Proceedings of the 5th International Conference on Automotive User Interfaces and
Interactive Vehicular Applications (AutomotiveUI ,13), October 28–30, 2013, Eindhoven, The Netherlands.

Evaluating Multimodal Driver Displays of Varying Urgency

Ioannis Politis
Glasgow Interactive Systems Group
School of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
Tel: +44 (0)141 330 3541
I.Politis.1@research.gla.ac.uk

Stephen Brewster
Glasgow Interactive Systems Group
School of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
Tel: +44 (0)141 330 4966
Stephen.Brewster@glasgow.ac.uk

Frank Pollick
School of Psychology
University of Glasgow
58 Hillhead Street
Glasgow G12 8QB, UK
Tel: +44 (0)141 330 3945
Frank.Pollick@glasgow.ac.uk

ABSTRACT
Previous studies have evaluated Audio, Visual and Tactile warnings for drivers, highlighting the importance of conveying the appropriate level of urgency through the signals. However, these modalities have never been combined exhaustively with different urgency levels and tested while using a driving simulator. This paper describes two experiments investigating all multimodal combinations of such warnings along three different levels of designed urgency. The warnings were first evaluated in terms of perceived urgency and perceived annoyance in the context of a driving simulator. The results showed that the perceived urgency matched the designed urgency of the warnings. More urgent warnings were also rated as more annoying but the effect of annoyance was lower compared to urgency. The warnings were then tested for recognition time when presented during a simulated driving task. It was found that warnings of high urgency induced quicker and more accurate responses than warnings of medium and of low urgency. In both studies, the number of modalities used in warnings (one, two or three) affected both subjective and objective responses. More modalities led to higher ratings of urgency and annoyance, with annoyance having a lower effect compared to urgency. More modalities also led to quicker responses. These results provide implications for multimodal warning design and reveal how modalities and modality combinations can influence participant responses during a simulated driving task.

General Terms
Design, Human Factors

Keywords
Multimodal interaction; warnings; audio; visual; tactile; perceived urgency; perceived annoyance; recognition time; simulator

1. INTRODUCTION
Multimodal displays are increasingly used to alert drivers about situations of varying importance. Numerous studies have investigated the effectiveness of different modalities as well as modality combinations in warnings. However, the candidate modalities have rarely been exhaustively combined with their urgency varied, to assess their subjective and objective responses.

Ho & Spence [8] investigated the use of car horn sounds and verbal cues as warning signals. It was found that participants' responses to a critical event were more rapid when the cues were coming from the direction of the event (front or back) and when their attention was directed to the appropriate direction through a verbal cue. Ho, Tan & Spence [10] studied a set of spatially predictive vibrotactile cues (indicating the correct direction of the threat in 80% of the cases), as well as non-predictive ones (indicating the correct direction of the threat in 50% of the cases). They found that both spatially predictive and non-predictive cues presented from the same direction as the approaching threat (front or back) can decrease drivers' reaction times during a simulated driving task, compared to cues presented from the opposite direction. Ho, Reed & Spence [9] found that audio-tactile presentation (vibration on the torso and car horn) of front to rear-end collision warnings can lead to faster reactions in a simulated driving task compared to the unimodal presentation of the warnings.

Utilizing audio, tactile and visual modalities for alerting drivers, Scott & Gray [23] found that tactile warnings on the abdomen, simulating seat belt warnings, can induce shorter reaction times during a critical driving situation compared to audio and visual warnings on the dashboard. Kern *et al.* [15] concluded that vibration on the steering wheel can lead to increased performance and user acceptance in a navigational task, when accompanied with multimodal information presentation including speech or visuals. Huang *et al.* [13] provide another example of using multiple modalities in isolation to alert drivers during a simulated driving task. They report that sound as well as tactile warnings on the steering wheel resulted to lower response times compared to visual warnings on the mirror, when alerting the drivers of a car potentially approaching from behind during overtaking. Murata, Kanbayashi & Hayami [20] presented a study using combinations of audio, tactile and visual modalities to alert drivers of an approaching hazard. The three modalities in isolation, as well as the combinations of audio + tactile, audio + visual and visual + tactile modalities were used in this study. All three modalities in combination were not used. It was found that tactile and audio + tactile warnings introduced shorter reaction times to hazards.

In the aforementioned studies, the signals were designed to convey a single level of urgency, usually related to several types of impending collisions. However, in a real driving situation, alerts may not always refer to situations that are equally urgent. Earlier work by Mollenhauer *et al.* [19] studied the use of visual and auditory displays to signify road signs during a simulated driving task. It was found that auditory presentation resulted in better recall of the signs but lower performance and increased ratings of perceived distraction and annoyance. When presenting only half the signals, sign recall and driving performance improved.

Kaufmann *et al.* [14] provided a set of guidelines regarding the use of the audio, tactile and visual modalities for three priority levels. These levels were defined as high priority, requiring immediate action, medium priority, with no immediate action necessary and low priority, with no immediate relevance to the driving task. It was suggested that audio and tactile modalities are suitable for high priority messages, visual and tactile for medium and au-

AutomotiveUI '13, October 28 - 30 2013, Eindhoven, Netherlands
Copyright 2013 ACM 978-1-4503-2478-6/13/10…$15.00.
http://dx.doi.org/10.1145/2516540.2516543

dio and visual for low priority ones. These suggestions were based on measures of driving performance of participants. Sullivan & Buonarosa [24] tested three suites of sound warnings, namely semantic (natural sounds, semantically associated with the events they signified), less-urgent-semantic (same as semantic, but modified for attributes such as pitch, pulse rate and duration) and abstract. It was found that semantic warnings induced the fastest reaction time and highest recognition rate. However, there was a possible limitation of less-urgent-semantic sounds, namely that editing natural sounds, may limit their association to the events.

Cao *et al.* [4] investigated audio and tactile cues designed to convey four urgency levels. Number of pulses and interpulse interval were manipulated to signify urgency. In addition, pitch was manipulated for the audio cues and intensity for the tactile ones. The main task was a visual tracking one and there were different levels of auditory distractions applied. There was no driving task simulated. Participants were asked to identify the urgency level of the cues and a general trend of higher priority – faster response was found, indicating that participants perceived the designed urgency of the cues. It was also found that vibration cues were identified more accurately, but sound cues more quickly. Sound cues were also reported as easier to distinguish than vibrations.

In terms of warning design several studies investigated how different signal parameters influence perception. Edworthy, Loxley & Dennis [6] demonstrated how parameters like higher fundamental frequency, higher speed and larger pitch range can increase the perceived urgency ratings of auditory warnings. Later, Edworthy *et al.* [5] observed significantly lower response times for warnings designed to be of a high urgency level, compared to those of medium and low urgency levels. Marshall, Lee & Austria [17] reported how parameters like higher pulse duration and lower interpulse interval can increase the ratings of urgency and annoyance of audio alerts. In their study, different sound cues were investigated in three contexts of varying urgency, namely impending collision, navigation and email messages. The results also indicated that perceived urgency was a more decisive factor for the ratings of appropriateness in high urgency situations, while perceived annoyance in low urgency situations.

In line with the above results, Gonzalez *et al.* [7] found that fundamental frequency, pulse rate and intensity of warning sounds, all positively influenced the ratings of urgency and annoyance. However, pulse rate was suggested as the most suitable in a context of conveying events of varying urgency, since it did not elicit as high ratings in annoyance compared to the ratings of urgency. A similar effect for the tactile modality was reported by Pratt *et al.* [21], where pulse rate was found to positively influence the ratings of perceived urgency, while having less impact on the ratings of perceived annoyance. Saket *et al.* [22] investigated the perceived urgency of vibrations on a mobile phone and found that shorter interpulse interval and pulse duration contributed to higher ratings of urgency. They also found that four different levels of urgency were clearly recognised by participants. Investigating the audio, tactile and visual modalities, Baldwin *et al.* [1] and Lewis & Baldwin [16] initiated the construction of a crossmodal urgency scale. Pulse rate (or flash rate for visuals) was found to be an effective measure of varying urgency across the three modalities. Moreover, sound intensity and frequency were found to be effective for audio signals, while word choice and colours for visual ones. It was found that for the auditory modality, an increase in perceived urgency created a higher increase in perceived annoyance, which was not the case for the tactile and visual modalities. Baldwin *et al.* [1] note that there is limited information regarding the impact of presenting warnings using multiple modalities to drivers and evaluating them in terms of urgency and annoyance, a factor we address in our studies.

In summary, although there have been several studies investigating the performance and perception of audio, visual or tactile warnings in the driving context, these have not been evaluated in all their multimodal combinations. Additionally, guidelines for designing messages of different urgency have not been applied multimodally in a driving simulator, so as to test their subjective and objective responses. Previous studies into the recognition of different levels of urgency of warnings have not used simulated driving as the main task. This indicates the relevance of investigating how all combinations of auditory, visual and tactile modalities will influence the ratings of urgency and annoyance in a simulated driving task. Moreover, response times for the recognition of the different signals will extend the available results for audio and investigate how they apply multimodally. Therefore, a set of warnings of varying urgency utilizing all aforementioned modalities was designed and tested in the context of a driving simulator in terms of perceived urgency, annoyance, and recognition times.

2. WARNING DESIGN

In line with [17], three levels of urgency were designed to signify situations of different importance. Level 1 signified situations of high urgency, such as an impending collision, Level 2 of medium urgency, such as low fuel and Level 3 of low urgency, such as an incoming message. Audio, visual and tactile modalities, as well as all of their combinations were used for the warnings (Audio (A), Visual (V), Tactile (T), Audio + Visual (AV), Audio + Tactile (AT), Tactile + Visual (TV), Audio + Tactile + Visual (ATV)). This resulted in 21 different warning signals, 7 signals with the above modalities × 3 levels of designed urgency.

The warnings consisted of pure tones, colours or vibrations, delivered as pulses to the participants, with a varying pulse rate depending on the level of urgency. Using such simple parameters allowed us to make the warnings as similar as possible across the modalities. In line with [1, 16], pulse rate was varied to signify escalating urgency across all modalities. Warnings of the same level had common characteristics of pulse rate regardless of modality. There were 8 pulses with 0.1 *sec* single pulse duration and 0.1 *sec* interpulse interval for Level 1, 5 pulses with 0.17 *sec* single pulse duration and 0.17 *sec* interpulse interval for Level 2 and 2 pulses with 0.5 *sec* single pulse duration and 0.5 *sec* interpulse interval for Level 3. All warnings lasted 1.5 *sec* each.

For auditory warnings, base frequency was also varied, in line with [1, 6, 16, 17] (1000 *Hz* for Level 1, 700 *Hz* for Level 2 and 400 *Hz* for Level 3). For visual warnings colour was also varied, in line with [1] (Red for Level 1, Orange for Level 2 and Yellow for Level 3[1]). A C2 Tactor from Engineering Acoustics[2] was used for the tactile stimuli, as is common in studies investigating tactile feedback, e.g. [11, 12]. The frequency of the tactile stimuli was kept constant at 250 *Hz*, which is the nominal center frequency of the C2 and the frequency at which the skin is most sensitive. Multimodal signals consisted of simultaneous delivery of unimodal ones to achieve a synchronous effect of sound, vibration, visuals or any of their combinations. Stimulus intensity was not varied in any of the modalities, to avoid causing discomfort to the partici-

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

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

pants, as suggested in studies of both Earcons and Tactons [2, 7, 11, 12].

To evaluate the warnings created, two experiments were designed. The goal of the first was to acquire participants' subjective ratings of perceived urgency and annoyance when exposed to the warnings, without being given any information about their designed urgency level. In the second, participants would attempt to identify the level of urgency of the same set of warnings with performance measured in terms of recognition time and accuracy.

3. EXPERIMENT 1

The first experiment investigated the subjective responses in terms of perceived urgency and perceived annoyance. In line with [17], it was hypothesized that the different levels of urgency designed in the warnings would influence the ratings of urgency and annoyance by the participants. Also, ratings of urgency and annoyance would differ depending on the modalities used. Moreover, in order to investigate the robustness of the warnings across different contexts, all responses were acquired both in the presence and in the absence of a driving simulator. The expectation was that if the participants became immersed in the context of a driving simulator, this would influence their ratings.

3.1 Design

A 7×3×2 within subjects design was followed for this experiment, with Modality, Designed Urgency and Context as the independent variables and Perceived Urgency and Perceived Annoyance as the dependent variables. Modality had 7 levels: A, T, V, AT, AV, TV, ATV. Designed Urgency had 3 levels: Level 1 (High Urgency), Level 2 (Medium Urgency) and Level 3 (Low Urgency). Context had 2 levels: Driving Simulator and No Driving Simulator.

3.2 Participants

Twenty participants (6 female) aged between 19 and 32 years old ($M = 22.4$, $SD = 4.3$) took part in this experiment. They all held a valid driving license and had between 1 and 13 years of driving experience ($M = 3.45$, $SD = 3.31$). They were either undergraduate or postgraduate University students. All participants reported either normal or corrected to normal vision and no injuries around the abdominal area, where vibrations were delivered. One participant reported moderately reduced hearing from one ear, which however did not hinder everyday activities. Therefore, data of this participant were retained.

3.3 Equipment

The experiment took place in a dedicated University room, where the participants sat on a padded chair in front of a desk with a 27-inch computer screen (Dell 2709W). The computer was running driving simulator software, depicting a three lane road in a rural area and a front car maintaining a steady speed. The simulator has been previously used in many research studies, for example [3]. As in [3], safety cones were placed on either side of the central lane, to reinforce lane keeping. Sound was delivered through a set of headphones (Sennheiser HD 25-1). Vibration was delivered through a C2 Tactor from Engineering Acoustics, attached on an adjustable waist belt. The belt was placed by the participants in the middle of the abdominal area, to simulate a vibrating seat belt, similar to [23]. Visuals were delivered through coloured circles that flashed in the top center of the screen, sized 400×400 pixels (12×12 *cm*). The circles did not obstruct the front car and were designed to simulate the feedback of a Head-Up Display (HUD). Participants used a mouse to submit their ratings. Figure 1.a depicts the setup of Experiment 1 and 1.c the waist belt and Tactor.

Figure 1. The setup of Experiments 1 (a) and 2 (b), the waist belt with the Tactor (c) and the steering wheel used in Experiment 2, with the response buttons highlighted (d).

3.4 Procedure

Participants were welcomed and provided an introduction to the experiment. In order to cover any noise from the Tactor, car sound was heard through the headphones throughout the experiment. The car sound was an extract from a recording of a vehicle idling, retrieved from the Internet[3].

Before beginning the ratings, all 21 signals were played once to the participants, always in the following order: A → V → T → AV → AT → TV → ATV for Level 1, then the same order for Level 2 and then for Level 3. If needed, sound and vibration were adjusted so as to achieve comfortable levels. No information about the levels of designed urgency was given to the participants. Next, the warning signals were played to the participants in a random order and with a random interval of any integral value between (and including) 8 – 20 *sec*, similarly to [8, 9, 10]. Each stimulus was played 3 times. This resulted in a total of 63 stimuli played to the participants. After each stimulus was played, participants were asked to rate the perceived urgency and annoyance of the stimulus on a scale of 0 to 100, in line with [17] (0 for lowest, 100 for highest). This was done by manipulating the value of a slider in a window that appeared on the screen, after each stimulus had finished playing.

The above procedure was repeated in two Contexts. In the first context, participants rated the stimuli in front of a blank computer screen and in the second context they were looking at the driving simulator with a car that was accelerating and then maintaining a speed of about 60 *mph*. We chose not to let participants control the vehicle in this case because pilot tests showed that it was not practical to manipulate the steering wheel and then switch to the mouse rating for this many repetitions. However, they were asked to imagine they were driving the car on the simulator. The above contexts were balanced across participants.

After rating the stimuli both in front of the simulator and in front of a blank screen, the experiment was concluded and participants

[3] http://soundfxcenter.com/transport/car/020ff2_Compact_Car_Idle_Sound_Effect.mp3

were debriefed about the purpose of the study. The experiment lasted about 45 minutes in total and participants received £6.

3.5 Results

3.5.1 Perceived Urgency

Data for perceived urgency were analysed using a three-way repeated measures ANOVA, with Context, Modality and Level as factors. Mauchly's test showed that the assumption of sphericity had been violated for Modality and Level, therefore Degrees of Freedom were corrected with Greenhouse–Geisser sphericity estimates. The main effect of Level was significant ($F(1.26, 74.06) = 213.41$, $p < 0.001$). Contrasts revealed that warnings of Level 1 were rated as significantly more urgent than warnings of Level 2 ($F(1, 59) = 293.88$, $r = 0.91$, $p < 0.001$), which in turn were rated as significantly more urgent than warnings of Level 3 ($F(1, 59) = 92.15$, $r = 0.69$, $p < 0.001$). The mean ratings of urgency across levels can be found in Figure 2.

The main effect of Modality was also found to be significant ($F(3.43, 202.07) = 73.64$, $p < 0.001$). Contrasts revealed that ATV warnings were rated as significantly more urgent than TV ones ($F(1, 59) = 34.28$, $r = 0.61$, $p < 0.001$), AV as significantly more urgent than AT ($F(1, 59) = 31.17$, $r = 0.59$, $p < 0.001$) and V was significantly more urgent than A($F(1, 59) = 16.19$, $r = 0.46$, $p < 0.001$). The mean ratings of perceived urgency are shown in Figure 3.

Finally, there was no significant main effect of Context ($F(1, 59) = 2.341$, $p = 0.131$). There was a significant interaction effect between Level and Modality ($F(6.56, 387.14) = 16.4$, $p < 0.001$). The interactions with the highest effects were that the visual warnings elicited significantly higher ratings of urgency compared to audio for Level 1 compared to Level 2, and that for the AT warnings in Level 1, ratings were lower compared to V, an effect that was reversed for Level 2 (see also Figure 4).

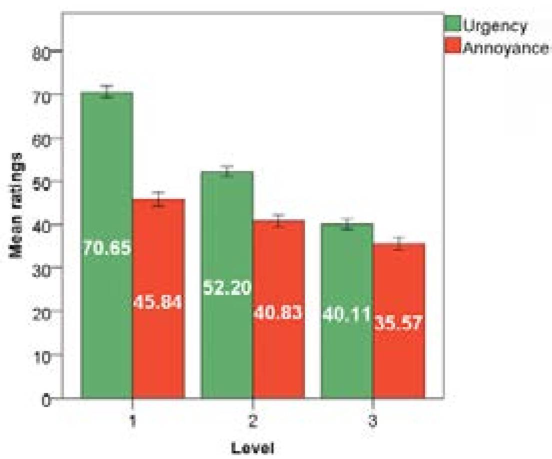

Figure 2. The mean ratings of urgency and annoyance across levels for Experiment 1. For all graphs, error bars show the Standard Error of ratings.

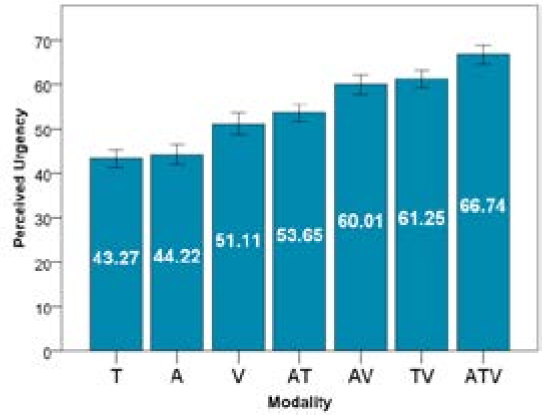

Figure 3. The mean ratings of perceived urgency across modalities for Experiment 1, sorted by their mean values.

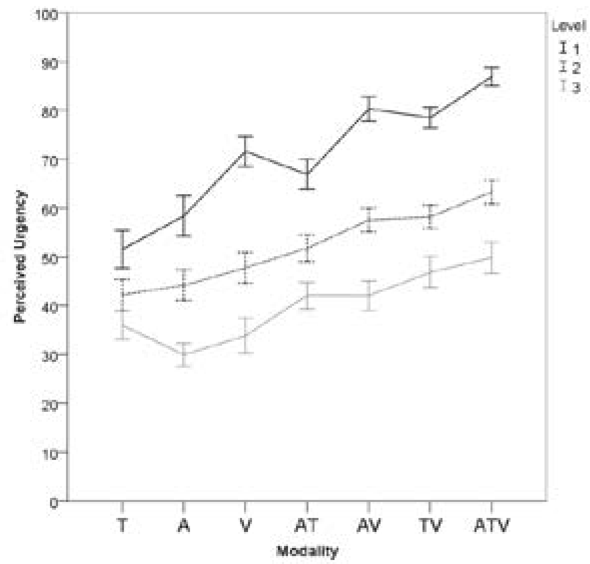

Figure 4. The interaction between Level and Modality for Experiment 1. Modalities are sorted by their mean values of perceived urgency.

3.5.2 Perceived Annoyance

Data for perceived annoyance were also analysed using a three-way repeated measures ANOVA, with Context, Modality and Level as factors. Mauchly's test showed that the assumption of sphericity had been violated for Modality and Level, therefore Degrees of Freedom were corrected with Greenhouse–Geisser sphericity estimates. The main effect of Level was found to be significant ($F(1.65, 97.56) = 37.76$, $p < 0.001$). Contrasts revealed that warnings of Level 1 were rated as significantly more annoying than warnings of Level 2 ($F(1, 59) = 16.42$, $r = 0.47$, $p < 0.001$), which in turn were rated as significantly more annoying than warnings of Level 3 ($F(1, 59) = 34.74$, $r = 0.61$, $p < 0.001$). The mean ratings of annoyance across levels can be found on Figure 2.

The effect of Modality was also significant ($F(2.87, 169.37) = 27.54$, $p < 0.001$). Contrasts revealed that ATV warnings were rated as significantly more annoying than AT ones ($F(1, 59) = 5.49$, $r = 0.29$, $p < 0.05$), TV warnings as significantly more annoying than T ($F(1, 59) = 20.56$, $r = 0.51$, $p < 0.001$), AV warnings as significantly more annoying than A ($F(1, 59) = 6.93$, $r = 0.32$, $p < 0.05$) and A warnings as significantly more annoying than V ($F(1,59) = 8.81$, $r = 0.36$, $p < 0.05$). The mean ratings of annoyance can be found on Figure 5. Finally, there was no significant main effect of Context ($F(1,59) = 0.842$, $p = 0.362$). There was a significant interaction effect between Level and Modality ($F(8.11, 478.29) = 3.742$, $p < 0.001$). Contrasts revealed that the T warnings elicited significantly lower ratings of annoyance compared to the AV ones for Level 1, an effect that was reversed for Level 2 (see also Figure 6).

As evident from Experiment 1, the perceived urgency of the warnings matched their designed urgency, since there was a clear difference of participants' ratings along the three urgency levels. This means that the urgency of the cues we designed was clearly identified, even without any training, suggesting that our design was effective. Although perceived annoyance did increase when warnings became more urgent, this effect was not as strong as urgency. In terms of modalities, the urgency ratings increased as more modalities were used, and the ratings of annoyance were higher for signals using the tactile modality.

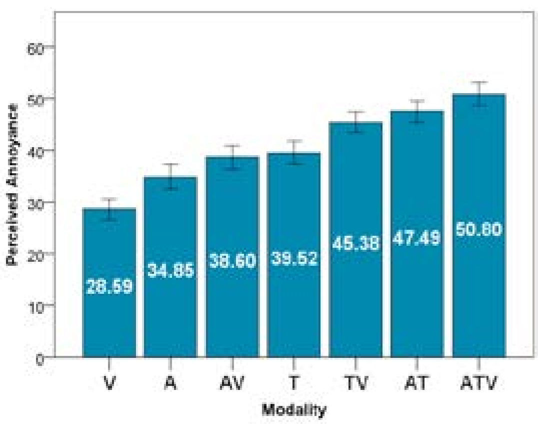

Figure 5. The mean ratings of annoyance across modalities for Experiment 1, sorted by their mean values.

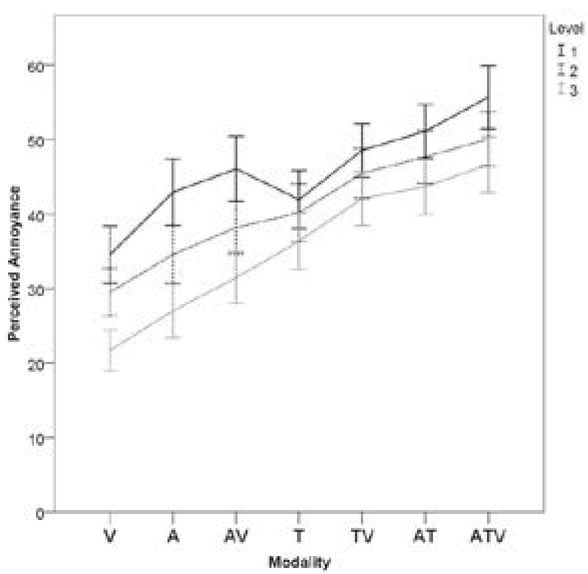

Figure 6. The interaction between Level and Modality for Experiment 1. Modalities are sorted by their mean values of perceived annoyance.

4. EXPERIMENT 2

For the second experiment, the same warnings as the first were evaluated in terms of recognition time and accuracy. In line with [4, 5], it was hypothesized that the designed urgency and modality of warnings would influence their recognition time. As in Experiment 1, all responses were acquired both in the presence and in the absence of a driving simulator. The expectation was that if participants were engaged in a primary driving task, this would influence their responses.

4.1 Design

A 7×3×2 within subjects design was followed for this experiment, with Modality, Designed Urgency and Context as the independent variables and Recognition Time and Recognition Accuracy as the dependent variables. All participants from Experiment 1, except one, participated to Experiment 2 over the period of a week. This resulted in nineteen participants (6 female) aged between 19 and 32 years ($M = 22.52$, $SD = 4.38$). The only difference in equipment between the experiments was that instead of a mouse, participants used a Logitech G27 gaming steering wheel[4] to control the simulator and to provide their responses. The simulator logged participants' inputs at a frequency of 50 *Hz*. Figure 1.b depicts the setup of the experiment and 1.d the steering wheel.

4.2 Procedure

Participants were welcomed and provided an introduction to the experiment. As in Experiment 1, car sound was heard through the headphones throughout the experiment to cover any sound from the Tactor. Before beginning the task, a training session was performed, where all 21 signals were played once to the participants. A label with the text "Level 1: Signals of HIGH urgency e.g. Impending Collision" was presented and then all signals of Level 1, (A → V → T → AV → AT → TV → ATV). This was followed by a label with the text "Level 2: Signals of MEDIUM urgency e.g. Low Fuel" and the signals of Level 2 in the same order. Finally, a label with the text "Level 3: Signals of LOW urgency e.g. Incoming Message" was shown, followed by the signals of Level 3. The whole training lasted about 80 *sec* in total. Any adjustments to sound or vibration were also performed at this part to ensure participants were comfortable.

For the main experiment, the warning signals were played to the participants in a random order and with a random interval of any integral value between (and including) 8 – 20 *sec*, as in Experiment 1. Each stimulus was played 3 times. This resulted in a total of 63 stimuli. Participants were asked to identify the level of urgency of each stimulus by pressing one of three buttons on the steering wheel as quickly as possible. Buttons were labelled with numbers (1, 2 or 3) according to the urgency levels – topmost for Level 1, middle for Level 2, bottom for Level 3 (see Figure 1.d).

The above procedure was repeated in two Contexts, balanced across participants. In the first, participants responded to the stimuli in front of a blank screen and in the second they were steering a car in the simulator, which was accelerating and then maintaining a speed of about 60 *mph*. Participants were instructed to maintain a central position in the lane. The accelerator and brake pedals were not used. Similar to [3], noise was added to the vehicle dynamics so that steering was required to keep the vehicle in the centre of the road and create a realistic driving task. The experiment lasted about 45 minutes and participants received £6.

[4] [illegible]

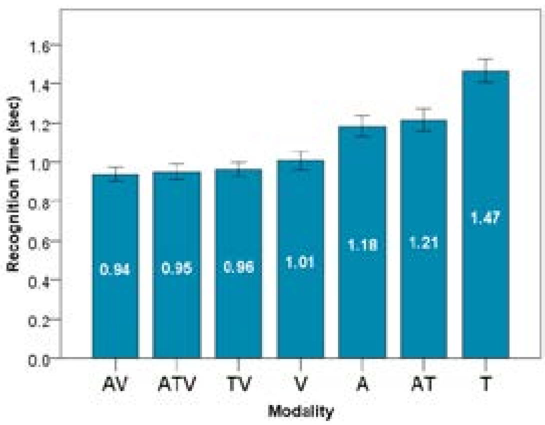

Figure 7. The mean recognition times across modalities for Experiment 2, sorted by their mean values.

4.3 Results

4.3.1 Recognition Time

All data for recognition time were analysed using a three-way repeated measures ANOVA, with Context, Modality and Level as factors. Mauchly's test showed that the assumption of sphericity had been violated for Modality, therefore Degrees of freedom were corrected with Greenhouse–Geisser sphericity estimates. There was a significant main effect of Level ($F(1.87, 89.65) = 147.65$, $p < 0.001$). Contrasts revealed that warnings at Level 1 elicited significantly quicker responses than warnings at Level 2 ($F(1, 48) = 284.63$, $r = 0.93$, $p < 0.001$) and at Level 3 ($F(1, 48) = 210.23$, $r = 0.90$, $p < 0.001$). There was no significant difference in recognition times between Levels 2 and 3 ($F(1, 48) = 0.554$, $p = 0.46$).

There was a significant main effect of Modality ($F(3.56, 170.77) = 71.00$, $p < 0.001$). Contrasts revealed that T warnings elicited significantly slower responses compared to AT warnings ($F(1, 48) = 52.60$, $r = 0.72$, $p < 0.001$) and responses to A warnings were significantly slower compared to V warnings ($F(1, 48) = 18.71$, $r = 0.53$, $p < 0.001$) The mean response time across modalities can be found in Figure 7. There was no significant main effect of Context ($F(1, 48) = 0.241$, $p = 0.63$). There was a significant interaction between Level and Modality ($F(7.81, 374.94) = 3.09$, $p < 0.001$), indicating that the V modality elicited significantly quicker responses compared to T for Level 2 than it did for Level 1. Moreover, while responses for AT warnings were quicker compared to A for Level 1, this effect was reversed for Level 2. Finally, responses were quicker for V compared to A for Level 3 and this effect was enhanced for Level 2.

4.3.2 Recognition Accuracy

In all, there were 2390 participant responses and only 4 cases where the participants failed to respond. Nine responses were excluded from the analysis, since the tactor was audible by the participant. This was due to a misplacement of the tactor and was reported by the participant during the trial, both when it started and when it stopped being audible. For the rest of the responses, 2255 were valid (94.7%) and 126 incorrect (5.3%). Data for recognition accuracy were treated as dichotomous (with values "correct" or "incorrect") and analysed with Cochran's Q tests. These revealed that participants made significantly more mistakes when the simulator was present ($Q(1) = 23.02$, $p < 0.001$). Moreover, significantly more misrecognitions were made for Level 2 compared to Level 1 ($Q(1) = 16.07$, $p < 0.001$) and significantly more misrecognitions were made for Level 3 compared to Level 2 ($Q(1) = 11.78$, $p < 0.001$). Finally, it was found that participants made significantly more mistakes in the T modality compared to A ($Q(1) = 17.04$, $p < 0.001$), AT ($Q(1) = 12.76$, $p < 0.001$) and AV ($Q(1) = 12.00$, $p < 0.001$).

The results of Experiment 2 suggest that there was a clear advantage for warnings of Level 1 in terms of recognition time and accuracy, whereas there was no strong difference when comparing Levels 2 and 3. AV and ATV warnings were the quickest to be recognised, and tactile warnings were the slowest and the least accurate in terms of recognition.

5. NUMBER OF MODALITES

Some participants commented that their responses were influenced by how many modalities were present in the signals, namely one modality (A, T, V), two modalities (AT, AV, TV) or three modalities (ATV). To investigate this further, a separate analysis was performed on the number of modalities for both experiments.

Data for perceived urgency from Experiment 1 were analysed using a one-way repeated measures ANOVA, with Number of Modalities (NoM) as a factor. Mauchly's test showed that the assumption of sphericity had been violated for NoM, therefore Degrees of freedom were corrected with Greenhouse–Geisser sphericity estimates. The main effect of NoM was found to be significant ($F(1.84, 660.33) = 203.46$, $p < 0.001$). Contrasts revealed that warnings with three modalities elicited significantly higher ratings of urgency compared to warnings with two modalities ($F(1, 359) = 86.21$, $r = 0.44$, $p < 0.001$), which in turn were rated as significantly more urgent than warnings with one modality ($F(1, 359) = 144.80$, $r = 0.54$, $p < 0.001$). Figure 8 shows the ratings of urgency and annoyance in Experiment 1, for different NoM.

Data for perceived annoyance from Experiment 1 were analysed using a one-way repeated measures ANOVA, with NoM as a factor. Mauchly's test showed that the assumption of sphericity had been violated for NoM, therefore Degrees of freedom were corrected with Greenhouse–Geisser sphericity estimates. The main effect of NoM was found to be significant ($F(1.89, 676.84) = 83.82$, $p < 0.001$). Contrasts revealed that warnings with three modalities elicited significantly higher ratings of annoyance compared to warnings of two modalities ($F(1, 359) = 50.69$, $r = 0.35$, $p < 0.001$), which in turn were rated as significantly more annoying than warnings of one modality ($F(1, 359) = 40.69$, $r = 0.32$, $p < 0.001$).

Data for recognition time from Experiment 2 were analysed using a one-way repeated measures ANOVA, with NoM as a factor. The main effect of NoM was found to be significant ($F(2,678) = 47.16$, $p < 0.001$). Contrasts revealed that warnings with three modalities elicited significantly quicker responses compared to warnings with two modalities ($F(1,339) = 17.08$, $r = 0.22$, $p < 0.001$), which in turn elicited significantly quicker responses than warnings with one modality ($F(1, 339) = 30.16$, $r = 0.29$, $p < 0.001$). Finally, there was no significant difference in terms of recognition accuracy between warnings with one, two or three modalities. Figure 9 shows the recognition times in Experiment 2, for different NoM.

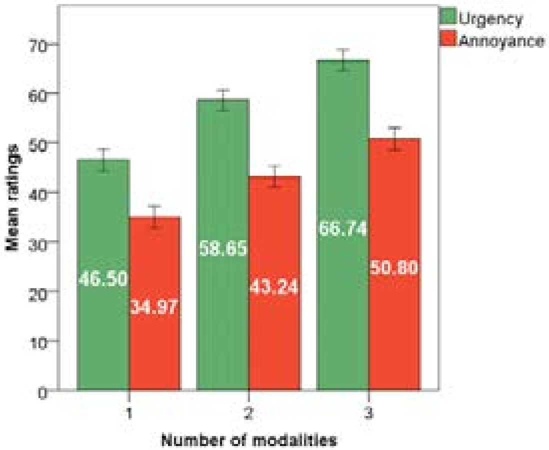

Figure 8. The mean ratings of urgency and annoyance for different NoM in Experiment 1.

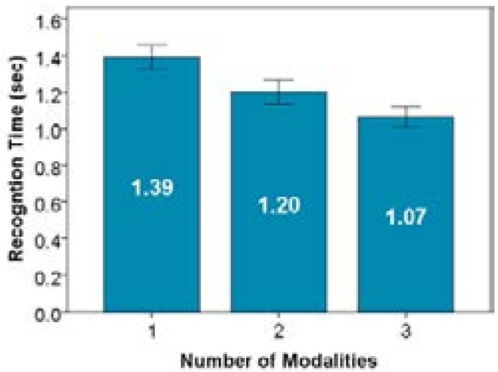

Figure 9. The mean recognition times for different NoM in Experiment 2.

6. DISCUSSION

As evident in both experiments, the warnings succeeded in clearly conveying three different urgency levels to participants. Ratings of perceived urgency were in accordance with the urgency designed in the warnings. This result extends existing work by testing perceived urgency in all combinations of the Audio, Tactile and Visual modalities. This was also done in the context of a driving simulator. As a guideline, car warning designers can utilise interpulse interval to vary urgency across any of the above modalities or their combinations. Frequency for audio and colour for visual signals can also be used as a means to manipulate urgency of warnings. Although warnings were rated as more annoying as their designed urgency increased, the values and effect sizes of observed differences in perceived annoyance across levels were lower compared to their perceived urgency. This result is in line with [7, 17, 21] and indicates that urgency was a more decisive factor than annoyance in the ratings. This is important, since warnings that annoy the driver can be less effective or may even be disabled (see also [18]). As long as warning designers are cautious not to overload the drivers with alarms of low importance, annoyance can be kept at manageable levels as urgency increases.

In terms of recognition times, warnings of high urgency were recognised both more quickly and more accurately compared to warnings of medium and low urgency. This result extends existing work like [4, 5] by testing in the context of a driving simulator and using all combinations of Audio, Visual and Tactile modalities. More urgent warnings created more quick and more accurate responses in this study. These results suggest that high urgency warnings were effectively associated with a high urgency situation (impending collision) and performed better compared to warnings associated with medium or low urgency situations (low fuel, incoming message). This indicates the robustness of the warnings and their appropriateness for use in a critical driving task.

The modalities used affected the results in two main ways. Firstly, warnings involving visual cues were perceived as more urgent in all signals with equal numbers of modalities. Namely, ATV warnings were rated as the most urgent, TV and AV as more urgent than AT, and V as more urgent than A and T. These results were all statistically significant. This shows the strength of the visual modality as a means to design urgent messages. However, in the experiments presented here there was not much visual attention required for the driving task. A further study will investigate the performance of all of these cues in a more complex simulator driving situation to see if the benefits of the visual cues still hold. Secondly, warnings using the tactile modality ware rated as more annoying. This was verified by participants' comments, mentioning that Tactile was often not liked.

In terms of recognition times, again signals using the Visual modality lead to quicker responses, while signals involving Tactile lead to slower responses and more mistakes, a result that was in line with [4]. Taken together, these results highlight that Visual, used in isolation or combined with Audio or Tactile is a promising modality for conveying urgency both quickly and accurately. Tactile on the other hand should be used with caution, as it may create higher annoyance and slower and less accurate responses. Further studies will investigate if the problems with tactile are fundamental, or just with the design used in these experiments. We will investigate different body locations for tactile feedback (for example seat or shoulder) and use other features like roughness (see also [12]) to see if the limitations can be ameliorated.

An interesting effect observed was the influence of NoM in ratings as well as recognition times. Warnings with three modalities were rated more urgent than warnings with two modalities and the latter more urgent than warnings with one modality. The effect of annoyance was also present as more modalities were used, but not as strong. Finally, warnings with three modalities created quicker responses compared to warnings with two modalities, which in turn were quicker compared to warnings with one modality. Interestingly, there was no difference in terms of recognition accuracy as more modalities were used. In studies like [9, 15, 25], modality combinations presented better results than the modalities in isolation, but no study to our knowledge has found as clear results in terms of how the number of modalities used affects responses for driver displays. A clear guideline for warning design is that NoM can be used to convey urgency without sacrificing recognition accuracy. They can also create responses that vary according to the NoM used. Future studies will investigate whether this effect can be replicated when using richer cues in the driving context, for example speech, text or tactile roughness.

Additionally, there was little effect of whether the simulator was present or absent in this study. The strongest effect found was that participants made significantly more mistakes in recognizing modalities when the simulator was present, which could be justified by the higher cognitive load required with the driving task. Although the driving task used was relatively simple, this result illustrates the robustness of the cues across contexts. Finally, it is noted that the results of this study were acquired using a simulated driving task, thus the degree to which they can be generalized to a real driving situation should be investigated.

7. CONCLUSIONS

The work reported in this paper utilised all multimodal combinations of Audio, Visual and Tactile modalities to alert drivers to events with varying urgency. It was found that the cues were clearly identified both in terms of perceived urgency as well as recognition time. Perceived annoyance was not as high, indicating the appropriateness of the cues for the driving context. These findings extend available results in all modality combinations used and in the context of a driving simulator. The strength of cues involving visual display in conveying messages quickly and accurately, as well as some limitations of utilising tactile cues for warnings were also highlighted. In addition, more modalities meant quicker and more accurate responses, as well as higher perceived urgency, without a large increase in perceived annoyance. The potential of using the number of modalities to convey urgency is a new result for automotive warning design. Taken together, the results of this study can inform the design of effective multimodal warnings for drivers.

8. ACKNOWLEDGEMENTS

This work was partly funded by Freescale Semiconductor Inc.

9. REFERENCES

[1] C. L. Baldwin, J. L. Eisert, A. Garcia, B. Lewis, S. M. Pratt, C. Gonzalez. 2012. Multimodal urgency coding: auditory, visual, and tactile parameters and their impact on perceived urgency. *Work: A Journal of Prevention, Assessment and Rehabilitation* 41, 1, 3586-3591.

[2] S. A. Brewster, P. C. Wright, A. D. Edwards. 1993. An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings INTER CHI'93*. ACM Press, 222-227.

[3] D. Brumby, V. Seyedi. 2012. An empirical investigation into how users adapt to mobile phone auto-locks in a multitask setting. In *MobileHCI 2012*. ACM Press, 281-290.

[4] Y. Cao, F. Van Der Sluis, M. Theune, A. Nijholt. 2010. Evaluating informative auditory and tactile cues for in-vehicle information systems. In *Proceedings Automotive User Interfaces and Interactive Vehicular Applications*. ACM Press, 102-109.

[5] J. Edworthy, E. Hellier, K. Walters, B. Weedon, A. Adams. 2000. The Relationship between Task Performance, Reaction Time, and Perceived Urgency in Nonverbal Auditory Warnings. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 674-677.

[6] J. Edworthy, S. Loxley, I. Dennis. 1991. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *The Journal of the Human Factors and Ergonomics Society* 33, 2, 205-231.

[7] C. Gonzalez, B. A. Lewis, D. M. Roberts, S. M. Pratt, C. L Baldwin, 2012. Perceived urgency and annoyance of auditory alerts in a driving context. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 1684-1687.

[8] C. Ho, C. Spence. 2005. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology Applied* 11, 3, 157.

[9] C. Ho, N. Reed, C. Spence. 2007. Multisensory in-car warning signals for collision avoidance. *The Journal of the Human Factors and Ergonomics Society* 49, 6, 1107-1114.

[10] C. Ho, H. Z. Tan, C. Spence. 2005. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 6, 397-412.

[11] E. E. Hoggan, S. A. Brewster. 2006. Crossmodal icons for information display. In *ACM CHI'06*, ACM Press, 857-862.

[12] E. Hoggan, R. Raisamo, S. A. Brewster. 2009. Mapping information to audio and tactile icons. In *Proceedings of ACM ICMI 2009*, ACM Press, 327-334.

[13] Y. Huang, C. Tsai, J. Kuo, F. Wu. 2011. The Comparison of Different Sensory Outputs on the Driving Overtake Alarm System. *Universal Access in Human-Computer Interaction. Context Diversity,* 290-297.

[14] C. Kaufmann, R. Risser, A. Geven, R. Sefelin. 2008. Effects of simultaneous multi-modal warnings and traffic information on driver behaviour. In *Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems,*33-42.

[15] D. Kern, P. Marshall, E. Hornecker, Y. Rogers. 2009. Enhancing Navigation Information with Tactile Output Embedded into the Steering Wheel. In *Pervasive 2009*. Springer, 42-58.

[16] B. A. Lewis, C. L. Baldwin. 2012. Equating Perceived Urgency Across Auditory, Visual, and Tactile Signals. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 1307-1311.

[17] D. C Marshall, J. D. Lee, P. A. Austria. 2007. Alerts for In-Vehicle Information Systems: Annoyance, Urgency, and Appropriateness. *The Journal of the Human Factors and Ergonomics Society* 49, 1, 145-157.

[18] D. McKeown, S. Isherwood, 2007. Mapping candidate within-vehicle auditory displays to their referents. *The Journal of the Human Factors and Ergonomics Society* 49, 3, 417-428.

[19] M. A. Mollenhauer, J. Lee, K. Cho, M. C. Hulse, T. A. Dingus. 1994. The effects of sensory modality and information priority on in-vehicle signing and information systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 1072-1076.

[20] A. Murata, M. Kanbayashi, T. Hayami. 2012. Effectiveness of automotive warning system presented with multiple sensory modalities. *In Proceedings of SICE Annual Conference 2012*. IEEE, 920-925.

[21] S. M. Pratt, B. A. Lewis, B. N. Peñaranda, D. M. Roberts, C. Gonzalez, C. L. Baldwin. 2012. Perceived Urgency Scaling in Tactile Alerts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 1303-1306.

[22] B. Saket, C. Prasojo, Y. Huang, S. Zhao. 2013. Designing an effective vibration-based notification interface for mobile phones. In *Proceedings CSCW 2013*. ACM Press, 149-1504.

[23] J. J. Scott, R. Gray. 2008. A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *The Journal of the Human Factors and Ergonomics Society* 50 2, 264-275.

[24] J. M. Sullivan, M. L. Buonarosa. 2009. Acquisition, response, and error rates with three suites of collision warning sounds. In *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* , 284-290.

[25] J. B.Van Erp, H. A. Van Veen. 2004. Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour* 7, 4, 247-256.