Language-Based Multimodal Displays for the Handover of Control in Autonomous Cars

Ioannis Politis[1], Stephen Brewster[2]
Glasgow Interactive Systems Group
School of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
[1]I.Politis.1@research.gla.ac.uk
[2]Stephen.Brewster@glasgow.ac.uk

Frank Pollick
School of Psychology
University of Glasgow
Glasgow, G12 8QB, UK
Frank.Pollick@glasgow.ac.uk

ABSTRACT

This paper presents a first evaluation of multimodal language-based warnings for handovers of control in autonomous cars. A set of possible handover situations varying in urgency is described. A set of multimodal, language-based warnings for these situations is then introduced. All combinations of audio, tactile and visual warnings for handovers were evaluated in terms of perceived urgency, annoyance and alerting effectiveness. Results showed clear recognition of the warning urgency in this new context, as well as low perceived annoyance overall, and higher perceived effectiveness for critical warnings. The time of transition from self-driving to manual mode in the presence of the warnings was then evaluated. Results showed quicker transitions for highly urgent warnings and poor driving performance for unimodal visual warnings. These results provide a novel set of guidelines for an effective transition of control between car and driver in an autonomous vehicle.

Author Keywords

Multimodal feedback; autonomous cars; handover; warnings; audio; visual; tactile; speech; Tactons; urgency.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces. - Auditory (non-speech) feedback; Haptic I/O; Voice I/O.

INTRODUCTION

Autonomous driving, the ability of a car to partially or fully drive itself, is gaining attention both in industry and with the public. The move towards autonomous vehicles is predicted to have a large effect on the car market [9]. This is, however, not without concerns over safety, highlighting the importance of carefully designing interfaces for autonomous cars. Authorities like NHTSA [12] and SAE [18] categorise the levels of autonomy in a vehicle. They stress that autonomy is not a binary state (no autonomy to full autonomy) but rather a staged process that will happen over time. The next generation of vehicles will be at Level 3 Limited Self-Driving Automation [12]. There, "the driver is not constantly expected to monitor the roadway, but to still be available for occasional control with sufficiently comfortable transition times". This point of transition of control is referred to as a handover. It is essential to warn the driver effectively in such a handover situation and design warnings that will enable safe transitions. There have been some studies on handover warnings using a limited number of modalities [13], looking into message content [8] and investigating transition times [5]. However, there has been no study on designing effective multimodal warnings for diverse handover situations when the driver of an autonomous car is busy with a demanding side task. The lack of work on this important topic motivated our study.

Figure 1: The setup of the experiment on autonomous handovers. In a critical event which the car could not address, a handover to the driver was enforced (a). In non-critical events, a handover to the driver was requested (b),(c). When in autonomous mode (d), drivers were playing a tablet game (e). See also http://youtu.be/ni048BpTDG8 for a video of the experimental setup.

In this paper, we first provide a set of situations varying in urgency that cover a range of possible transitions of control from the driver to the car and *vice versa*. We then present a study investigating a set of warnings designed for these situations, using all combinations of audio, tactile and visual modalities varying in urgency. We evaluate the warnings in a driving simulator with participants playing a game on a tablet, to simulate future driving situations. We finally present a first set of guidelines to inform drivers about autonomous car handovers in an effective manner.

Multimodal Displays and Urgency

The use of multimodal displays to alert drivers about different road situations has shown benefits in the past [7]. Since the cues should convey the appropriate level of urgency to the drivers, Baldwin & Moore [1] showed that the use of the word "Danger" increased ratings of perceived urgency for speech. "Warning" and "Caution" created intermediate results and "Notice" led to the lowest urgency ratings. Baldwin [2] also observed lower reaction times to urgent words, presented with higher signal intensity. Edworthy *et al.* [3] found that urgently spoken signal words were perceived as more urgent and appropriate.

Using all combinations of audio, visual and tactile modalities, Politis, Brewster & Pollick [15] evaluated a set of multimodal abstract warnings across three levels of urgency. They found that an increased number of modalities increased ratings of urgency and annoyance. They then evaluated the warnings in the presence of a critical driving event and found lower reaction times [14]. The use of Speech Tactons, tactile cues modelled on the structure of speech, was found to improve the recognition of warnings' urgency and perceived effectiveness [16]. Finally, comparing a set of language-based messages to abstract ones, they found similar performance as well as a slight improvement of driving metrics when using language-based cues [17].

In this paper, we build on these warnings and put them in the context of autonomous cars. No study has yet tested them in this context, yet there is need for effective warnings in situations where the car and driver will share control. Since we wish to describe why a handover occurs, we use speech to accurately inform the drivers about the events that lead to this handover.

Control Handovers in autonomous cars

Driving under high automation is clearly different to manual driving. De Winter *et al.* [21] state that more non-driving tasks are performed in highly-automated driving situations, such as using in-vehicle entertainment systems. A differential effect on situational awareness can also be observed; if drivers are instructed to detect objects in the environment, situational awareness increases with automated driving. This effect is reversed, however, if they are engaged in non-driving tasks. This highlights the design challenge to keep the driver attentive when automation is not complete and handovers may occur. Interestingly, this is still a relevant topic also in the field of aviation [4].

Merat *et al.* [11] found that drivers were quicker to resume control from highly automated driving when automation was switched off regularly compared to when automation switch-off was triggered when drivers disengaged their attention from the road. In their study, the current driving mode was indicated through a text display lighting up when in automated driving. However, there was no emphasis on how to inform the driver about the transition from manual to automated modes. We address this gap in our study. Gold *et al.* [5] investigated the behaviour of drivers when requested to return to driving due to automation failures. Drivers engaged in a tablet side task were warned through a pure tone and a visual icon that they needed to take over control due to an unexpected event and the required times for such transitions were investigated. There was no comparison of warning modalities in this study. It was found that when warnings were presented closer in time to an imminent handover (5 *sec* before) reactions were quicker but driving behaviour was less accurate compared to when they were presented earlier (7 *sec* before). In our study, we used a similar distraction task on a tablet and compared a wide set of warning modalities. We also addressed scenarios where drivers resumed control and *vice versa*.

Regarding warning design for transitions between manual and autonomous modes, literature is limited. Koo *et al.* [8] used speech warnings when the car took over from the driver due to an unexpected event. They found that messages describing why an event happened (e.g. "Obstacle ahead") were preferred and led to better driving performance than messages describing how it was addressed (e.g. "The car is braking"). Urgency in the messages was not varied and all the driving situations tested were critical. Naujoks, Mai & Neukum [13] studied the opposite case, where a driver needed to take over from a vehicle due to an automation failure. They presented audio (a pure tone) together with visual warnings (flashing icon on the dashboard) during takeover requests. They found that bimodal presentation of warnings led to shorter handover times and better driving behaviour compared to only visual presentation.

As evident from the above studies, there is distinct lack of research on how to design effective warnings for control handovers, particularly when the driver is involved in a secondary task. In our study, we address this by designing a set of language-based messages across all combinations of audio, tactile and visual modalities, taking into account the urgency of the takeover situation. We design multimodal warnings that address situations where control is handed from car to driver, as well as from driver to car. We first evaluate these warnings to assess their perceived urgency, annoyance and alerting effectiveness. We then study in a simulator the time taken and driving behaviour when returning to driving from a tablet-based game, for all modalities and urgency levels, to gain a deeper understanding of effective warning design for autonomous cars.

WARNING DESIGN

We designed a set of six speech messages covering a range of possible handovers of control between the car and the driver. We used three different Levels of Designed Urgency (LDU) for the envisioned situations (Level High - L_H, Level Medium - L_M and Level Low - L_L), as in previous studies [16,17]. We also used situations where either the car would hand over control to the driver (CD) or the driver to the car (DC). This resulted to six speech messages, presented in Table 1 (L_HCD, L_MCD, L_LCD, L_HDC, L_MDC and L_LDC). The messages used were adjusted from [10], containing a set of in-vehicle messages prioritized according to SAE J2395 [19]. Adjustments were in order to avoid resemblance between messages in terms of rhythm. High priority messages in [10] were mapped to L_H, intermediate priority ones to L_M and low priority ones to L_L. The word *"Danger!"* was added before each L_H message, *"Warning!"* before L_M and *"Notice!"* before L_L, since this has shown to provide distinctively different urgency ratings in previous studies [1,6,16]. Finally, in L_H the handover was enforced, since imminent actions would be needed in such critical situations, while in L_M and L_L, the handover was only requested. This was reflected in the text of the messages, concluding with whether the messages were an enforced handover or handover request. In line with [8], the messages explained why a handover was necessary, rather than how the handover would happen.

Handover of Control

		CD	DC
Urgency	L_H	Danger! Collision Imminent You have control! **D:** 2.7 *sec* **P:** -0.0 *dBFS* **AF:** 371 *Hz*	Danger! Object in roadway I have control! **D:** 3.0 *sec* **P:** -0.2 *dBFS* **AF:** 346 *Hz*
	L_M	Warning! GPS signal weak Want to take over? **D:** 3.8 *sec* **P:** -11.0 *dBFS* **AF:** 309 *Hz*	Warning! Dense fog ahead May I take over? **D:** 3.4 *sec* **P:** -9.4 *dBFS* **AF:** 296 *Hz*
	L_L	Notice! Toll ahead, 5 pounds Want to take over? **D:** 4.7 *sec* **P:** -18.5 *dBFS* **AF:** 212 *Hz*	Notice! New email from John May I take over? **D:** 4.2 *sec* **P:** -20.4 *dBFS* **AF:** 211 *Hz*

Table 1: The messages designed, using situations of High (L_H), Medium (L_M) and Low urgency (L_L). The handovers of control were from Car to Driver (CD) or from Driver to Car (DC). For each message the duration (D), peak (P) and average frequency (AF) of the audio are reported.

The messages were recorded by a female voice actor using a Rode NT2-A[1] condenser microphone. In line with [6,16], the actor was instructed to speak messages of L_H urgently, as if a loved one was in imminent danger. L_M messages were spoken non-urgently, as if in a friendly conversation with nothing interesting about the situation and L_L messages were spoken in a monotone, deadpan manner. L_H messages were slightly modified to remove pauses between sentences so as to decrease duration. As tactile equivalents of the audio warnings, we used Speech Tactons with a C2 tactor[2], which were constructed following the procedure described in [16]. Finally, for the visual warnings, the text of the warnings was displayed for the duration of the utterance and varied in colour, in line with [17] (Red for L_H, Orange for L_M and Yellow for L_L[3]). For all modifications, Praat[4] and Audacity[5] software were used. Although there are possible limitations of using speech messages as alerts (e.g. see [17]), we saw benefit in creating a good initial set of warnings, which were not available in this use case.

We presented the designed 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 42 different cues were created[6], 7 cues with all modalities (A, T, V, AT, AV, TV, ATV) × 3 Levels of Designed Urgency (L_H, L_M, L_L) × 2 Situations (CD, DC). These warnings were evaluated in two experiments, looking into subjective and objective responses of participants when exposed to the cues.

EXPERIMENT 1: RATING THE DESIGNED WARNINGS

The first experiment investigated subjective responses of participants when exposed to the warnings. This has never been studied before and is essential in order to provide insights on how such warnings would be perceived. A 7×3×2 within subjects design was used with Modality, LDH and Situation as the independent variables and Perceived Urgency (PU), Perceived Annoyance (PA) and Perceived Alerting Effectiveness (PAE) as the dependent ones.

Procedure

Twenty one participants (3 female) aged between 18 and 29 years ($M = 21.00$, $SD = 2.84$) took part in this experiment. They all held a valid driving license and had between 1 and 8 years of driving experience ($M = 3.36$, $SD = 2.01$). All were right handed University students and reported normal

---

[1] http://www.rode.com/microphones/nt2-a

[2] http://www.atactech.com/PR_tactors.html

[3] Red was *RGB*(255,0,0), Orange was *RGB*(255,127,0) and Yellow was *RGB*(255,255,0).

[4] http://www.fon.hum.uva.nl/praat/

[5] http://audacity.sourceforge.net/

[6] All warnings can be found in http://goo.gl/IhzXoQ

vision and hearing. The experiment took place in a University room where participants sat in front of 27-inch Dell 2709W monitor and a PC running the experimental driving simulator. Auditory cues were displayed through a set of Sennheiser HD 25-1 headphones. Tactile cues through a wristband on participants' left hand with a C2 tactor attached to it, in line with [16]. Visual cues were coloured text appearing for the duration of the utterance of the audio in the top centre of the screen, simulating a Head up Display (HuD), in line with [17]. They were sized 228×700 pixels (about 7×21 cm). Participants provided all responses using a mouse. To cover any tactor noise, car sound was played during the experiment.

After being welcomed and explained the experimental procedure, the 42 cues were displayed in a random order to participants to familiarize them with the signals. For each cue, they could either repeat it or proceed to the next when they felt familiar with it. Afterwards, they were again presented with the cues when sitting in front of a driving simulator software depicting a rural road scene with a straight road and a car in front. This software has been used in many studies, e.g. [14]. The participants' car was self-driving. They were asked to imagine they were sitting in the driver's seat of an autonomous vehicle, wearing a wrist mounted device like a smart watch for vibration. Participants rated all cues in terms of PU, PA and PAE, by completing a 5-point Likert scale, in line with [2,16]. In all ratings, the scale was labelled: Not at all (1), Slightly (2), Moderately (3), Very (4) and Extremely (5). Each cue was presented twice, resulting to 82 trials.

Results

Perceived Urgency
Data for PU were analysed using a three-way repeated measures ANOVA, with Modality, LDU and Situation as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(3.20,131.25) = 29.88$, $p < 0.001$). Contrasts revealed that modalities were rated for PU in the following order[7]: T and V lower than A ($F(1,41) = 20.11$, $r = 0.57$, $p < 0.001$), A lower than TV and AV ($F(1,41) = 7.80$, $r = 0.40$, $p < 0.05$), TV and AV lower than AT ($F(1,41) = 5.51$, $r = 0.34$, $p < 0.05$) and AT lower than ATV ($F(1,41) = 7.62$, $r = 0.40$, $p < 0.05$). See Figure 2.a for mean values of PU across modalities. There was a significant main effect of LDU ($F(1.58,64.56) = 306.02$, $p < 0.001$). Contrasts revealed that levels were rated in the following order: L_L lower than L_M ($F(1,41) = 151.02$, $r = 0.89$, $p < 0.001$) and L_M lower than L_H ($F(1,41) = 282.06$, $r = 0.93$, $p < 0.001$) (see Figure 2.b). There was a significant main effect of Situation ($F(1,41) = 24.52$, $p < 0.001$). Contrasts revealed that DC was rated lower than CD (Mean values of DC: 2.79, of CD: 2.99, $F(1,41) = 24.52$, $r = 0.61$, $p < 0.001$).

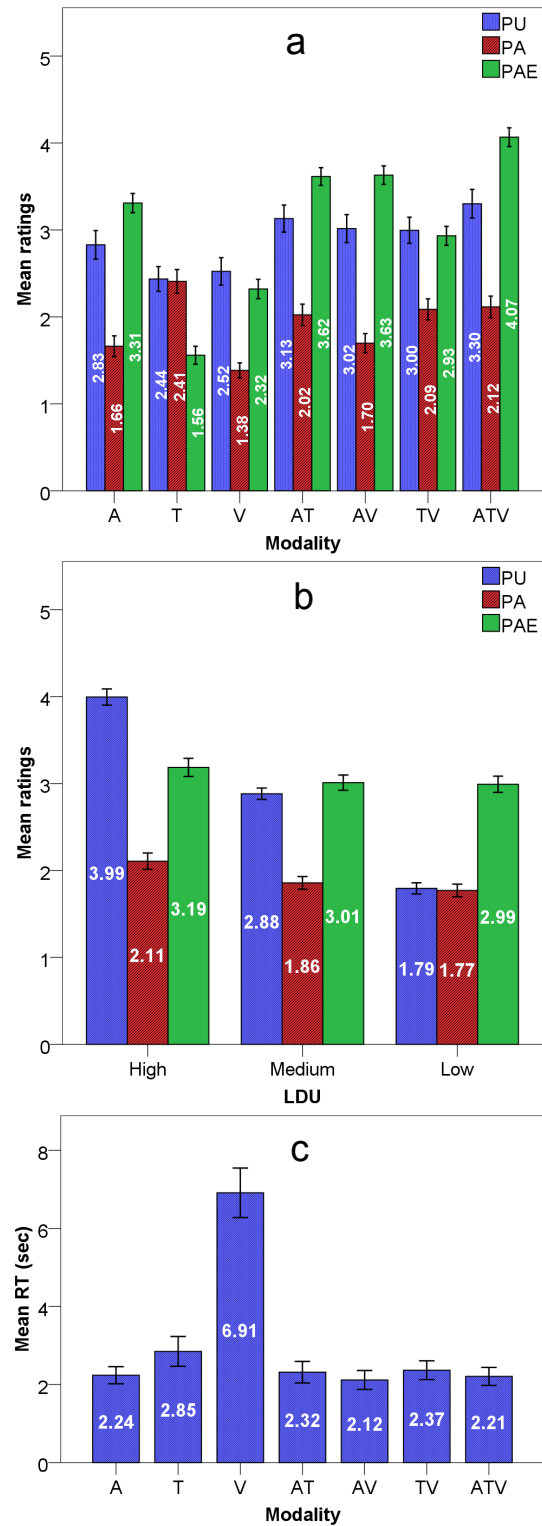
[7] All differences described are significant.

Figure 2: Mean ratings of Perceived Urgency (PU), Perceived Annoyance (PA) and Perceived Alerting Effectiveness (PAE) across modalities (a) and LDU (b) for Experiment 1. Response Times (RT) across modalities for Experiment 2 (c). Error bars indicate 95% confidence intervals.

There was a significant interaction between Modality and Level ($p < 0.001$), indicating that ratings were not differing in L_M between modalities T, V and A and that ratings for L_H were higher in A compared to TV and for L_M higher in TV compared to AV. Finally, that ATV was not rated higher than AT in L_L. A significant interaction between Level and Situation ($p < 0.001$), indicated that the described differences in ratings of Situations were weaker in L_M. Finally, a significant interaction between all factors ($p < 0.05$), indicated that in CD and L_H AV was rated higher than AT.

Perceived Annoyance
Data for PA were analysed using a three-way repeated measures ANOVA, with Modality, LDU and Situation as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(2.82,115.63) = 20.52$, $p < 0.001$). Contrasts revealed that modalities were rated for PA in the following order: V lower than A and AV ($F(1,41) = 9.70$, $r = 0.44$, $p < 0.01$), A and AV lower than AT, TV and ATV ($F(1,41) = 14.21$, $r = 0.51$, $p < 0.01$) and AT, TV and ATV lower than T ($F(1,41) = 4.63$, $r = 0.32$, $p < 0.05$) (see Figure 2.a). There was a significant main effect of LDU ($F(1.17,48.11) = 7.10$, $p < 0.05$). Contrasts revealed that L_L and L_M were rated lower than L_H ($F(1,41) = 8.09$, $r = 0.41$, $p < 0.001$, see Figure 2.b). There was a significant interaction between Modality and Level ($p < 0.05$), indicating that ratings for L_M were as low as L_L for ATV but as high as L_H for T.

Perceived Alerting Effectiveness
Data for PAE were analysed using a three-way repeated measures ANOVA, with Modality, LDU and Situation as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(3.38,138.75) = 146.70$, $p < 0.001$). Contrasts revealed that modalities were rated for PAE in the following order: T lower than V ($F(1,41) = 43.34$, $r = 0.72$, $p < 0.001$), V lower than TV ($F(1,41) = 76.83$, $r = 0.81$, $p < 0.001$), TV lower than A ($F(1,41) = 12.23$, $r = 0.48$, $p < 0.01$), A lower than AT and AV ($F(1,41) = 17.23$, $r = 0.54$, $p < 0.001$), AT and AV lower than ATV ($F(1,41) = 32.64$, $r = 0.67$, $p < 0.001$) (see Figure 2.a). There was a significant main effect of LDU ($F(1.36,55.63) = 4.78$, $p < 0.05$). Contrasts revealed that L_L and L_M were rated lower than L_H ($F(1,41) = 8.34$, $r = 0.41$, $p < 0.01$) (see Figure 2.b). There was a significant main effect of Situation ($F(1,41) = 5.45$, $p < 0.05$). Contrasts revealed that CD was rated lower than DC (Mean values of CD: 3.02, DC: 3.10, $F(1,41) = 5.45$, $r = 0.34$, $p < 0.05$). There was a significant interaction between Modality and Level ($p < 0.001$), indicating that for modality V ratings were lower in L_H compared to L_M and L_L, while they were similar across levels for modality T and TV. They also revealed that for modality A ratings in L_H were higher than L_M and L_L, which did not differ to each other. For AV, this higher rating of L_H was even more pronounced. However in modality AT, other than the difference between L_H and the

other two levels, L_M also rated higher than L_L. Finally, there was a significant interaction between Modality and Situation ($p < 0.05$), indicating that the described differences in ratings for situations CD and DC were more pronounced in modalities V and A.

The results of the first experiment showed clearly that participants identified the designed urgency of all the signals and rated highly urgent warnings as more effective and more annoying, but with low values of annoyance overall, similar to [16]. Multimodal warnings were rated as more urgent and more effective, while unimodal tactile warnings as the most annoying and least effective. To investigate how the cues supported handover situations, the messages relating to cases where the driver needed to take control (situation CD) were assessed in Experiment 2.

EXPERIMENT 2: TAKING CONTROL FROM THE CAR
The second experiment investigated how quickly and effectively distracted participants would be able to resume control in an autonomous vehicle. We focused on situation Car to Driver (CD). This was either requested (L_M, L_L) or enforced (L_H). We designed a task where the driver would be distracted by a tablet game while not driving but would need to return to driving periodically. We investigated how quickly and accurately this transition would happen and how it would affect driving after it happened. A 7×3 within subjects design was used, with Modality and LDU as the independent variables and Response Time (RT), Response Accuracy (RA) and Lateral Deviation after Handover (LDaH) as the dependent ones.

Procedure
Experiment 2 followed immediately after Experiment 1. Participants and equipment were identical, with the addition of a Logitech G27 gaming wheel and pedals, and a ViewSonic View Pad 10pro 10-inch tablet computer. For this experiment, a side-task on the tablet, in line with [5], was introduced to participants. This would decrease their engagement with the driving task and provide a more challenging transition back to driving [21]. It is also likely that drivers may play games on tablets when not driving. The side task was the Concentration memory game previously used in the context of multimodal home reminders [20] (see Figure 1.e). It was chosen since it has a well-defined set of performance metrics and requires high levels of concentration. The driving scene used was similar to Experiment 1, depicting a curvy rural road with a car in front (Figure 1.d). After getting familiarized with the game and the driving simulator, participants were asked to focus on playing the game, unless interrupted by a warning. They were instructed to rest their left hand on the table and their feet on the floor. They were able to use their right hand to play the game by using the tablet, which was placed on a table top stand on the right of the screen, to better accommodate their dominant hand. If they matched all cards in the 3×8 grid, the game would reload automatically with a different random set of cards. While playing the game, the participants'

car was self-driving in autonomous mode in the centre of the lane at a speed of around 70 mph. This mode simulated Level 3 Automation by NHTSA [12] and did not require any driver intervention, but did require availability for occasional control.

At random intervals of any integral value between (and including) 27–32 sec (in line with [14,17] and [11], where automation was regularly disengaged) a warning was presented. If it was an L_L or L_M warning (messages L_LCD and L_MCD), control from the driver was requested. This simulated events, where taking control of the vehicle was not critical. In this case, participants were instructed to press a labelled button on the steering wheel to come back to driving, in line with [13] (see Figure 1.b, 1.c for the visual warnings for these situations). If the warning presented was a L_H one (message L_HCD), control from the driver was required. This simulated automation failures the vehicle could not correct and therefore a switch to manual mode would be needed. To create a more critical situation, the car in front in the L_H case started braking along with the warning presentation, as in [14,17] (see Figure 1.a for the visual warning for this situation). Participants were then handed control and were instructed to brake immediately with their right foot and then return to normal driving. Once the participants stepped on the brake, the car in front would return to its original position away from the participants' car.

Each warning was presented twice in the above setting, resulting in a total of 42 presentations (7 Modalities × 3 Levels of Designed Urgency × 2 presentations). When back to driving, participants were able to steer using the wheel for 10 sec (they did not need to use the accelerator pedal). During this period they were asked to maintain a central lane position. After 10 sec, the car automatically took control and the next trial started. The mode was always indicated on the screen with a car icon for autonomous mode or a person icon for manual (see Figures 1.a-1.d, top right). Even though in a real driving scenario, drivers might not wish to regain control in non-critical situations, participants were still asked to react as quickly as possible in all cases, to measure response time.

Participants' RT was calculated from the onset of a stimulus until the participants pressed the button on the steering wheel (for L_L, L_M) or pressed the brake pedal (for L_H). If participants did not respond to a cue, their RA was 0. Otherwise, their RA was 1 if they performed the right action in the first instance (pressing the pedal or the button) otherwise it was 0 (if they performed the wrong action initially, even if they later corrected it). Their LDaH was the RMSE of their lane position values, logged for 10 sec after they pressed the button on the steering wheel (for L_L, L_H) or 10 sec after the onset of a L_H stimulus and start of the braking event of the lead car (for L_H). The value of 10 sec was chosen since it has shown to be an adequate time to come back to driving in handover situations [11]. Experiments 1 and 2

together lasted about 60 minutes and participants were then debriefed and paid £6.

Results

Response Time
In all there were 882 trials. If participants did not respond to a cue (which was the case in 83 trials – 9.4%), their RT was adjusted to the maximum available time to respond, 10 sec, to allow for a two factor ANOVA analysis. Data for RT were analysed using a two-way repeated measures ANOVA, with Modality and LDU as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(2.35,96.19) = 99.22$, $p < 0.001$). Contrasts revealed that AV, ATV, A, AT and TV caused quicker responses than T ($F(1,41) = 4.98$, $r = 0.33$, $p < 0.05$) and T created quicker responses than V ($F(1,41) = 127.67$, $r = 0.87$, $p < 0.001$). See Figure 2.c for mean values of RT across modalities. There was a significant main effect of LDU (Mean values of L_H: 2.15 sec, L_M: 3.41 sec, L_L: 3.45 sec, $F(1.55,63.47) = 37.27$, $p < 0.001$). Contrasts revealed that L_H cues caused quicker responses than L_M and L_L ones ($F(1,41) = 47.19$, $r = 0.73$, $p < 0.001$). There was a significant interaction between Modality and LDU ($p < 0.05$), indicating that contrary to L_H, TV created quicker responses than AV in L_M.

Response Accuracy
The values of RA for Modalities were as follows: V: 40%, T: 88%, TV: 92%, ATV: 93%, A: 94%, AT: 95% and AV: 98%. Data for RA were treated as dichotomous and analysed with Cochran's Q tests. It was found that modality V was the least accurate compared to all other modalities (all comparisons were significant with $p < 0.001$ and $52.94 \le Q(1) \le 71.05$). It was also found that T was less accurate than AT ($Q(1) = 5.40$, $p < 0.05$) and AV ($Q(1) = 9.94$, $p < 0.01$). Finally, AV was more accurate than TV ($Q(1) = 5.33$, $p < 0.05$) and ATV($Q(1) = 5.44$, $p < 0.05$). The resulting values of RA for LDU were as follows: L_H: 83%, L_M: 86%, L_L: 89%. Cochran's Q tests revealed that L_H was less accurate than L_L ($Q(1) = 4.63$, $p < 0.05$).

Lateral Deviation
Data for LDaH were analysed using a two-way repeated measures ANOVA, with Modality and LDU as factors. Due to sphericity violations, degrees of freedom were corrected using Greenhouse–Geisser estimates. There was a significant main effect of Modality ($F(1.28,52.54) = 12.03$, $p < 0.001$). Contrasts revealed that V warnings resulted in higher LDaH compared to all other modalities (AV, A, AT, TV, T and ATV, $F(1,41) = 11.62$, $r = 0.47$, $p < 0.01$). There was a significant main effect of LDU ($F(1.02,41.78) = 36.06$, $p < 0.001$). Contrasts revealed that L_H warnings led to higher LDaH compared to L_M and L_L ($F(1,41) = 35.10$, $r = 0.68$, $p < 0.001$). There was a significant interaction between LDU and Modality ($F(1.25,51.20) = 18.71$, $p < 0.001$). Contrasts revealed that for L_H, LDaH values increased in modality V, while for L_M and L_L they decreased.

Finally, the results of the tablet game performance were as follows: 134.70 *sec* mean time to complete one game, 0.54 *Clicks per Second* and 0.33 *Superfluous Views per Click*[8]. These are similar to [20], showing that participants were attentive to the game.

DISCUSSION

The results of Experiment 1 showed clearly that participants were able to identify the designed urgency of the cues. This confirmed results of previous studies like [2,6,16] and showed the cues would be suitable for autonomous handovers. Ratings were higher as the number of modalities increased, in line with [15,16]. Finally, situation CD, in which the driver needed to act during a handover, was rated higher in terms of urgency compared to DC. This could reflect an increased alertness on behalf of the driver when his/her intervention is needed, as opposed to when the handover is suggested or performed by the car. It is an interesting finding, since it highlights how drivers' perception of urgency was affected by their involvement in the situation.

In terms of perceived annoyance, the observed values were low overall (between "Slightly" and "Moderately"). Cues in L_H were rated more annoying, confirming previous results [15,16]. The highest rated modality for annoyance was T, indicating that Speech Tactons are more acceptable when used in conjunction with other modalities, as in [16]. This was especially true in L_H and L_M, where the message would be more important to act upon in a real situation compared to L_L. It is interesting that annoyance ratings of T surpassed even trimodal ratings in this study, while many participants reported that Speech Tactons were less understood when used alone. This raises caution on the use of Speech Tactons unimodally for autonomous handovers.

Ratings of alerting effectiveness confirmed the importance of informing drivers about critical events, as in [16]. L_H was rated higher than L_M and L_L, showing that alerts about most critical situations are considered more effective. There was a particular preference to cues containing audio, which was also confirmed by many participants' comments. Unimodal T and V cues, but also TV ones were rated as less effective compared to A. Visual cues were rated low in L_H, indicating that participants appreciated more salient ways to be informed during critical situations. Finally, situation DC was rated as more effective compared to CD, especially in modalities V and A. This indicates that participants showed a preference for simpler cues for situation DC, which indicates that since this situation does not require driver action it can be communicated with less modalities.

Regarding handover times in Experiment 2, it was found that in L_H participants returned to driving significantly quicker than in L_M and L_L. This is a novel finding, confirming that reaction to a critical situation can be reflected in participant responses in this context. Although there were different motor requirements to perform each task (using the foot to brake vs the hand to press a button), this indicates that participants had increased alertness in L_H situation. This was also supported by some participants' comments after the experiment, mentioning that they felt inclined to respond more quickly in an urgent event.

A significant increase in handover time was observed for the V condition which had a much higher RT (also observed in [13]). This reflects the visual attention required by the game, which led to a high number of missed responses, and low response accuracy values. Drivers took an average of 6.9 *sec* to notice visual warnings about handover, which would be catastrophic in real driving. Thus, great care must be taken using unimodal visual displays when the driver's visual attention is focused towards another device. Future studies will investigate whether this problem can be ameliorated by presenting the warnings in the interaction area, for example placing some warnings on the tablet or moving the game interaction to a HuD. However, with the current arrangement, we do not recommend the use of unimodal visuals for autonomous vehicles handovers. The second worse performing modality both in terms of RT as well as RA was T, again highlighting similar limitations to [16] when interpreting vibration in unimodal presentation of Speech Tactons. One limitation of our study was that the game was purely visual. It is common for games to use audio and tactile feedback too. A further study will use a game with more feedback types to study if these interfere with the different modalities of feedback from the car.

In terms of LDaH, it had highest values for the visual modality, in line with [13]. This is due to the high number of failures to respond to handovers signified visually in L_H (15 out of 42 trials). These failures are critical, since they would leave the car uncontrolled. Of course, this particular transition would be much more difficult to test in a real driving setting. Again, the high distraction caused by the side task is reflected in this finding. Presenting handover warnings visually on a HuD would not help overcome this distraction and thus such visual feedback is not recommended.

The fact that LDaH was decreased for L_M and L_L is actually a negative result, indicating a high number of failures to respond to a unimodal visual warning. If participants did not respond, the car would remain under autonomous mode and values of LDaH were zero. These decreased LDaH values would not necessarily lead to problems, since, in the real world, driver intervention is not essential. However, they still indicate that unimodal visuals are not salient enough to attract attention during a handover. Finally, future studies will also compare abstract warnings with language-based ones to investigate any added benefits.

---

[8] Superfluous views show 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 match it to another picture was a superfluous view.

CONCLUSIONS

This paper provided subjective and objective results for multimodal warnings signifying the handover of control between car and driver in autonomous driving. It was found that drivers clearly identified the urgency of the cues and rated multimodal cues as more urgent and more effective. Unimodal audio and visual cues were rated as less annoying but less effective than multimodal ones. When evaluating the handover time and accuracy for distracted drivers to take control of a vehicle, it was found that they were quicker with multimodal cues of high designed urgency. Unimodal visual cues were especially poor since they did not attract drivers' attention back to the road. We therefore suggest the use of multimodal informative cues for critical handovers and we advise against the use of unimodal visuals for such a case.

REFERENCES

1. Baldwin, C.L. and Moore, C.Perceived Urgency, Alerting Effectiveness and Annoyance of Verbal Collision Avoidance System Messages. *HFES Annual Meeting 46*, 22 (2002), 1848–1852.

2. Baldwin, C.L.Verbal collision avoidance messages during simulated driving: perceived urgency, alerting effectiveness and annoyance. *Ergonomics 54*, 4 (2011), 328–337.

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

4. Geiselman, E.E., Johnson, C.M., and Buck, D.R.Flight Deck Automation: Invaluable Collaborator or Insidious Enabler? *Ergonomics in Design: The Quarterly of Human Factors Applications 21*, 3 (2013), 22–26.

5. Gold, C., Damböck, D., Lorenz, L., and Bengler, K."Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications (2013), 1938–1942.

6. Hellier, E., Edworthy, J., Weedon, B., Walters, K., and Adams, A.The Perceived Urgency of Speech Warnings: Semantics versus Acoustics. *Human Factors 44*, 1 (2002), 1–17.

7. Ho, C., Reed, N., and Spence, C.Multisensory In-Car Warning Signals for Collision Avoidance. *Human Factors 49*, 6 (2007), 1107–1114.

8. Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., and Nass, C.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 (IJIDeM)*, (2014), 1–7.

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

10. Lee, J.D., Bricker, D.L., and Hoffman, J.D.A Dynamic Programming Algorithm for Scheduling In-Vehicle Messages. *IEEE Transactions on Intelligent Transportation Systems 9*, 2 (2008), 226–234.

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

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

13. Naujoks, F., Mai, C., and Neukum, A.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 2014*, (2014), 431–438.

14. Politis, I., Brewster, S., and Pollick, F.Evaluating Multimodal Driver Displays under Varying Situational Urgency. *CHI 2014*, ACM Press (2014), 4067 – 4076.

15. Politis, I., Brewster, S., and Pollick, F.Evaluating Multimodal Driver Displays of Varying Urgency. *Automotive UI 2013*, ACM Press (2013), 92 – 99.

16. Politis, I., Brewster, S., and Pollick, F.Speech Tactons Improve Speech Warnings for Drivers. *Automotive UI 2014*, ACM Press (2014), 1–8.

17. Politis, I., Brewster, S., and Pollick, F.To Beep or Not to Beep ? Comparing Abstract versus Language-Based Multimodal Driver Displays. *CHI 2015*, (2015), In Press.

18. SAE J3016.*Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.* 2014.

19. SAE.Its In-Vehicle Message Priority Standard. *J2395*, (2002), 1–7.

20. Warnock, D., Mcgee-lennon, M., and Brewster, S.The Role of Modality in Notification Performance. (2011), 572–588.

21. De Winter, J.C.F., Happee, R., Martens, M.H., and Stanton, N.A.Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour 27*, (2014), 196–217.