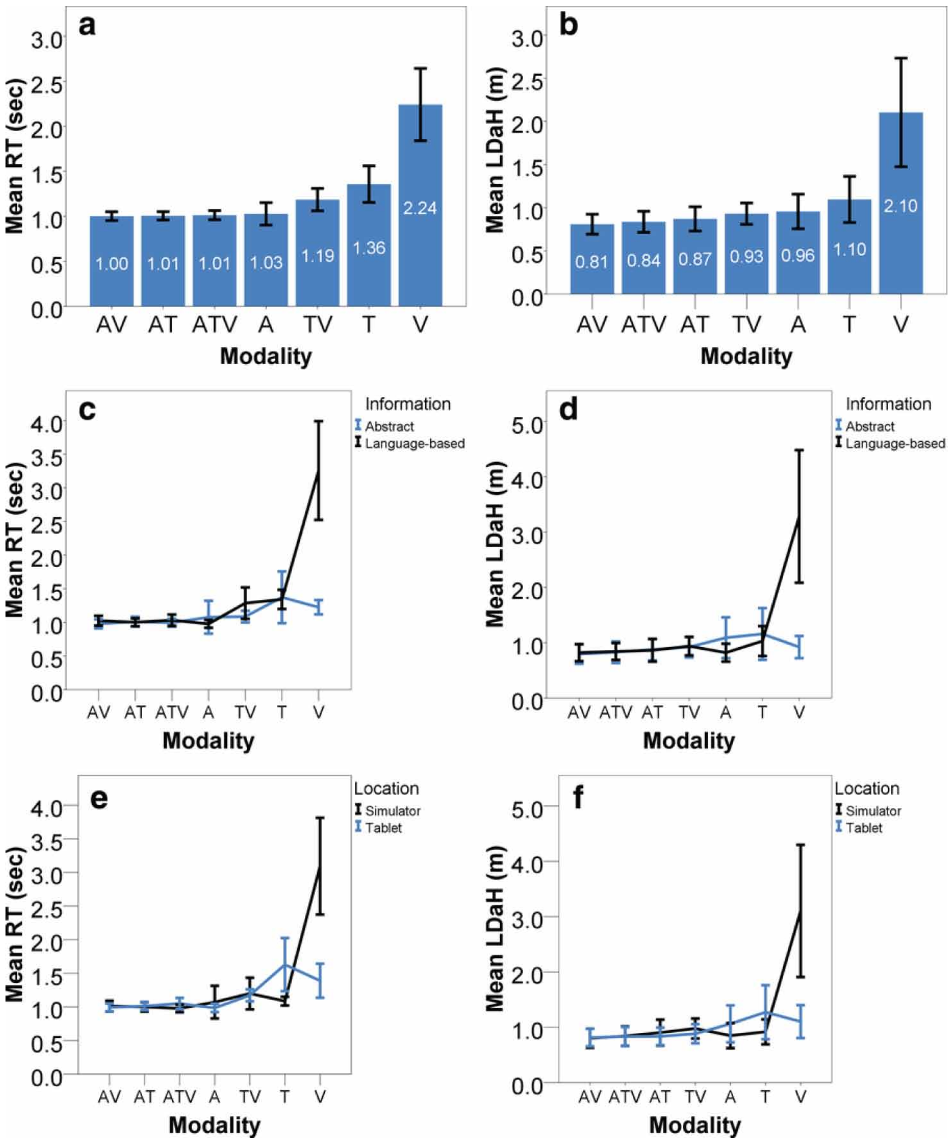
The results of the experiment confirmed the observed limitation of visual language-based cues coming from the simulator in Politis et al. (2015a). These cues created the longest response times, the least accurate responses and disturbed the driving the most. However, the intervention of this study, i.e. moving the cues from the simulator to the tablet and adding abstract cue designs, positively influenced metrics and addressed the problem with visual cues in Politis et al. (2015a).

Response Time

A notable difference to Politis et al. (2015a) is the lower handover times observed in this study. This was partly because all handovers were critical, requiring imminent attention. Another reason could be the simplicity of the task, which in contrast with Politis et al. (2015a) was always the same and did not involve different types of responses. The order of modalities in terms of average response times was similar to Politis et al. (2015a), which, in combination with the better performance of cues including audio compared to unimodal tactile and visual cues, increases confidence in the advantage of audio cues for signifying handovers in autonomous cars. This extends the findings of previous studies (Naujoks et al., 2014; Politis et al., 2015a) by providing a more elaborate examination of warning modalities for this situation. It also introduces an extensive set of possible cues to be used as warnings during an automation failure, extending the case presented by Gold & Bengler (2014).

In terms of warning designs, language-based warnings showed a disadvantage, which was mainly observed in the simulator location and the visual modality. This confirms the findings of Politis et al. (2015a) and once again shows that the problem with the visual warnings was ameliorated by moving them to the area of the gaming interaction. This also extends findings in Politis et al. (2015b), where

Figure 3. (a) Mean Response Time (RT) and (b) mean Lateral Deviation after Handover (LDaH) across Modalities. (c) The interaction between Modality and Information for RT and (d) for LDaH. (e) The interaction between Modality and Location for RT and (f) for LDaH. Error bars indicate 95% confidence intervals.



abstract and language-based cues showed similar performance in a critical task. In our study, we examined these cues when delivered from different locations in simulated driving. It was found that abstract and language-based cues are equally effective when coming from the game location, while language-based ones present limitations when delivered away from it. This extends findings of previous studies (Naujoks et al., 2014; Telpaz et al., 2015; Walch et al., 2015) by investigating a much wider set of modalities to inform about imminent handovers. As a guideline, in a vehicle where the drivers could be inattentive to the road but still expected to intervene periodically, it would be essential to capture their visual attention. Achieving this by interrupting the game on the tablet showed good results in our study. Abstract cues also showed a possibility of capturing attention when delivered from the simulator, possibly because of their pulsating design. Investigating this further by using eye-tracking techniques would be promising. We note that, as in some previous studies (Politis et al., 2015a, 2015b), language-based cues were slightly longer and this might have created an advantage for abstract cues. However, reactions were required immediately for both cue designs, and similar results were achieved for both designs when coming from the tablet.

A further comment related to the location used for informing drivers is that tactile messages delivered on the hand that was interacting with the tablet showed a disadvantage compared to the hand that was assigned for simulator cues. Additionally, when tactile messages were delivered in combination with visual ones, the bimodal presentation was beneficial when coming from the simulator for language-based cues, but problematic when coming from the tablet for abstract cues. The fact that combining visual and tactile modalities for language-based warnings showed an advantage when coming from the simulator, reveals that this bimodal presentation may have been clarifying the message content which was not salient enough when delivered only visually, as in Politis et al. (2014b). In contrast, the limitation of unimodal tactile presentation from the tablet could reveal unfamiliarity of this type of warning, since we chose a novel location for vibration, even though participants were trained with these cues until they felt confident with them. Future studies could experiment on different locations for tablet vibrations, e.g. the finger, and with more extensive training.

The limitations of bimodal tactile and visual presentation from the tablet could also reveal a high cognitive load when being occupied with a non-driving task while still being expected to periodically return to driving. This is in line with some participants' comments, mentioning that, even when playing the game, their attention was still partially on the road. Similar effects were observed in Politis et al. (2014a) and in the visual modality when combined with other modalities. Since our study was a more demanding one, requiring attention to both the road and a game, seems to have created this effect of increased attentional demand. In line with Politis et al. (2014a), we suggest the use of a limited number of modalities in warnings unless the event to be signified is critical. Even when critical, when a warning is delivered through a tablet, we suggest a preference for audio and visual modalities.

Response Accuracy

The results of RA showed that the visual modality created the least accurate responses, in line with Politis et al. (2015a). Abstract cues and cues coming from the tablet created more accurate responses, indicating the advantage of adding a new cue design and cue location compared to Politis et al. (2015a). This further supports the guideline of using the area where interaction takes place to warn the drivers of imminent events, as well as an abstract urgent cue design. It can also inform designs of previous studies (Krome et al., 2015; Terken et al., 2013), by combining a gaming interaction in the car with more critical interventions. It is worth noting that, although the described results are significant, the overall RA is much higher than in Politis et al. (2015a) (1.3% of responses were inaccurate here, as opposed to 9.4% in that study). This improvement can be attributed to the new cue design and cue location, but also to the simpler nature of the response task.

Lateral Deviation after Handover

Results of LDaH confirm the observed disadvantage of the V warnings found in the reaction time analysis, which is also in line with previous studies (Naujoks et al., 2014; Politis et al., 2015a). The disadvantage was stronger in the simulator and language-based condition, as in RT. This again shows the benefit of this new setup, which improved LDaH, and thus reduced driver distraction during critical events requiring intervention. When coming from the tablet, language-based tactile cues showed a limitation in terms of LDaH, which is in line with the slower responses observed in RT. This highlights the caution needed when using speech Tactons unimodally, also observed in other studies (Politis et al., 2014b, 2015a, 2015b).

We stress that in the few cases where there was an absence of response, the effects would be catastrophic. This is because the vehicle would be uncontrolled, as the automation failure would have disabled autonomous driving and the enforced handover would have been missed by the driver.

Warning designers should aim to eliminate such cases by creating salient handover warnings that will be noticed by drivers.

Finally, in terms of the game performance, the results of the tablet game were as follows[4]: 142.76 sec mean time to complete one game, 0.45 Clicks per Second and 0.27 Superfluous Views per Click. These are similar to previous work (Politis et al., 2015a; Warnock et al., 2011), showing that participants were attentive to the game and confirming the demanding nature of this task, making it a good choice for use in driving experiments.

General Discussion

As evident from the results, there is potential in warning drivers not only using the conventional methods available in cars, but also at the area of attention focus. In our study this was a tablet, but one can easily imagine other locations away from the central area of attention, such as the car centre stack. Synchronising these devices with the car warning mechanism would increase saliency of warnings and enable drivers to return to driving promptly in an autonomous car. If this is not possible, the good results observed with abstract warnings coming from the simulator shows benefit in using multimodal messages to capture peripheral attention in critical situations, as also observed in Spence (2010). The saliency of cues including audio can also be used, by combining visual and audio warnings in the area of attention focus in critical cases. Future work should use shorter speech messages conveying handovers and investigate if their effectiveness compared to abstract ones will improve.

To explore further locations, future work should also explicitly compare the presentation of warnings on mobile devices versus on the centre stack, which is another possible location for playing games in autonomous cars. Systems such as Apple CarPlay (2015) and Android Auto (2015a) are gaining popularity with users and car manufacturers. These systems link mobile devices to car systems so could potentially capture and display car warnings and messages on phones or tablets in the car. This could be on a device used by the driver, or even devices used by other passengers that might be connected to the car. Our guidelines are relevant to these applications, as well as to app designers who consider an autonomous car driver as a possible part of their user group.

CONCLUSION

This paper presented a study of critical handovers in an autonomous car. Participants were occupied with a tablet game, an activity very likely to occur as drivers become less engaged on the road and driving requires less involvement. Handovers were signified by multimodal combinations of abstract and language-based cues. Delivering the warnings with abstract cues including audio and visuals from the area of the game captured visual attention when signifying a handover of control. Therefore, we suggest the utilisation of this area when a driver is distracted in an autonomous vehicle. Since in a real driving situation there may or may not be a side task, we suggest the synchronization of mobile devices used by the driver with the autonomous vehicle so that warnings and notifications from the car can be presented where the driver's attention is focused, increasing warning saliency.

REFERENCES

Apple. (2015). CarPlay. Retrieved from http://www.apple.com/uk/ios/carplay/

Baldwin, C. L. (2011). Verbal collision avoidance messages during simulated driving: Perceived urgency, alerting effectiveness and annoyance. *Ergonomics*, *54*(4), 328–337. doi:10.1080/00140139.2011.558634 PMID:21491275

Edworthy, J., Hellier, E., Walters, K., Clift-Mathews, W., & Crowther, M. (2003). Acoustic, semantic and phonetic influences in spoken warning signal words. *Applied Cognitive Psychology*, *17*(8), 915–933. doi:10.1002/acp.927

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *33*(2), 205–231. doi:10.1177/001872089103300206 PMID:1860703

Gold, C., & Bengler, K. (2014). Taking Over Control from Highly Automated Vehicles. In Advances in Human Aspects of Transportation: Part II (pp. 64–69).

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? In *HFES Annual Meeting* (Vol. 57, pp. 1938–1942). SAGE Publications. http://doi.org/doi:<ALIGNMENT.qj></ALIGNMENT>10.1177/1541931213571433

Google. (2015a). Android Auto. Retrieved from https://www.android.com/auto/

Google. (2015b). Google Self-Driving Car Project - Monthly reports. Retrieved from http://www.google.com/selfdrivingcar/reports/

Ho, C., Reed, N., & Spence, C. (2007). Multisensory In-Car Warning Signals for Collision Avoidance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(6), 1107–1114. doi:10.1518/001872007X249965 PMID:18074709

Ho, C., Tan, H. Z., & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(6), 397–412. doi:10.1016/j.trf.2005.05.002

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2014). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing*. Doi:<ALIGNMENT.qj></ALIGNMENT>10.1007/s12008-014-0227-2

Krome, S., Goddard, W., Greuter, S., Walz, S. P., & Gerlicher, A. (2015). A Context-Based Design Process for Future Use Cases of Autonomous Driving: Prototyping AutoGym. Proceedings of AutomotiveUI 2015 (pp. 265–272). ACM Press. doi:10.1145/2799250.2799257

Kyriakidis, M., Happee, R., & De Winter, J.C.F. (2014). Public opinion on automated driving : Results of an international questionnaire among 5,000 respondents.

Lee, J. D., Bricker, D. L., & Hoffman, J. D. (2008). A Dynamic Programming Algorithm for Scheduling In-Vehicle Messages. *IEEE Transactions on Intelligent Transportation Systems*, *9*(2), 226–234. doi:10.1109/TITS.2008.922876

Lindgren, A., Angelelli, A., Mendoza, P. A., & Chen, F. (2009). Driver behaviour when using an integrated advisory warning display for advanced driver assistance systems. *IET Intelligent Transport Systems*, *3*(4), 390–399. doi:10.1049/iet-its.2009.0015

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics*, *44*(4), 425–442. doi:10.1080/00140130010011369 PMID:11291824

Merat, N., Jamson, A. H., Lai, F. F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 274–282. doi:10.1016/j.trf.2014.09.005

Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction Becomes Engagement in Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1676–1680). SAGE.

Mok, B. K., Johns, M., Lee, K. J., Ive, P. H., Miller, D., & Ju, W. (2015). Timing of Unstructured Transitions of Control in Automated Driving. *Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1167–1172). IEEE. doi:10.1109/IVS.2015.7225841

National Highway Traffic Safety Administration. (2013). *Preliminary Statement of Policy Concerning Automated Vehicles*. Washington, DC.

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE '14* (pp. 431–438).

Neubauer, C., Matthews, G., & Saxby, D. (2014). Fatigue in the Automated Vehicle: Do Games and Conversation Distract or Energize the Driver? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 2053–2057). SAGE. doi:10.1177/1541931214581432

Pfromm, M., Khan, R., Oppelt, S., Abendroth, B., & Bruder, R. (2015). Investigation of Take-Over Performance of Driving Tasks by the Driver due to System Failure of Semi-Automated and Assisted Driving. *Proceedings of the 19th Triennial Congress of the IEA* (pp. 1–8).

Politis, I., Brewster, S., & Pollick, F. (2013). Evaluating Multimodal Driver Displays of Varying Urgency. Proceedings of the AuomotiveUI '13 (pp. 92–99). ACM Press; doi:10.1145/2516540.2516543

Politis, I., Brewster, S., & Pollick, F. (2014a). Evaluating Multimodal Driver Displays under Varying Situational Urgency. Proceedings of CHI '14 (pp. 4067 – 4076). ACM Press. doi:10.1145/2556288.2556988

Politis, I., Brewster, S., & Pollick, F. (2014b). Speech Tactons Improve Speech Warnings for Drivers. Proceedings of the AutomotiveUI '14 (pp. 1–8). ACM Press. doi:10.1145/2667317.2667318

Politis, I., Brewster, S., & Pollick, F. (2015a). Language-Based Multimodal Displays for the Handover of Control in Autonomous Cars. Proceedings of the Automotive UI '15 (pp. 3–10). ACM Press. doi:10.1145/2799250.2799262

Politis, I., Brewster, S., & Pollick, F. (2015b). To Beep or Not to Beep? Comparing Abstract versus Language-Based Multimodal Driver Displays. Proceedings of CHI '15 (pp. 3971–3980). ACM Press.

SAE. (2002). *Its In-Vehicle Message Priority Standard J2395*.

SAE J3016, & J3016, S. (2014). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*.

Serrano, J., Di Stasi, L. L., Megías, A., & Catena, A. (2011). Effect of directional speech warnings on road hazard detection. *Traffic Injury Prevention*, *12*(6), 630–635. doi:10.1080/15389588.2011.620661 PMID:22133340

Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*(1), 182–200. doi:10.1111/j.1749-6632.2010.05440.x PMID:20392281

Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic Seat for Automated Driving : Preparing the Driver to Take Control Effectively. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 23–30). ACM Press. doi:10.1145/2799250.2799267

Terken, Z., Haex, R., Beursgens, L., & Szostak, D. (2013). Unwinding After Work: An In-Car Mood Induction System for Semi-Autonomous Driving. *AutomotiveUI*, *246–249*, 246–249. doi:10.1145/2516540.2516571

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous Driving: Investigating the Feasibility of Car-Driver Handover Assistanc. Proceedings of Automotive UI '15 (pp. 11–18). ACM Press. doi:<ALIGNMENT.qj></ALIGNMENT>10.1145/2799250.2799268

Warnock, D., McGee-Lennon, M., & Brewster, S. (2011). The Role of Modality in Notification Performance. In Interact (pp. 572–588). Springer. doi:10.1007/978-3-642-23771-3_43

Zhao, S., Brumby, D. P., Chignell, M., Salvucci, D., & Goyal, S. (2013). Shared Input Multimodal Mobile Interfaces: Interaction Modality Effects on Menu Selection in Single-Task and Dual-Task Environments. *Interacting with Computers*, *25*(5), 386–403. doi:10.1093/iwc/iws021

ENDNOTES

1. http://www.aactechnologies.com/category/45
2. http://www.microsoft.com/surface/en-us/products/surface-pro-3
3. http://www.betrononline.com/p/9037748/betron-pop-up-portable-travel-speaker-black.html
4. The metrics used were taken from (Warnock et al., 2011) and were the average time to complete a game, the clicks per second and the superfluous views. Superfluous views indicate how many decisions (pictures tapped) were not successful. When a picture was viewed it was marked as 'seen'. Every subsequent viewing of that picture failing to create a match was a superfluous view.

Ioannis Politis is a Research Associate at the Engineering Design Centre, Department of Engineering, University of Cambridge. He is investigating Human-Computer Interaction topics for autonomous cars, funded by the Engineering and Physical Sciences Research Council and Jaguar - Land Rover. He completed his PhD at the University of Glasgow, Multimodal Interaction Group, funded by Freescale Semiconductor (now NXP). During his PhD, he investigated the use of multimodal displays varying in urgency and message content, to alert drivers of manual and autonomous cars. Before his PhD, he had worked as Usability Engineer at Philips Healthcare, and as Postgraduate Design Trainee at Eindhoven University of Technology.

Stephen Brewster is a Professor of Human-Computer Interaction in the School of Computing Science at the University of Glasgow, UK. He runs the internationally leading Multimodal Interaction Group. His research focuses on multimodal HCI, or using multiple sensory modalities and control mechanisms (particularly hearing, touch and gesture) to create a rich, natural interaction between human and computer. His work has a strong experimental focus, applying perceptual research to practical situations. A long term theme has been haptics, starting with force-feedback and more recently tactile displays. He has authored over 350 papers and is a member of the CHI Academy.

Frank Pollick is a professor of psychology at the University of Glasgow. He completed undergraduate studies in Physics and Biology at MIT, and masters studies in Biomedical Engineering at Case Western Reserve University before obtaining his PhD in Cognitive Sciences from The University of California, Irvine. He then moved to Kyoto, Japan where he worked at the Advanced Telecommunication Research Institute (ATR) as a researcher in the Human Information Processing laboratory. Since 1997 he has been at the University of Glasgow. His research interests include multisensory processing and how we understand the actions of others as well as how perceptual abilities change with expertise and conditions such as autism.