To Beep or Not to Beep? Comparing Abstract versus Language-Based Multimodal Driver Displays

Ioannis Politis[1], Stephen Brewster[2]
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

Multimodal displays are increasingly being utilized as driver warnings. Abstract warnings, without any semantic association to the signified event, and language-based warnings are examples of such displays. This paper presents a first comparison between these two types, across all combinations of audio, visual and tactile modalities. Speech, text and Speech Tactons (a novel form of tactile warnings synchronous to speech) were compared to abstract pulses in two experiments. Results showed that recognition times of warning urgency during a non-critical driving situation were shorter for abstract warnings, highly urgent warnings and warnings including visual feedback. Response times during a critical situation were shorter for warnings including audio. We therefore suggest abstract visual feedback when informing drivers during a non-critical situation and audio in a highly critical one. Language-based warnings during a critical situation performed equally well as abstract ones, so they are suggested as less annoying vehicle alerts.

Author Keywords
Multimodal feedback; warnings; audio; visual; tactile; speech; Tactons; urgency; recognition; response.

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

INTRODUCTION

Informing drivers has become easier with the availability of rich in-car displays. Car manufacturers use these displays to present drivers with a wide range of information, such as vehicle-related updates or collision warnings. Messages can be abstract, e.g. repeated pulses with no semantic association to the events signified [27] or more informative, e.g. speech, having a higher association to the event [29]. Additionally, such warnings can use any of the audio, tactile or visual modalities. Previous experiments have evaluated the performance of abstract versus more informative audio [11,18,25] or tactile cues [13]. However, no research has studied all multimodal combinations of these warnings and how their simultaneous presentation affects responses. This is important in order to provide guidelines on the effectiveness of these two types of messages and the best modalities to utilize for message display. Further, Speech Tactons, tactile patterns synchronous to speech [29], presented promising results when combined with speech warnings. However, they have never been tested in the driving context, which may affect performance.

In this paper, we address the above and present two experiments investigating responses to abstract versus language-based multimodal warnings varying in urgency in a driving simulator for the first time. Abstract warnings were repeated pulses along the audio, visual and tactile modalities, as well as all their bimodal and trimodal combinations. Language-based warnings were delivered in the same modalities and were speech, text or Speech Tactons. All warnings were evaluated in terms of recognition time of the cues' urgency and response time to high urgency cues.

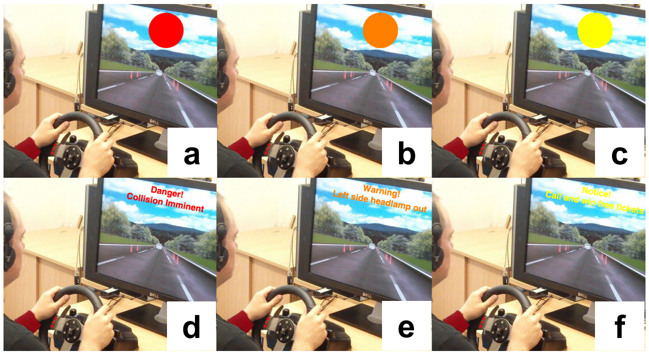

Figure 1: The experimental setup. The visual signals are Abstract L_H (a), L_M (b) and L_L (c) and Language-based L_H (d), L_M (e) and L_L (f).

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04…$15.00
http://dx.doi.org/10.1145/2702123.2702167

In summary, we can derive the following guidelines from this work:

- Abstract cues have quicker recognition in a low criticality task, i.e. recognizing warning urgency with no critical event on the road;
- Multimodal cues including visuals are suitable for the same task, since participants rely on a visual interpretation of the cues;
- Multimodal cues including audio create quicker responses in a high criticality task, i.e. responding to a car in front braking sharply;
- Abstract and language-based cues have similar response times when a critical event is presented and can be used interchangeably. However, the use of language-based cues marginally improves driving performance;
- In high urgency situations, the use of warnings leads to a slight degradation of steering performance, but warnings are still suggested, since they improve response times.

RELATED WORK

Using multimodal warnings to alert drivers

Previous work has shown the utility of multimodal warnings to alert drivers to various situations on the road. Ho, Tan & Spence [19] used simple spatial vibrotactile cues, coming from the direction of an approaching threat, in order to decrease reaction times to a simulated critical event while driving. Ho & Spence [18] found advantages in reaction times when using a car horn sound and the words *"front"* or *"back"* indicating the direction of a threat. All sounds performed better when coming from the direction of a rapidly approaching vehicle (front or back), and, in this case, speech related messages led to shorter reaction times compared to car horn sounds. Ho, Reed & Spence [16] elaborated on these results by showing that reactions can be quicker when audio and tactile messages were delivered in combination, with a car horn sound used along with a simple vibrotactile cue.

Comparing unimodal messages in the audio, tactile and visual modalities, Scott & Gray [31] observed quicker responses to vibrotactile messages compared to simple visual cues and simple tones. Serrano *et al.* [32] verified the advantage of directional speech cues for identifying whether a presented road scene was hazardous or not. The above studies show the benefits of multimodal cues for driving, especially when the direction of the cues corresponds to the direction of the impending danger. However, the cues where either unimodal or bimodal in one case, and no systematic comparison between abstract and language-based cues was attempted. In our study, we used the benefits of directionality by presenting our warnings from the appropriate direction, in this case the front. We also compared all unimodal, bimodal and trimodal combinations of audio, visual and tactile cues, as well as abstract, i.e. repeated pulses, versus language-based messages, to explore fully the effectiveness of these different warning types.

Designing urgency in warnings

Other than alerting drivers, car warnings should reflect the urgency of the signified situation. There is a rich body of literature on how to design differently urgent abstract warnings. Edworthy, Loxley & Dennis [10] found that higher fundamental frequency, speed and pitch range increased the perceived urgency ratings for auditory warnings. Edworthy *et al.* [9] found shorter response times to highly urgent warnings compared to medium and low urgency ones. Marshall, Lee & Austria [23] showed that higher pulse duration and lower interpulse interval increased ratings of urgency of audio alerts. Gonzalez *et al.* [12] further described the influence of higher fundamental frequency, pulse rate and intensity to the ratings of urgency and annoyance of sound warnings. Pratt *et al.* [30] observed that pulse rate also increased the ratings of urgency for tactile alerts. Lewis & Baldwin [20] created a crossmodal urgency scale, where pulse rate (or flash rate for visual signals) was suggested as an effective means to vary urgency in all of these modalities. Increased sound intensity and frequency were effective for audio signals, while colours were used for visual ones.

In terms of language-based warnings, urgency, annoyance and alerting effectiveness have been investigated in the past. Baldwin & Moore [1] suggested the use of the word *"Danger"* to increase ratings of perceived urgency of speech. *"Warning"* and *"Caution"* showed intermediate results, while *"Notice"* was perceived as the least urgent. Higher S/N ratio increased ratings of urgency, regardless of the content of the messages. Baldwin [2] also observed lower reaction times to highly urgent words, presented with high signal intensity. Hellier *et al.* [15] demonstrated how urgently spoken signal words increased ratings of urgency compared to non-urgently spoken ones, which in turn had higher ratings compared to words spoken in a monotone manner. Edworthy *et al.* [8] showed that signal words spoken urgently are perceived as more urgent and appropriate.

Using all combinations of audio, visual and tactile modalities and the above guidelines, Politis, Brewster & Pollick [27] designed a set of abstract warnings. They consisted of repeated pulses, across three levels of urgency. It was found that perceived urgency increased and recognition time decreased, as modalities used increased from one, to two, to three. Using the same set of warnings, Politis, Brewster & Pollick evaluated the influence of a critical event, i.e. a simulated lead car braking, along with the exposure to the signals [28]. They found improved reaction times when the warnings accompanied such an event. Finally, using a set of language-based messages along three different urgency levels, Politis, Brewster & Pollick [29] found that when speech was accompanied by Speech Tactons the recognition of warnings' urgency and their perceived effectiveness improved. However, in [29] there was no driving task, which would increase ecological validity and is vital if the cues are to be used in real driving. In our study, we combine the use of abstract warnings from [27,28] and language-based ones from [29] to make a comparison across all combinations of audio, tactile and visual modalities, along three different urgency levels using a driving simulator. We evaluate the cues in terms of recognition and reaction times, providing new quantitative results for both high and low criticality driving tasks, which has not been attempted in the past.

Comparing Abstract and Informative Warnings

A direct comparison between some abstract and more informative warnings has been investigated. McKeown & Isherwood [25] experimented with more complex sound cues of varying content in order to alert drivers. They used abstract sounds, environmental sounds, Auditory Icons and speech. When participants were matching these sounds to the appropriate driving events, abstract sounds had the highest response times and lowest accuracy, while speech and Auditory Icons had the lowest response times and the highest accuracy. Speech was perceived as more pleasant and less urgent compared to the abstract sounds. McKeown, Isherwood & Conway [24] later compared repetitive pulses and a gunshot sound with the sound of screeching brakes and found lower response times when participants reacted to the latter, which had a higher association with driving. Although useful, the above studies only attempt comparisons in audio. We extend this by using all combinations of audio, visual and tactile cues to assess the influence of modality and designed urgency on response performance.

Cao *et al*. [6,7] used speech, abstract audio cues and visuals to present road obstacle warnings in a simulator. Speech combined with pictures led to high recall of the signified events when asked about them after the experiment, and low reaction times. The use of speech along with images was therefore suggested by the authors for tasks not requiring imminent responses, such as navigation. In more demanding situations, like low visibility and under fatigue, speech and images were also perceived as most useful. One of the limitations of these studies, as mentioned by the authors, is the relatively long utterances of the speech cues, some of which were as long as ten words. We address this in our study, by using speech cues whose text is between three and six words long.

More recently, there has been interest in evaluating audio and tactile signals whose intensity and location change with time. Gray [14] found low reaction times and high response accuracy to audio warnings with looming intensity, i.e. intensity increasing as the danger was approaching. These warnings outperformed abstract pulses and a car horn sound. Ho, Spence & Gray [17] confirmed the good results of looming intensity for audio, while no additional benefit of looming intensity was found in the tactile modality. Gray, Ho & Spence [13] however found decreased response times in the tactile modality compared to constant pulses, when looming intensity was combined with apparent motion towards the drivers' head, created by a vertical array of three tactors attached on the abdomen and activated in an upward manner. A variation of apparent motion was tested by Meng *et al.* [26] by activating vibrotactile cues first on participants' hands and then on their torso, creating a sense of cues moving towards the torso. This intervention produced lower response times compared to static cues, while looming intensity of vibration showed again no additional benefits. The above studies present an interesting application of varying audio and tactile intensity or location to alert drivers. Although not looming or moving, our warnings use changes in intensity, which increases with urgency.

The above work provides an indication of the potential of warnings with a stronger semantic association to the event signified compared to abstract ones. However, no attempt has been made to test the effectiveness of these warnings multimodally, something we address in our study. We compare abstract and language-based warnings, taking into account their designed urgency by evaluating our cues in three different urgency levels. Other than Edworthy, Walters & Hellier [11], who found no difference in perceived urgency between speech and non-speech audio warnings, no other study has directly compared these two types of alerts taking into account the designed urgency of the cues, identifying in this way the best modalities for each case.

WARNING DESIGN

In order to compare responses to abstract versus language-based warnings, cues from [28] and [29] were used, utilizing respectively repeated pulses and language-based messages and presented 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).

Abstract Warnings

The abstract warnings consisted of repeated tones and were similar to [28]. As in [28], three Levels of Designed Urgency (LDU) were created, indicating conditions varying in importance. L_H (Level High) signified situations of high urgency, such as an impending collision, L_M (Level Medium) situations of medium urgency, such as a broken headlamp and L_L (Level Low) situations of low urgency, such as an advertisement. There were 21 signals: 7 signals with the above modalities (A, T, V, AT, AV, TV, ATV) × 3 Levels of Designed Urgency. The warnings consisted of pure tones, colours or vibrations delivered as repeated pulses. Pulse rate increased as signals became more urgent, as in [20,28]. Warnings of the same urgency level had the same pulse rate, independent of modality. 8 pulses having 0.1 *sec* single pulse duration and interpulse interval were used for L_H, 5 pulses having 0.17 *sec* single pulse duration and interpulse interval for L_M and 2 pulses having 0.5 *sec* single pulse duration and 0.5 *sec* interpulse interval for L_L. All warnings had 1.5 *sec* duration. Auditory warnings were varied additionally in base frequency, as in [10,20,23]

(1000 Hz for L_H, 700 Hz for L_M and 400 Hz for L_L).Visual warnings were also varied in colour, in line with [20,28] (Red for L_H, Orange for L_M and Yellow for L_L [1]). A C2 Tactor from Engineering Acoustics[2] was used for the tactile warnings. Tactile warnings had a frequency of 250 Hz, the nominal centre frequency of the C2. The above warnings showed significantly different ratings of perceived urgency in [28] and were selected as good candidate abstract signals to convey differently urgent events multimodally.

Contrary to the fixed intensity of [28], we decided to decrease the intensity of audio and tactile cues as their designed urgency decreased for three reasons. Firstly, annoyance levels in [28] were higher in the tactile modality, an effect that was ameliorated by varying intensity as urgency decreased in [29]. Secondly the good recognition results achieved for language-based cues in [29] provided a good potential for a similar result for the abstract tones. Finally, we wished to have a fair comparison between abstract and language-based warnings and avoid any observed effects to be accounted on different intensities. Therefore, we used the intensity of speech cues of the same urgency level, described below, also in the abstract cues. Thus, in both audio and tactile cues, L_H messages had a peak of -1.9 *dBFS*, L_M had a peak of -11.1 *dBFS* and L_L had a peak of -16.5 *dBFS*. Simultaneous delivery of unimodal signals was used in the multimodal ones, to create a synchronous effect of sound, vibration, visuals and all their combinations.

Language-based warnings

The language-based warnings used were the best performing cues in terms of recognition accuracy from [29]. Three speech messages designed to convey three different urgency levels, L_H, L_M and L_L were used: *"Danger! Collision Imminent"* for L_H, *"Warning! Left side headlamp out"* for L_M and *"Notice! Call and win free tickets"* for L_L. All messages were recorded by a female voice actor, who in line with [15] was instructed to narrate the message of L_H in an urgent manner, as if a loved one was in imminent danger. The L_M message was spoken non-urgently, as if in a friendly conversation with nothing interesting about the situation and the L_L message was spoken in a monotone, deadpan manner. The L_H message was 1.7 *sec* long and had a peak of -1.9 *dBFS* and an average frequency of 377 *Hz*. The L_M message was 2.7 *sec* long, had a peak of -11.1 *dBFS* and an average frequency of 285 *Hz*. Finally, the L_L message was 3.7 *sec* long, had a peak of -16.5 *dBFS* and an average frequency of 202 *Hz*.

For the Speech Tactons, all stimuli designed were auditory, to be used with a C2 tactor. To construct the auditory cues, the fundamental frequency F_0 (pitch) of each sample of the speech recordings was obtained, which resulted in alternating pure tones for each utterance. Then, the changes in intensity of the original sound files were used in the tones. All tactile cues retained the rhythm and intensity variations of the original recordings. The resulting values of average frequency of all tactile cues never differed to the average frequency of the audio more than ±10Hz.

Finally, for the visual cues, the text of the warnings was displayed in the same colour as the abstract cues of the respective LDU (Red for L_H, Orange for L_M and Yellow for L_L). A possible limitation of this approach is that the effects of text meaning and text colour were not measured separately. However we chose to maintain a consistent colouring between abstract and language based visual cues for simplicity, in line with [20]. 21 different language-based cues were created, 7 cues with all modalities (A, T, V, AT, AV, TV, ATV) × 3 Levels of Designed Urgency. For all modifications, Praat[3] and Audacity[4] software were used.

In all, there were 42 different warnings, 21 abstract and 21 language-based ones[5]. These warnings were evaluated in two experiments, looking into how quickly and accurately participants would respond when exposed to them.

EXPERIMENT 1: RECOGNITION TIME

The first experiment evaluated how quickly and accurately participants were able to recognize the level of urgency of the presented multimodal warnings. A 7×3×2 within subjects design was used with Modality, Level of Designed Urgency (LDU) and Information as the independent variables and Recognition Time (RecT) and Recognition Accuracy (RecA) as the dependent variables. Modality had 7 levels (A, T, V, AT, AV, TV, ATV), LDU had 3 levels (L_H, L_M, L_L) and Information had 2 levels (Abstract, Language-based). There were the following hypotheses:

- RecT will be influenced by Modality (H_{1a}), LDU (H_{1b}) and Information (H_{1c});
- RecA will be influenced by Modality (H_{2a}), LDU (H_{2b}) and Information (H_{2c}).

Procedure

Twenty participants (10 female) aged between 20 and 38 years (*M* = 25.05, *SD* = 5.11) took part in this experiment. They all held a valid driving license and had between 1 and 20 years of driving experience (*M* = 6.05, *SD* = 5.23). There were two left handed participants and all reported normal hearing and normal or corrected to normal vision. They were either University students or employees.

[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

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

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

[5] All warnings are available at http://goo.gl/XHViGY

The experiment took place in a University room, where participants sat in front of 27-inch Dell 2709W monitor and a PC running the simulator software (see Figure 1). In the software, a three lane road in a rural area was depicted, with a lead car maintaining a steady speed in the central lane. This simulator has been used in several previous research studies, e.g. [4]. As in [4], safety cones were placed on either side of the central lane, to reinforce lane keeping. Participants used a Logitech G27 gaming wheel and pedals to steer the simulated vehicle and to brake. Inputs were logged with a frequency of 50 *Hz*. Participants wore a set of Sennheiser HD 25-1 headphones and a wristband on their left wrist with a C2 Tactor attached on the inside of the band, in line with [29,30]. This simulated tactile feedback being presented by a smart watch. To cover any noise from the Tactor, car sound was played throughout the experiment. For two participants, sound and vibration were slightly adjusted to maintain comfortable intensities. Visual abstract cues were delivered through coloured circles that flashed in the top central area of the screen, and were sized 400×400 pixels (about 12×12 cm). Visual language-based cues were coloured text displaying each warning, which appeared once and for as long as the warning was uttered in the top central area of the screen, and were sized 200×800 pixels (about 24×6 cm). The visual cues did not obstruct the lead car and were designed to simulate a Head-Up Display. Abstract and language-based visual cues were also designed so as to occupy roughly the same area on the screen (about 144 cm^2). Figure 1 shows the setup and visual cues.

Participants were welcomed and provided with a brief introduction to the experiment. Afterwards, the participants were exposed to all the warnings as follows: first, a label with the text *"Level High (H) Warnings of HIGH urgency will follow"* appeared on the screen, then the 7 abstract warnings of L_H were played once to half of the participants, in the following order: A → T → V → AT → AV → TV → ATV and then the 7 language-based warnings of L_H were played in the same order of modality. Afterwards, a label with the text *"Level Medium (M) Warnings of MEDIUM urgency will follow"* appeared and then the 7 abstract warnings of L_M were displayed followed by the 7 language-based L_M ones, keeping the same order for modalities. Finally, a label with the text *"Level Low (L) Warnings of LOW urgency will follow"* appeared, followed by the 7 L_L abstract and then the 7 L_L language-based warnings as above. To the other half of the participants, first the language-based cues were played in each LDU and then the abstract ones, in the same manner as above. This procedure was chosen to minimize any order effects when presenting abstract and language-based cues, while still presenting them in a memorable way. The training lasted about 6 *min* for each participant. Participants were then asked to drive for 90 *sec*, to get accustomed to the simulator.

In the main part of the study, participants were presented with a driving scene, where they drove a simulated vehicle along a straight rural road following a car in front. Participants were able to steer the vehicle but did not use the accelerator pedal. The vehicle controlled by the participants maintained a constant speed of just above 70 *mph*. This speed was chosen in order to exceed the UK motorway speed limit (70 *mph*) creating a hazardous driving situation and requiring the drivers' attention. While steering the vehicle, the warnings were displayed to the participants in a random order and with a random interval of any integral value between (and including) 11–19 *sec*. These values were chosen to be similar with previous driving studies with repeated exposure to stimuli, e.g. [27]. Each stimulus was played twice. This resulted in a total of 82 stimuli (42 warnings × 2 presentations). Participants were asked to identify the urgency level of each stimulus by pressing one of three buttons on the steering wheel as quickly as possible. Buttons were labelled with letters (H, M or L) according to the urgency levels – topmost for L_H, middle for L_M, bottom for L_L. Participants were asked to maintain a central lane position. The whole experiment lasted about 30 *min* and participants were then prepared for the next experiment, which followed immediately.

Results

Recognition Time

All data for recognition time were analysed using a three-way repeated measures ANOVA, with Modality, Level and Information as factors. Mauchly's test showed that the assumption of sphericity had been violated for Modality and the interaction between Modality and LDU and Modality and Information. Therefore Degrees of freedom were corrected with Greenhouse–Geisser sphericity estimates.

There was a significant main effect of Modality ($F(3.01,102.33) = 103.34$, $p < 0.001$). Contrasts revealed that AV, ATV, V and TV warnings elicited significantly quicker responses compared to A and AT ($F(1,34) = 22.77$, $r = 0.59$, $p < 0.001$), which in turn had quicker responses compared to T ones ($F(1,34) = 106.78$, $r = 0.87$, $p < 0.001$). There was a significant main effect of LDU ($F(2,68) = 74.13$, $p < 0.001$). Contrasts revealed that L_H warnings were recognised quicker compared to the L_M and L_L ones ($F(1,34) = 89.05$, $r = 0.85$, $p < 0.001$). There was a significant main effect of Information ($F(1,34) = 37.55$, $p < 0.001$). Contrasts revealed that Abstract warnings were recognised quicker than Language-based ones (1.41 *sec* on average for Abstract vs. 1.57 *sec* for Language-based warnings, $F(1,34) = 37.55$, $r = 0.72$, $p < 0.001$). As a result hypotheses H_{1a}, H_{1b} and H_{1c} were accepted. See Figure 2 for Recognition Times across Modalities (a) and LDU (b).

There was a significant interaction between Modality and LDU ($F(7.28,247.60) = 2.63$, $p < 0.05$). Contrasts revealed that A warnings had slower recognition times compared to TV ones for L_M compared to L_H ($F(1,34) = 14.55$, $r = 0.55$, $p < 0.05$). Also, that although for TV warnings L_M cues were quicker than L_L in recognition, this effect was reversed for A ($F(1,34) = 4.73$, $r = 0.35$, $p < 0.05$). Finally, that AT warnings had quicker recognition times compared to T ones for L_H compared to L_M ($F(1,34) = 6.04$, $r = 0.39$, $p < 0.05$). There was a significant interaction between Modality and Information ($F(3.68,125.07) = 26.82$, $p < 0.001$). Contrasts revealed that for TV warnings Abstract signals were recognised quicker than Language-based ones, but this was reversed for A warnings ($F(1,34) = 11.87$, $r = 0.51$, $p < 0.05$). Also, while for AT warnings Language-based signals were recognised quicker than Abstract ones, this was reversed for T ($F(1,34) = 52.37$, $r = 0.78$, $p < 0.05$).

Recognition Accuracy

In all, there were 1512 participant responses and only 1 trial where a participant failed to respond. For the rest, 1366 responses were correct (90.4%) and 145 incorrect (9.6%). 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 in modality T compared to all the rest of the modalities. Specifically, out of 228 trials for each modality, there were 80 mistakes for T versus 16 for A ($Q(1) = 47.63$, $p < 0.001$), 17 for V ($Q(1) = 46.69$, $p < 0.001$), 18 for AT ($Q(1) = 46.89$, $p < 0.001$), 9 for AV ($Q(1) = 63.81$, $p < 0.001$), 12 for TV ($Q(1) = 57.80$, $p < 0.001$) and 13 for ATV ($Q(1) = 54.08$, $p < 0.001$). Contrasts also revealed that there were significantly more mistakes in L_L compared to L_H and L_M. Specifically, out of 532 trials for each level, there were 80 mistakes for L_L versus 38 for L_H ($Q(1) = 16.04$, $p < 0.001$) and 47 for L_M ($Q(1) = 10.78$, $p < 0.001$). There was no significant difference in number of mistakes between Abstract and Language-based cues. Specifically, out of 798 trials for each type of Information, there were 88 mistakes for Abstract versus 77 for Language-based cues, $Q(1) = 0.83$, $p = 0.36$. As a result, hypotheses H_{2a} and H_{2b} were accepted and H_{2c} was rejected.

EXPERIMENT 2: RESPONSE TIME

The second experiment evaluated how quickly participants were able to respond to presented multimodal warnings of high urgency (L_H). Other than their response time to this task, two driving metrics suggested in studies such as [5,27] were used. These were the Root Mean Square Error (RMSE) of the vehicle's lateral deviation and steering angle. Lower lateral deviation and variation of the steering angle can indicate lower driver distraction [21,22]. Also, the variable Time was used to measure the effect of the warning presentation on participants' driving behaviour by comparing the metrics described above before and after the presentation of the warnings.

A 7×3×2×2 within subjects design was used with Modality, LDU, Information and Time as the independent variables. Response Time (ResT), Lateral Deviation (LatDev) and Steering Angle (SteAng) were the dependent variables. As in the previous experiment, Modality had 7 levels (A, T, V, AT, AV, TV, ATV), LDU had three levels (L_H, L_M, L_L) and Information had 2 levels (Abstract, Language-based).

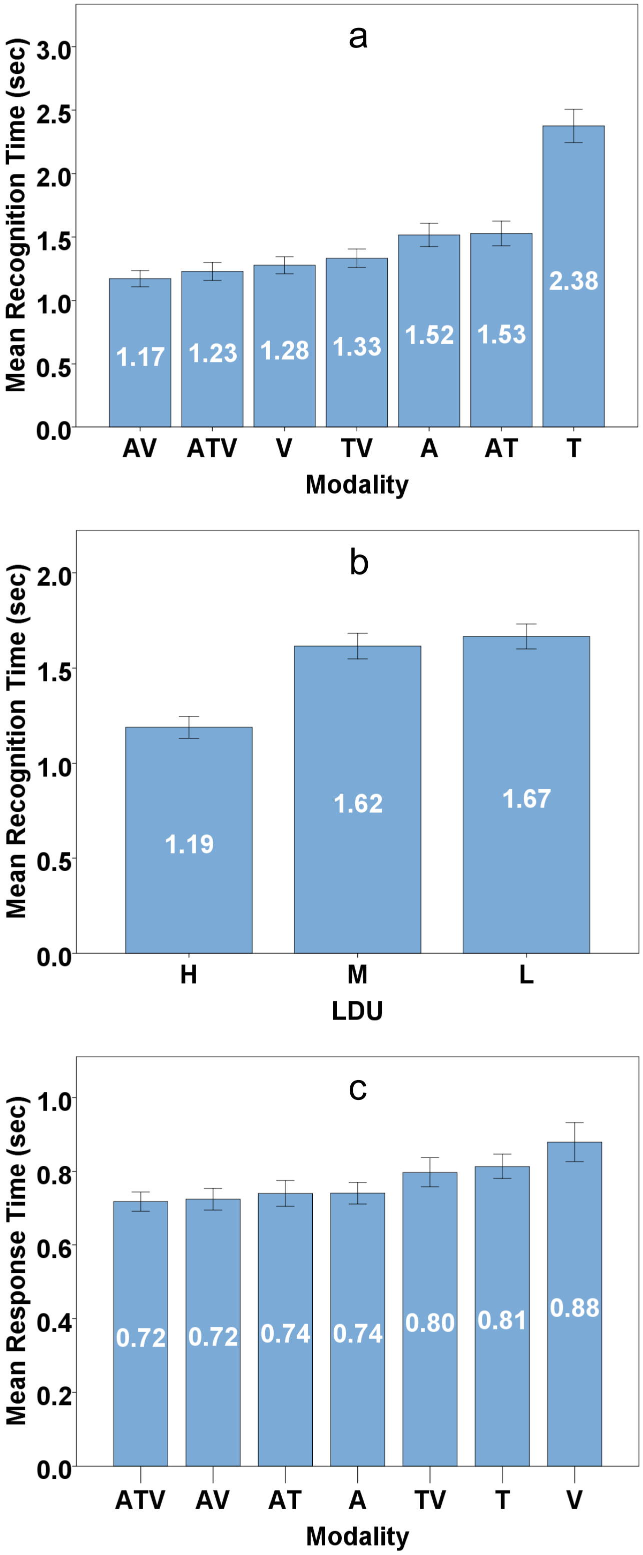

Figure 2: Recognition times for Experiment 1 across Modalities (a), Level of Designed Urgency (b) and response times for Experiment 2 across Modalities (c). Graphs are sorted by mean values. Error bars represent 95% Confidence Intervals.

Time had 2 levels: Before cue was presented and After cue was presented. There were the following hypotheses:

- ResT when reacting to L_H warnings and a car braking event will be influenced by Modality (H_{3a}) and Information (H_{3b});
- LatDev when reacting to L_H warnings and a car braking event will be influenced by Modality (H_{4a}), Information (H_{4b}) and Time (H_{4c});
- SteAng when reacting to L_H warnings and a car braking event will be influenced by Modality (H_{5a}), Information (H_{5b}) and Time (H_{5c});
- LatDev when exposed to L_M and L_L warnings without a car braking event will be influenced by Level (H_{6a}), Modality (H_{6b}), Information (H_{6c}) and Time (H_{6d});
- SteAng when exposed to L_M and L_L warnings without a car braking event will be influenced by Level (H_{7a}), Modality (H_{7b}), Information (H_{7c}) and Time (H_{7d}).

Procedure

Participants and equipment were identical to the previous experiment. It took place after participants completed the previous experiment and had a short break. Participants were presented with the same driving scene showing a vehicle maintaining a constant speed of just above 70 *mph*. Other than steering the vehicle, participants were able to respond by pressing the brake pedal. While steering the vehicle, the warnings were again displayed to the participants in a random order and with a random interval of any integral value between (and including) 11–19 *sec*. Each stimulus was again played twice, resulting in a total of 82 stimuli (42 warnings × 2 presentations). When there was a L_H warning, the vehicle in front started braking towards the participant vehicle along with the presentation of the warning. In case of an L_M or L_L warning, the vehicle in front continued driving and did not brake.

Participants were asked to maintain a central lane position throughout the experiment. They were instructed to respond by pressing the brake pedal as quickly as possible when there was a L_H warning presented along with the car in front braking. Finally, they were instructed to ignore the L_M and L_L warnings and not to respond to them. This process was chosen because responding to some warnings and ignoring others would create an increased workload for participants, requiring higher attention. Also, as shown in [28], the presentation of warnings along with a critical event resulted in quicker response times, which is desired in this situation. Finally, testing responses to L_H warnings was considered as more ecologically valid, since participants would not have to respond promptly to L_M and L_L warnings in a real setting.

Participants' ResT was calculated from the onset of the L_H stimulus and start of the braking event of the lead car, until the participant first pressed the brake pedal. Their LatDev and SteAng were logged for 4.7 *sec* (from 5.7 *sec* to 1 *sec* before any stimulus was displayed), forming their baseline value for driving performance. They were logged again for 4.7 *sec* immediately after the stimulus to assess the warning effects on driving. The value of 4.7 *sec* was chosen, since it was the duration of the longest of all cues (3.7 *sec*), increased by 1 *sec*. Thus, any effects occurring throughout the longest possible duration of a cue would be recorded.

For both LatDev and SteAng, the RMSE values were then computed from the logged values. As a result, out of the 82 overall trials, there were 28 values of ResT ([7 Abstract L_H cues + 7 Language-based L_H cues] × 2 presentations). Also, since LatDev and SteAng were logged in all cases (L_H, L_M and L_L cues), for each of the 82 trials there were two values for their LatDev (baseline value and value after the cue was displayed) and two values for their SteAng (baseline value and value after the cue was displayed). The whole experiment lasted about 30 *min* and participants were then debriefed about the purpose of both experiments and paid £6 for their participation.

Results

Response Time

Data for response times to L_H cues were analysed using a two-way repeated measures ANOVA, with Modality and Information 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 Modality ($F(4.17,154.37) = 18.83$, $p < 0.001$). Contrasts revealed that ATV, AV, AT and A warnings caused quicker responses compared to TV and T ones ($F(1,37) = 7.46$, $r = 0.41$, $p < 0.05$), which in turn had quicker responses compared to T warnings ($F(1,37) = 6.92$, $r = 0.40$, $p < 0.05$). There was no significant effect of Information ($F(1,37) = 1.37$, $p = 0.25$). As a result, H_{3a} was accepted and H_{3b} was rejected. See Figure 2 for Response Times across Modalities (c).

Lateral Deviation and Steering Angle

Data for LatDev when reacting to L_H cues by braking were analysed using a three-way repeated measures ANOVA, with Modality, Information and Time as factors. Time had two levels: Before cue was displayed (baseline data) and After cue was displayed. There was a significant main effect of Time ($F(1,39) = 5.65$, $p < 0.05$). Contrasts revealed that LatDev was higher after the cues were displayed (0.47 *m* on average before the L_H cue and the car braking event vs. 0.51 *m* after the cue and the event, $F(1,39) = 5.65$, $r = 0.36$, $p < 0.05$). There was an effect of Information which approached significance ($F(1,39) = 4.00$, $p = 0.053$), suggesting that the average LatDev both before and after the exposure to Language-based cues may be lower compared to Abstract cues, but not significantly so. No other significant effects were observed.

Data for SteAng when reacting to L_H cues by braking were also analysed using a three-way repeated measures ANOVA, with Modality, Information and Time as factors.

As above, Time had two levels: Before cue was displayed (baseline data) and After cue was displayed. There was a significant main effect of Time ($F(1,39) = 26.63$, $p < 0.001$). Contrasts revealed that SteAng was higher after the cues were displayed (0.07 *rad* on average before the L_H cue and the car braking event vs. 0.08 *rad* after the cue and the event, $F(1,39) = 26.63$, $r = 0.64$, $p < 0.001$). There were no other significant effects observed.

Data for LatDev and SteAng when exposed to L_M and L_L cues without reacting were analysed using a four-way repeated measures ANOVA, with Modality, Level, Information and Time as factors. There were no significant effects observed. As a result, H_{4c} and H_{5c} were accepted and H_{4a}, H_{4b}, H_{5a} and H_{5b} were rejected. Further, H_{6a} - H_{6d} and H_{7a} - H_{7d} were all rejected.

DISCUSSION

Recognition Time and Accuracy

Results for recognition time showed an advantage of abstract cues when identifying the level of designed urgency (LDU). This can be partly explained by the fact that these cues were shorter in length overall. In studies like [6], speech cues also created longer response times compared to abstract ones. We note, however, that our task was an identification one, requiring recall of the cues' LDU. As will be discussed later, this is different to a simple response task, in which abstract and language-based cues performed similarly in our study. As a guideline, abstract multimodal cues can cause quicker identification compared to language-based ones in a non-critical task. Although we designed the speech cues so that the distinctive word describing their LDU (*Danger*, *Warning* or *Notice*) came first in the message, it seemed that the cue length still required more time to interpret compared to the short abstract pulses. Interestingly, language-based warnings performed better in terms of RecT compared to abstract ones in modalities A and AT, indicating that when there is sound conveying the information, speech can also be a good means to do so. However, our above guideline still holds, since the main effect showed better performance of abstract cues overall. In future work, even simpler language-based cues could be evaluated to compare their recognition time with abstract ones.

In terms of modalities, results are very similar to [28], where only abstract cues were used. As in [28], the visual modality seemed to play an important role in participants recognizing the cues' LDU, since the modalities with the shortest recognition times all included visuals. Other than V, all other better performing cues were multimodal, similar to studies like [27,28,29], where the presence of more modalities enhanced responses. The presence of V in the group of best performing modalities confirms the role of visuals when interpreting such messages. However, V alone cannot be recommended for critical situations as it has been shown that it suffers in terms of performance when a visual critical event occurs [27]. Also in our study, as will be discussed later, V cues performed worse when users were reacting to an imminent collision event. As a guideline, abstract cues including visuals can be used to quickly inform about non-critical driving events. Combining results of perceived annoyance in [28], which increases as modalities used increase, we recommend bimodal rather than trimodal cues for this case.

A disadvantage of unimodal tactile cues in terms of recognition time was found. As also mentioned by several participants, cues were harder to identify when not accompanied by sound and / or visuals clarifying their meaning. This was also found in [28], where only abstract cues were evaluated, indicating that this difficulty holds also when language-based cues are used. The results of recognition accuracy, where the T cues were the worst performing compared to all other modalities, also add to this observed disadvantage. Thus, we recommend avoiding the delivery of messages through vibration alone when recognition time is important, since this may slow down their interpretation.

Finally, recognition of L_H cues was quickest, confirming that they were conveying an increased level of urgency. This was again in line with [28] and showed that the design used was effective in conveying high urgency in both abstract and language-based warnings. Combined with the recognition accuracy results, where L_L cues performed the poorest, we are confident that the cues designed afford quick recognition in more urgent situations. In terms of low urgency situations, we conclude that cues should be used cautiously, since a driver response is not essential in such cases (e.g. for an advertisement) and combined with the low performance observed, it may also be disruptive.

Response Time, Lateral Deviation and Steering Angle

Results for response time were similar to [27], with warnings including audio creating quickest responses compared to the rest. With the exception of A, all warnings in the quickest performing group were multimodal. This is in line with [27], although in that study all best performing warnings were multimodal. This enhanced performance of warnings including audio could be attributed to the reliance on cues different than visual for reacting to a visual critical event. As shown in [27], visual cues can suffer in terms of response times when users were exposed to a critical event in the simulator. This was confirmed in our study, since unimodal visual cues performed the poorest and only when accompanied by sound or by sound & vibration did they create quick responses. As a guideline, sound is a viable means of creating quick responses in highly urgent situations. We note that the task used in this study was a response one, where participants did not have to evaluate the cues' content, but rather automatically react. In this way, we were able to assess performance of cues in the presence of an event requiring an imminent response, where identifying the cues' content may be less critical.

Interestingly, there was no difference in performance between abstract and language-based cues. This is an indication that the designed language-based warnings perform as well as the abstract ones in this case. This is in line with [11], where no difference was found in terms of how urgent abstract and language-based warnings were perceived. Although not a perception task, our study showed similar response results for these two types of warnings. It could be concluded that as long the cues' content is clear, the response to such high urgency warnings is more affected by the modality they are delivered in (multimodally and including audio) than by their content (language-based or abstract sounds). Considering the results of [29], where language-based cues received low annoyance ratings overall, these cues seem to present an advantage over abstract pulses in a critical situation. As will be described below, a trend towards better lane keeping performance when exposed to the language-based warnings is an additional indication of this possible advantage.

In terms of lateral deviation and steering angle, the results showed that the exposure to L_H cues led to higher values and poorer lane-keeping. This is in line with [27], where the presence of cues did not improve or slightly worsened these metrics. We confirm therefore that the presence of cues along with critical events can create a distraction to the driving task. This is expected, since it is an additional factor for the driver to address. It has also been confirmed by studies like [3], where a startling effect of beeping cues was observed, leading to degradation of driving metrics. Additionally, since there is a physical reaction to the cues with braking, some increase in the driving metrics values is justified. As long as this increase is not dramatic, and as long as the set of cues improves response performance compared to the absence of them, as has been shown to do in [27], this is a necessary drawback when exposed to critical warnings. This also suggests that the use of warnings should be scarce, unless they signify critical events.

As described earlier, there was marginally better overall driving performance with the language-based cues, however the results did not reach significance. Therefore, we cannot provide a definite guideline on their advantage in this case, but they seem to create a trend towards better lane keeping behaviour. This could be addressing the problem of beeping cues in [3], since speech may avert startling effects created by abstract sounds. This new finding could be further examined in future work, by investigating the use of less prominent speech cues in critical situations, or using abstract looming warnings found in studies like [17], where intensity in the cues changes with time.

The presence of L_M and L_L cues, which had to be ignored, did not disturb the driving metrics. This is an important finding, since non-critical warnings should not add additional burden to the main task of driving. Participants were very accurate in discriminating the L_H cues from the L_M and L_L ones and in reacting to L_H. In only one case out of 1120 trials did a participant mistakenly react to a L_M cue and in no case did anyone react to a L_L one. These are encouraging results for all the cues designed, showing their suitability for use in contexts of intermediate or low criticality, which may occur more frequently when driving.

Finally, we note that since the results of this study were acquired using a simulated driving task, their generalizability to a road situation should be investigated in the future.

CONCLUSIONS

This paper describes two experiments presenting the first evaluation of responses to abstract versus language-based multimodal car warnings of varying urgency. All multimodal combinations of audio, tactile and visual warnings were evaluated in the driving context. Two tasks were used; a recognition task, where the cues' urgency was identified with no critical event present, and a response task, where responses to high urgency warnings were measured in the presence of such an event. An advantage of abstract warnings and warnings including visuals in the recognition task was observed. Cues including audio performed better in the response task. In both tasks, multimodal cues were the best performing ones, with the exception of unimodal visuals for recognition and unimodal audio for response. Driving behaviour, although slightly worsened by all cues in the critical situation, was marginally better when using language-based cues compared to abstract ones. These results show the benefit of using abstract cues in non-critical situations and a possible advantage of language-based cues in a critical situation. The derived guidelines can aid car warning designers and extend available knowledge by directly comparing these different ways of informing drivers.

ACKNOWLEDGEMENTS

The study was partly funded by Freescale Semiconductor Inc., Automotive Microcontroller Product Group.

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. Biondi, F., Rossi, R., Gastaldi, M., and Mulatti, C. Beeping ADAS: Reflexive effect on drivers' behavior. *Transportation Research Part F: Traffic Psychology and Behaviour 25*, (2014), 27–33.
4. Brumby, D. and Seyedi, V. An empirical investigation into how users adapt to mobile phone auto-locks in a multitask setting. *MobileHCI 2012*, ACM Press (2012), 281 – 290.
5. Brumby, D.P., Davies, S.C.E., Janssen, C.P., and Grace, J.J. Fast or Safe? How Performance Objectives

Determine Modality Output Choices While Interacting on the Move. *CHI 2011*, ACM (2011), 473–482.

6. Cao, Y., Mahr, A., Castronovo, S., Theune, M., Stahl, C., and Müller, C.A. Local danger warnings for drivers: The effect of modality and level of assistance on driver reaction. *Intelligent User Interfaces*, ACM Press (2010), 239–248.
7. Cao, Y., Theune, M., and Müller, C. Multimodal Presentation of Local Danger Warnings for Drivers : A Situation-dependent Assessment of Usability. *Professional Communication Conference (IPCC)*, IEEE (2010), 226–229.
8. 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.
9. Edworthy, J., Hellier, E., Walters, K., Weedon, B., and Adams, A. The Relationship between Task Performance, Reaction Time, and Perceived Urgency in Nonverbal Auditory Warnings. *HFES Annual Meeting 44*, 22 (2000), 674–677.
10. Edworthy, J., Loxley, S., and Dennis, I. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors 33*, 2 (1991), 205 –231.
11. Edworthy, J., Walters, K., Hellier, E., and Weedon, B. Comparing Speech and Nonspeech Warnings. *HFES Annual Meeting 44*, 22 (2000), 746–749.
12. Gonzalez, C., Lewis, B.A., Roberts, D.M., Pratt, S.M., and Baldwin, C.L. Perceived Urgency and Annoyance of Auditory Alerts in a Driving Context. *HFES Annual Meeting 56*, 1 (2012), 1684–1687.
13. Gray, R., Ho, C., and Spence, C. A comparison of different informative vibrotactile forward collision warnings: does the warning need to be linked to the collision event? *PloS one 9*, 1 (2014), e87070.
14. Gray, R. Looming Auditory Collision Warnings for Driving. *Human Factors 53*, 1 (2011), 63–74.
15. 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.
16. Ho, C., Reed, N., and Spence, C. Multisensory In-Car Warning Signals for Collision Avoidance. *Human Factors 49*, 6 (2007), 1107–1114.
17. Ho, C., Spence, C., and Gray, R. Looming Auditory and Vibrotactile Collision Warnings for Safe Driving. *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.*, (2013), 551–557.
18. Ho, C. and Spence, C. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of experimental psychology. Applied 11*, 3 (2005), 157–74.
19. Ho, C., Tan, H.Z., and Spence, C. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Traffic Psychology and Behaviour 8*, 6 (2005), 397–412.
20. Lewis, B.A. and Baldwin, C.L. Equating Perceived Urgency Across Auditory, Visual, and Tactile Signals. *HFES Annual Meeting 56*, 1 (2012), 1307–1311.
21. Lindgren, A., Angelelli, A., Mendoza, P.A., and Chen, F. Driver behaviour when using an integrated advisory warning display for advanced driver assistance systems. *IET Intelligent Transport Systems 3*, 4 (2009), 390–399.
22. Liu, Y.C. Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics 44*, 4 (2001), 425–42.
23. Marshall, D.C., Lee, J.D., and Austria, P.A. Alerts for In-Vehicle Information Systems: Annoyance, Urgency, and Appropriateness. *Human Factors 49*, 1 (2007), 145–157.
24. McKeown, D., Isherwood, S., and Conway, G. Auditory Displays as Occasion Setters. *Human Factors 52*, 1 (2010), 54–62.
25. McKeown, D. and Isherwood, S. Mapping Candidate Within-Vehicle Auditory Displays to Their Referents. *Human Factors 49*, 3 (2007), 417–428.
26. Meng, F., Gray, R., Ho, C., Ahtamad, M., and Spence, C. Dynamic Vibrotactile Signals for Forward Collision Avoidance Warning Systems. *Human Factors*, (2014).
27. Politis, I., Brewster, S., and Pollick, F. Evaluating Multimodal Driver Displays under Varying Situational Urgency. *CHI 2014*, ACM Press (2014), 4067 – 4076.
28. Politis, I., Brewster, S., and Pollick, F. Evaluating Multimodal Driver Displays of Varying Urgency. *Automotive UI 2013*, ACM Press (2013), 92 – 99.
29. Politis, I., Brewster, S., and Pollick, F. Speech Tactons Improve Speech Warnings for Drivers. *Automotive UI 2014*, ACM Press (2014), 1 – 8.
30. Pratt, S.M., Lewis, B.A., Penaranda, B.N., Roberts, D.M., Gonzalez, C., and Baldwin, C.L. Perceived Urgency Scaling in Tactile Alerts. *HFES Annual Meeting 56*, 1 (2012), 1303–1306.
31. Scott, J.J. and Gray, R. A Comparison of Tactile, Visual, and Auditory Warnings for Rear-End Collision Prevention in Simulated Driving. *Human Factors 50*, 2 (2008), 264–275.
32. Serrano, J., Di Stasi, L.L., Megías, A., and Catena, A. Effect of directional speech warnings on road hazard detection. *Traffic injury prevention 12*, 6 (2011), 630–635.