The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario

Geitner, Claudia[a]; Biondi, Francesco[b]; Skrypchuk, Lee[c]; Jennings, Paul[a] and Birrell, Stewart[a]*

[a]*Warwick Manufacturing Group (WMG). University of Warwick, International Digital Laboratory, Coventry, CV4 7AL, UK.*

[b]*Department of Psychology, University of Utah, Salt Lake City, UT 84112, US*

[c]*HMI Research Team. Jaguar Land Rover. University of Warwick, International Digital Laboratory, Coventry, CV4 7AL, UK.*

* Corresponding authors: (S. Birrell) s.birrell@warwick.ac.uk

The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during automated driving

In an automated car, users can fully engage in a distractor task, making it a primary task. Compared to manual driving, drivers can engage in tasks that are difficult to interrupt and of higher demand, the consequences can be a reduced perception of, and an impaired reaction to, warnings. In this study we compared three in-vehicle warnings (auditory, tactile, and auditory-tactile) which were presented during three highly attention capturing tasks (visual, auditory, and tactile) while the user was engaged in a self-driving car scenario, culminating in an emergency brake event where the warning was presented. The novel addition for this paper was that three set paced, attention capturing tasks, as well the three warnings were all designed in a pilot study to have comparable workload and noticeability. This enabled a direct comparison of human performance to be made between each of the attention capturing tasks, which are designed to occupy only one specific modality (auditory, visual or haptic), but remain similar in overall task demand. Results from the study showed reaction times to the tactile warning (for the emergency braking event) were significantly slower compared to the auditory and auditory-tactile (aka multimodal or multisensory) warning. Despite the similar reaction times between the in-vehicle auditory warning and the multimodal warning, the multimodal warning led to a reduced number of missed warnings and fewer false responses. However, the auditory and auditory-tactile warnings were rated significantly more startling than the tactile alone. Our results extend the literature regarding the performance benefits of multimodal warnings by comparing them with in-vehicle auditory warnings in an autonomous driving context. The set-pace attention capturing tasks in this study would be of interest to other researchers to evaluate the interaction in an automated driving context, particularly with hard to interrupt and attention capturing tasks.

Keywords: multimodal warning, tactile warning, driver distraction; autonomous vehicle; take-over of automation

1. Introduction

The development towards autonomous vehicles is one of the key trends in the automotive industry (Woodward et al., 2017). With autonomous vehicles drivers' task-demands change from actively driving to passively monitoring, with input from the driver being required up to the highest level of automation (SAE J3016, 2016). Indeed, drivers might be freed from so much of the driving task in a highly automated car that it can result in 'underload' (Young and Stanton, 1997; Körber, Cingel, Zimmermann, and Bengler, 2015). This underload increases the temptation to engage in another task (Carsten et al., 2012; Strayer and Fisher, 2015), and subsequent diversion of attention between tasks can delay reaction times (RT) towards an in-vehicle warning (Spence and Driver, 1997; Merat, Jamson, and Carsten, 2012). A possible resultant effect might lead to a safety critical situation when the driver needs to manually respond to an automation failure. In contrast to driver distraction in manually driven cars, effects in automated vehicles could be worse because a distractor task may even become the primary task (Blanco et al., 2015) which can be hard to interrupt. In this study we evaluated such a situation (simulated by a set paced, highly attention capturing task) in a self-driving scenario. We also investigated the efficacy of multimodal warnings – one that combine two or more sensory modalities (in this study auditory and tactile feedback) – in returning the attention of the driver back to the road and the driving task. Furthermore, we compared an industry developed in-vehicle auditory warning against a multimodal variant, in order to determine whether the multimodal warning elicits the same advantages which previous literature found in self-paced tasks during manual driving, in our highly distracting, self-driving vehicle simulated scenario.

The fatigue generated from more passive, monitoring tasks required of a driver in an automated vehicle (Desmond and Hancock, 2001; Körber et al., 2015) can increase the temptation to engage in non-driving related tasks (Carsten et al., 2012; Llaneras, Sallinger and Green, 2013; Lee, 2014; Naujoks, Purucker, and Neukum, 2016). Llaneras et al. (2013) found that drivers engaged more often in riskier, secondary tasks (e.g. dialling a mobile phone or reading) that involved longer off-road glances in a semi-automated driving condition (adaptive cruise control and automatic steering) compared to a purely driver assistance condition (adaptive cruise control only). Thus, in an automated vehicle context, different distractor tasks (such as set paced tasks) may be necessary to understand delayed driver response times (Eriksson and Stanton, 2017) or potentially missed feedback.

In-vehicle haptic feedback can help to reduce driver workload when conducting an unfamiliar driving task (Birrell, Young, and Weldon, 2013), increase safety by reducing the stop time in an emergency brake event when the automation system fails, or decrease the time for transition in a take-over scenario (see literature below). Previously it has been shown in manual driving that multimodal warnings can lead to a faster RT in high-workload situations compared to unimodal warnings (Brown, 2005; Ho, Reed, and Spence, 2007; Biondi et al., 2017). Similar benefits were found in automated driving, where drivers required considerably shorter times to detect a visual warning when combined with another modality, compared to a unimodal visual warning (Naujoks, Mai, and Neukum, 2014; Blanco et al., 2015). Research by Naujoks et al. (2014) showed that drivers returned their hands to the steering wheel faster while reading when exposed to a multimodal, visual-auditory warnings (2.29 s) compared to visual only warnings (6.19 s). Pjetermeijer, Bazilinskyy, Bengler, and de Winter (2017) found that drivers touched the steering wheel faster with a multimodal, auditory-tactile warning verses an auditory or tactile only warnings. Thus, multimodal warnings might be beneficial for a highly-automated vehicle when drivers are not required to actually drive and can focus on another task.

Despite the apparent benefits of multimodal warnings, they have been underused in the automotive industry to date in favour of more traditional unimodal visual or auditory warnings. Traditional warnings may be suitable for when the driver's primary task is driving, as their visual attention will be directed mainly forwards, towards the in-vehicle systems and/or roadway. However, in automated vehicles the user may be engaged in visually and auditory demanding tasks (e.g. watching a film or using a laptop) with their head, body or even their seat not in a single, uniformly consistent direction. This may result in traditional warnings being missed or increased reaction times, as shown in laboratory settings with secondary tasks (Spence and Ho, 2017). Supplementing traditional warnings with haptic feedback, within the specific use case of highly automated vehicles, warrants further research. This paper extends this knowledge about multimodal warnings towards a performance comparison between a multimodal, a currently used in-vehicle auditory, and a tactile warning while completing a highly attention capturing tasks in three discrete modalities, in a self-driving vehicle simulated scenario.

Within autonomous vehicle research, multimodal warnings are typically compared in tasks with a "single distractor" condition such as reading (Naujoks et al., 2014), the Surrogate Reference Task (Pjetermeijer et al., 2017), or a phone conversation (Biondi et al., 2017). Blanco et al. (2015) conducted one of the few studies which looked at a set of secondary tasks (web-browsing, e-mail, and navigation) where the participants decided the pace of those tasks. While a self-paced task gives opportunities for disengagement, in a naturalistic setting drivers might engage in tasks that are harder to interrupt such as e.g. with goals to achieve (Lee, 2017) or interest (Horrey et al., 2017) such as some kind of game. Artificial tasks with a set pace can simulate this engagement which could be considered difficult to interrupt.

In this paper, we compared the performance of a multimodal warning to a traditional in-vehicle warning while the driver completed a series of highly attention-capturing, set-paced, primary tasks. Equally weighted and similar tasks were presented in three modalities which are typically used for in-vehicle interactions: visual, tactile, and auditory. We compared an auditory-tactile, auditory-only, and tactile-only, in-vehicle warnings for an emergency braking event during a simulated autonomous driving scenario. The principles behind the Multiple Resources Theory (MRT; Wickens, 2002) were used to infer predictions for the performance of each warning type. Building on the MRT for this autonomous vehicle context, two tasks should interfere the most when they require the same resource, e.g. the same sensory channel. Thus, an auditory task is expected to interfere with an auditory warning, and the tactile task is expected to interfere with a tactile warning etc. A multimodal (auditory-tactile) warning is predicted to have an advantage in both task conditions, due to the information redundancy offered by the second modality. Hence, multimodal warnings can be more efficient, because on a perceptual level the two components of a multimodal warning can enhance each other (King and Calvert, 2001).

2. Materials and Method

This study evaluated two research questions:

1) Does a multimodal (auditory-tactile) warning perform more efficient compared to unimodal warnings, specifically a traditional used auditory warning, in distracting conditions that utilise the same sensor modality as one of its components?
2) Does the subjective perception of a multimodal warning differ (positively in terms of noticeability or negatively in terms of, e.g., annoyance) over three distracting conditions compared to a unimodal auditory and a tactile warning?

2.1. Participants

Forty-five participants took part in this study (26 female and 19 male), with 36 of the participants (80%) aged between 20-39 years old. All participants had normal or corrected-to-normal vision, normal or corrected-to-normal hearing, and no known illness that could affect tactile perception. Candidates with diabetes were excluded because of a potential influence on haptic perception (Travieso and Lederman, 2007). Driving experience or a driver's license was not required because a self-driving car scenario was employed for this study.

The ethical review process was conducted and approved by the University of Warwick's Biomedical and Scientific Research Ethics Committee (REGO-2016-1741).

2.2. Design and procedure

The study used a 3 (warning type) x3 (task type) within-subjects factorial design with repeated measures. Dependent variables were reaction time (RT), and four subjective ratings of the warning in each scenario. The first factor was the warning type and the second factor the task type. The three warnings were a multimodal (auditory-tactile) warning, an auditory warning (from a commercially available car), and a tactile warning. In each driving scenario, only one type of warning cue (either audio, tactile or multimodal) was presented to the driver, however each warning was presented eight times during a single scenario, with the warnings appearing randomly, but separated by at least 10 seconds. The three set paced, attention capturing tasks ('Task' in Table 1) were presented in modalities which covered typical interface modalities and a range of tasks drivers might engage: reading (visual), listening to an audiobook (auditory), or physical interaction with a mobile phone (tactile). All tasks were serial presentation tasks which imposed a continuous level of demand (see section 2.3). For each factor combination (three warnings and three tasks), a slightly different scenario was designed, resulting in a total of nine self-driving, simulated scenarios (Table 1).

Table 1. Study scenarios.

Scenario	A	B	C	D	E	F	G	H	I
Task	Visual distraction	Auditory distraction	Tactile distraction	Visual distraction	Auditory distraction	Tactile distraction	Visual distraction	Auditory distraction	Tactile distraction
Warning	Multimodal	Multimodal	Multimodal	Audio	Audio	Audio	Tactile	Tactile	Tactile

The participants started the study with an introduction, then completion of the demographic questionnaire. The participants made themselves comfortable in the driving seat of the driving simulator (Figure 1), after which a training scenario enabled participants to practice each of the three tasks and experience the three warnings. After the training scenario, the study started and the scenarios (Table 1) were presented in a counterbalanced order for each participant. In addition to the subjective data collected (Table 2) at the end of each scenario, the objective parameter of reaction time (time taken to press the brake pedal after warning was displayed) performance was also measured. To remove the variability of brake pedal initiation latencies, participants were encouraged to rest their foot on the brake pedal during the course of a scenario. Each warning type was presented eight times per scenario. Each scenario lasted approximately 5 minutes, after which the participants rated the level of workload and their experience of the warning in the scenario (Table 2).

Figure 1. Study procedure.

2.3. Design of the attention capturing tasks

The three tasks were selected to be similar in nature, to proscribe a pace and deliver a continuous attentional demand which lowered the interruptibility of the task. All tasks were rapid serial presentation tasks (RSPT). The participant was presented with either visual or auditory stimuli in rapidly changing serial cues, e.g. a rapid series of numbers and letters, appearing for a predefined timeframe. After exceedance of this timeframe, the cue disappeared and no cue was presented for a predefined timeframe (between 80 and 350 ms, dependent on the task). Thereafter, the next cue appeared. When a target was identified, (in this study a number appeared within a rapidly scrolling series of letters) the participants should react by tapping on the screen of the recording device (a Microsoft Surface Pro 4 tablet PC, with 60 Hz screen refresh rate). This task can easily be transferred to the auditory and visual modality, and has been previously successfully employed in a variety of research projects (see Soto-Faraco and Spence, 2002; Ho, Tan and Spence, 2005). However, a significant contribution of this paper is to extend the previously utilised visual and auditory serial presentation tasks, into the tactile modality.

The visual task was adapted from Ho and Spence (2005). A stream of random letters appeared on the display of a tablet PC (Figure 2). Within the stream of letters, numbers appeared at random points of time. Whenever a number appeared on-screen the participant was required to tap on the screen as fast as possible. The auditory task was adapted from Soto-Faraco and Spence (2002), and used the same letters and numbers as the visual task. The task was similar to the visual one; the participants held a tablet, an audio stream of letters was played with a randomly occurring number as a target. Participants again responded by tapping on the screen as quickly as possible when they detected the number.

Figure 2. The visual task on a tablet PC (left), and the equipment for the tactile task (right).

The tactile task was designed to mimic the visual and auditory task. Due nature of presenting information in a tactile modality, it was not possible to present letters and numbers in this task. The tactile task needed to be a change detection from the previous stimuli without requiring intense learning. For the study, two motors (Lilypads) were connected to an Arduino system, and thence to a tablet computer (Figure 2). The participants placed the tablet on their lap, and held one of the motors in the left and the other in the right hand for the duration of the task. The following possible stimuli were presented in tactile task: left motor vibrating; right motor vibrating; both motors vibrating; no motor vibrating. Participants were required to detect when both motors vibrated at the same time, and then tap on the tablet to register a response.

In a pilot study, three settings in each of the three tasks (visual, auditory and tactile) were compared to identify settings across the task types which have a comparable level of demand (measured by a rating of workload). A participant experienced each of the nine task-settings once, for 2 minutes. After experiencing a task setting, the participant rated the experienced level of workload on a scale from zero (very low) to one hundred (very high) (Hill et al., 1992). The workload ratings were averaged for all participants for each task setting to determine the mean workload a setting imposed on the participants. The selected settings and their associated workload ratings were the following:

i. Visual task: workload *M* = 76; setting: 8 targets, signal appeared for 40 ms, inter-stimulus intervals 80 ms
ii. Auditory task: workload *M* = 74.5; settings: 8 targets, signal appeared for 120 ms, inter-stimulus intervals 150 ms
iii. Tactile task: workload *M* = 65; signal appeared for 40 ms, inter-stimulus intervals 350 ms

2.4. Design of the warnings

A pilot study investigated warnings in auditory-tactile, auditory, and tactile modalities as these do not require visual attention. Auditory and tactile modalities were selected as they have the advantage of being independent on where the users' glances are located.

Two unimodal warnings were investigated in this study: an auditory warning from commercially available cars and a tactile warning. The auditory warning was a 2 s beep presented in frequencies between 94-8000 Hz, pulsing seven times, presented over two loudspeakers left and right to the monitors. Research by Lees and Lee (2007) suggest that warnings should be presented at 10-15 decibel (dB) above the surrounding noise level. Consequently, the auditory warning was presented at 70 dB for this study. The tactile warning was constructed by taking the auditory warning and transforming this into low vibrational frequencies that were 'hapticly' perceptible through the seat of the development simulator (Figure 3). The ButtKicker Gamer 2 (with power amplifier BKA-130-C, providing a power output of 90 watts at 2 ohms) is a rotating motor on the back of the driver seat which converts audio signals (primarily bass sounds) into a low frequency vibration. The ButtKicker's transducer is fitted to the lower rear of the driver seat and vibrated the whole driver's seat as a tactile warning. Auditory and tactile warning were evaluated in a pilot study beforehand to be perceived as equally intense to avoid confounding the RT by a more intense warning.

The auditory and tactile unimodal warnings, as described above, were combined to form the multimodal warning. Two stimuli presented concurrently will have the strongest association the more characteristics they share (Spence and Ho, 2017). Hence, the auditory and tactile components of the multimodal warning were presented simultaneously and in parallel, both shared a similar pattern, but as the warning was a general, non-directional specific one there was limited need for it to be spatially linked (c.f Wilson, Reed and Braide, 2009; Spence and Ho, 2017).

2.5. Apparatus

The study was conducted in WMG's 3xD Development Simulator, consisting of a racing car seat on a metal frame, a racing steering wheel, gearbox, pedals and three monitors on which the virtual driving scenario was presented (Figure 3). The scenarios showed a car driving through a cross-country road and a small village. Medium density of traffic appeared in the opposite lane, flow varied, but with the same number of cars were present in all scenarios. The participant's car drove in autonomous mode with one car following and two lead cars in front (i.e. the third car in a convoy of four). The vehicles drove at 50 miles per hour (mph) on straight sections of roadway, and at approximately 25 mph around bends. Two seconds after a warning onset, the lead car started braking.

After each driving scenario, the warnings were rated in a questionnaire. The participants rated the noticeability, motivation to respond, annoyance, and how startling the warning was in the scenario (Table 2). The overall workload scale (Hill et al., 1992) was adapted to let the participant rate the perceived level of workload between 0 (very low) and 20 (very high). Campbell, Richard, Brown, and McCallum (2007) suggest that warnings should not be annoying or startling, though, they must be noticeable. Furthermore, Wogalter, Conzola, and Smith-Jackson (2002) suggest that motivation plays a role in response to the warning and this can, besides past experiences with the system, be influenced by the characteristics of the warning. Hence these additional criteria were also selected to be subjective rated in this study. A similar questionnaire with a seven-point rating scale was used by the authors in a previously published study (Geitner, Birrell, Krehl, and Jennings, 2018) and by (Brown, 2005).

Figure 3. The WMG 3xD Development Simulator.

Table 2. Questionnaire about the subjective perception of the warning cues

1) How would you rate the workload in the last scenario?

Very low																			Very high
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

2) How clearly was the warning cue noticeable?

Not very much						Very much
☐	☐	☐	☐	☐	☐	☐

3) How much did warning cue motivate you to respond?

Not very much						Very much
☐	☐	☐	☐	☐	☐	☐

4) How startling was the warning cue?

Not very much						Very much
☐	☐	☐	☐	☐	☐	☐

5) How annoying was the warning cue?

Not very much						Very much
☐	☐	☐	☐	☐	☐	☐

2.6. Data analysis

The study had a three (warning cue) by three (distractor tasks) factorial design. Data analysis was conducted in R (R Core Team, 2014). The main objective dependent variable was RT, with other subjective dependent variables were the ratings (Table 2) given to each warning after every scenario.

We define RT as the time passed from onset of the warning to the initiation of the brake pedal. The RT data was analysed for outliers and missed warnings, excluding RTs longer than 2.5s or shorter than 0.4s from the analysis. Overall, 27 values (less than 1% of the data) were discarded. Missing values were replaced by the mean RT value for the participant's other existing RTs in this scenario. The dataset met the criteria of Sphericity, tested with a Mauchly's test (Task: W=0.98, p=0.6; Cue: W=0.93, p=0.21; Task-cue: W=0.85, p=0.6). Following this a repeated measure ANOVA analysis was conducted, with the RT being the dependent variable, and cue type and task type the independent variables. Paired t-tests were applied as post-hoc comparisons (with Bonferroni correction applied).

The rating data was tested for normality with the Shapiro-Wilk test (Noticeability: $p<0.001$, Motivation: $p<0.001$; Annoyance: $p<0.001$; Startling: $p<0.001$). The ratings for noticeability, motivation, annoyance and startling were not normally distributed. A paired Wilcoxon signed rank test as a non-parametric statistic was then applied for a within-subject variable comparison of the ratings across the three warning cues.

3. Results

3.1. Reaction time (RT)

An initial three (warning cue) by three (distractor task) repeated measure ANOVA was conducted to evaluate main effects of task and warning type. The ANOVA revealed a significant main effect of task (F(2, 88), p<.001, generalised $\eta^2 = 0.81$), and a main effect of warning (F(2, 88), p<.001, generalised $\eta^2 = 0.03$), and an interaction effect between task and warning (F(4, 176), p<.001, generalised $\eta^2 = 0.01$). Given the presence of a main effect, RTs were then compared across the three warning types separately in each of the three task conditions, and presented as a box plot in Figure 4. The box plot used in Figures 4, 5 and 6 presents the median value as the middle line of the box, the first and third quartile by the lower and upper hinge, whiskers that extend from the hinges to values no larger than 1.5 times the Interquartile Range (IQR), with values that are out of that range as presented dots and represent outliers of the data.

RTs in the tactile task condition are shown in Figure 4. Participants reacted significantly faster to the multimodal (M = 1.28 s) and auditory warnings (M = 1.27 s) compared to the tactile warning (M = 1.33 s), p<0.005 and $p<0.001$ respectively. In the tactile task condition, the RT to the multimodal warning ($M = 1.28$ s) and to the auditory warning ($M = 1.27$ s) did not differ significantly, $p>0.05$.

RTs in the auditory task condition are shown in Figure 4 as overview. In the auditory task condition, RTs did not differ significantly across the three warning cues (multimodal: $M = 0.75$ s, SE = 0.008; auditory: $M = 0.78$ s, SE = 0.009; tactile: $M = 0.78$ s, SE = 0.01).

RTs in the visual task condition are shown in Figure 4 as overview. In the visual task condition, RTs were faster for the multimodal warning ($M = 1.25$ s) compared to the tactile warning ($M = 1.3$ s), $p<0.001$, and for the auditory warning ($M = 1.23$ s) compared to the tactile warning, $p<0.001$. RTs between multimodal warning and auditory warning did not differ significantly, $p>0.05$.

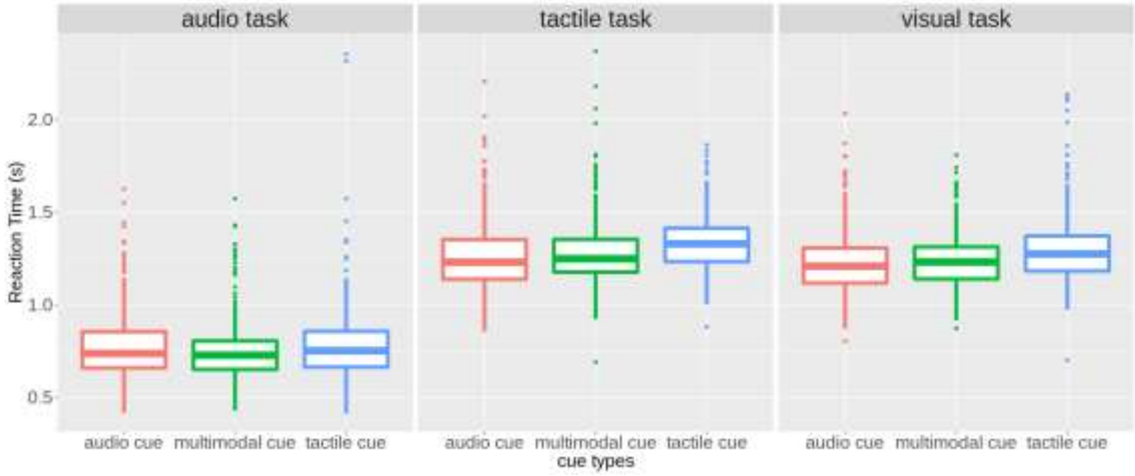

Figure 4. Reaction time to each warning cue (audio, tactile and multimodal) presented for all three set paced, attention capturing tasks conditions (audio, tactile and visual).

Overall, the RTs to all warning cues were shorter in the auditory task condition compared to the other two task conditions (Figure 4). This might be related to the nature of the task or task setting. An auditory task might be easier to combine with partial monitoring of the road and reacting to a warning. Critically, the auditory task involved the lowest percentage of correct target detections compared to the other tasks ranging between 25-30% of the targets (target detection rates: mean 29.4% (auditory task), mean 84.3% (visual task), and mean 45.6% (tactile task)). However, the low correct detection could also be a result of the task difficulty (high workload). To investigate the engagement of participants in the task in more detail, all interactions (taps) on the tablet were analysed over the course of the task. Those taps comprised correct responses to targets and false interactions, were separated into sub-sections of 30 seconds for analysis. Whenever at least one tap on the tablet occurred in such a 30 seconds interval, one was added to the engagement counter. The engagement in the auditory task was generally high, with only two incidences of no taps with a 30 seconds period recorded by participants.

In total 19 warnings were missed. The most warnings were missed in the auditory warning conditions (11 misses), with less in the multimodal (4 misses) and in the tactile (4 misses) conditions. The most false reactions (i.e. brake pedal being pressed when no warning was active) occurred in conditions with the tactile warning (43 false reactions) compared to the multimodal (20 false reactions) and to the auditory warning (19 false reactions).

3.2. Subjective perception

Overall, the three warnings were rated significantly different for noticeability (
Figure **5**). A Wilcoxon signed rank test revealed that the multimodal warning (*M* = 6.33) was rated as significantly more noticeable compared to the auditory (*M* = 5.96), V = 1646.5, *p*<.001), and the tactile warning (*M* = 5.56), V = 2499.5, *p*<0.001. The auditory warning was rated as significantly more noticeable compared to the tactile warning, V = 2237, *p*<0.001.

The multimodal warning (*M* = 6.24) was rated as significantly more motivating to respond than the auditory warning (*M* = 5.87), V = 2068.5, *p*<0.001, and the tactile warning (*M* = 5.65), V = 2437.5, *p*<0.001. The motivation to respond was not significantly different between auditory and tactile warnings (
Figure **5**).

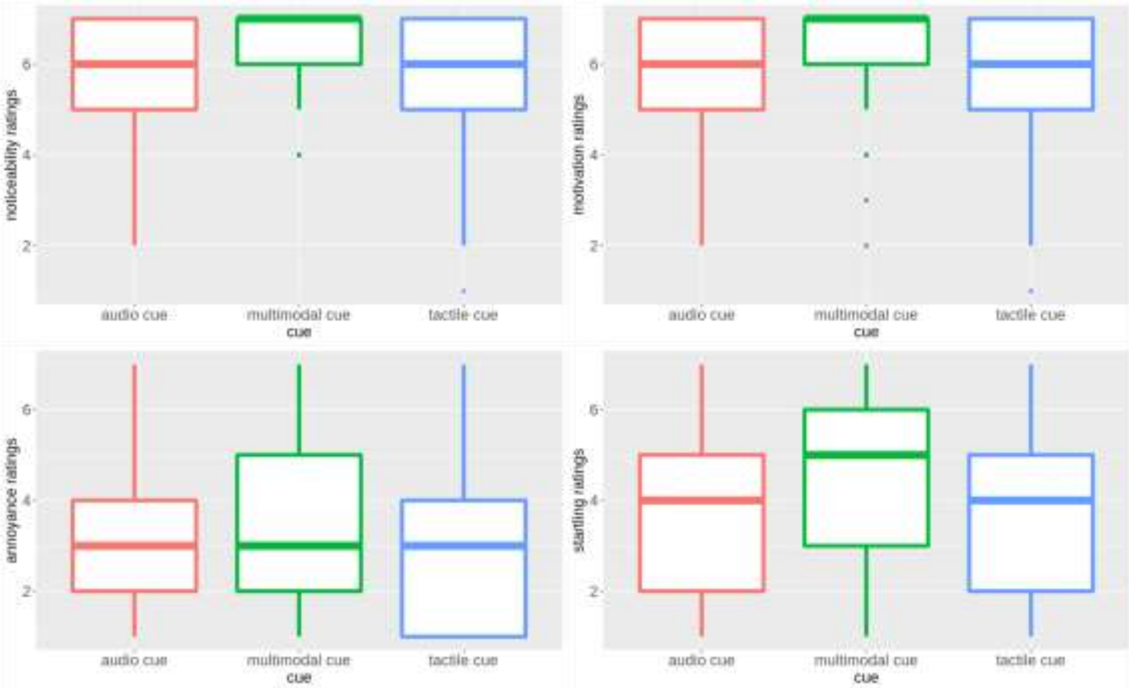

Figure 5. Subjective ratings (ranked between 1 and 7, as per Table 2) for each of the three cues: noticeability (top left), motivation (top right), startling (aka. startlement, bottom left), and annoyance (bottom right).

The multimodal warning (*M* = 4.46) was rated as being significantly more startling than the auditory warning (*M* = 4.07), V = 2545, *p*<0.001, or the tactile warning (*M*=3.95), V = 3298.5,

p<0.001. There was no significant different in the startling ratings between auditory and tactile warnings, p>0.05 (

Figure **5**).

There were no significant differences between the multimodal warnings (M = 3.25) and the auditory warnings (M = 3.25) when rated for annoyance (Figure 5). Both the multimodal warning (V = 2205.5, p=.015) and the auditory warning (V = 2157.5, p=.016) were rated as significantly more annoying compared to the tactile warning (M = 2.99).

The ratings for noticeability and startlement were similar across the task conditions for the multimodal warning, auditory warning and the tactile warning (

Figure **6**).

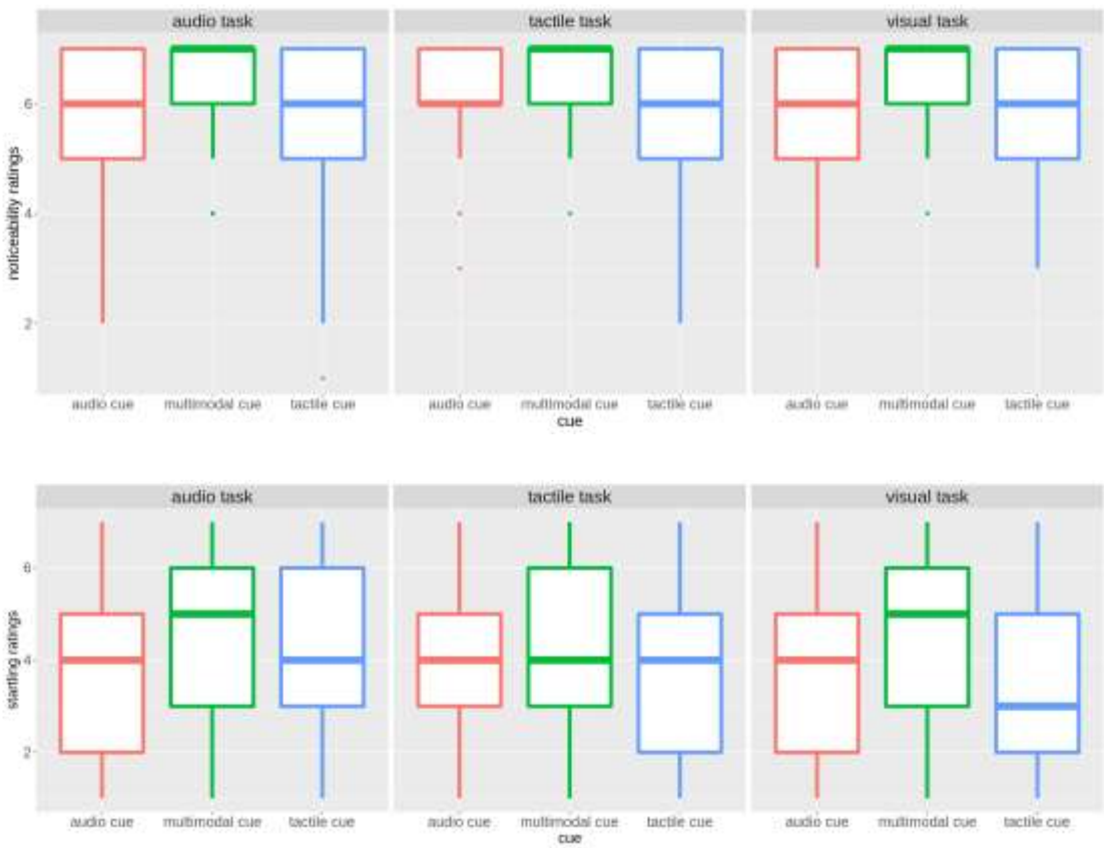

Figure 6. Subjective ratings (ranked between 1 and 7, as per Table 2) for each of the three cues: noticeability (top), and startling (aka. startlement, bottom).

4. Discussion and conclusion

In our autonomous driving scenario with a highly attention-capturing series of set pace tasks with similar workload, the auditory-tactile warning and the industry used auditory warning conditions showed similar reaction times (RT) – with both being significantly faster than the tactile only warning. Answering the first research question, the auditory-tactile warning resulted in fewer missed warnings and fewer false reactions. The reduction in false reactions in the multimodal condition is similar to Blanco et al. (2015) for a visual-haptic alert; where drivers gave incorrect responses during a cautionary take-over alert, 83.3% of these were in the unimodal (visual) condition, compared to just 16.7% in the multimodal condition. Within the automated driving context, and independent to warning type or modality, providing the driver with knowledge of the attributes and limitations of a low capability automated system led to an increase in false responses, as the drivers' were overly cautious about when the system might fail (Khastgir et al., 2018). The use of multimodal warnings, as shown in this paper, might support the driver with this uncertainty.

Participants reacted faster to multimodal and auditory only warnings compared to the tactile-only warning in the visual task condition. Although this cannot be predicted by the MRT it has been found in laboratory settings that the perception of visual stimuli can decrease the perception of tactile stimuli that are presented in parallel (Auvray et al., 2008; Murphy and Dalton, 2016). Auvray et al. (2008) evaluated how people detect changes between two patterns of tactile pulses presented on the finger. When participants wore an eye mask to eliminate the visual channel between being presented the two tactile patterns, they had more difficulties in detecting the change. Murphy and Dalton (2016) conducted a visual task, with high and low difficulties, responding to the presence or absence of a tactile stimulus in parallel. They reported that detection accuracy decreased under highly difficult visual tasks, compared with the low visual task. Both studies indicate that a visual task can decrease tactile perception – an effect also observed in this current study.

The multimodal warning was associated with a faster RT than the tactile warning in the tactile task condition, and also RTs to the auditory warnings were faster than to the tactile warning. This results are complementary to the MRT (Wickens, 2002), where different sensory channels are assumed to be separate resources. Both the tactile task and tactile only warning, utilised the same sensory channel, which increased the RT to the tactile warning. The auditory component of the multimodal warning and the auditory only warning utilised an alternative sensory channel, not occupied by the task, and hence the performance benefits were observed.

A similar effect of task interference would have been expected between auditory warning and auditory task, because warning and task utilise the same modality. However, in this condition there was no significant difference in RT between the three warnings. This reason could be due to a speed accuracy trade-off, as the auditory task conditions resulted in the shortest RTs towards the warning compared to the other two task conditions, but lowest detection rate of the targets in the distracter task. Another explanation might be that participants disengaged from the auditory task, and therefore were able to react faster to the auditory warning than would be expected. Participants gave higher than average numbers of false responses in the auditory task condition which indicates that although they remained physically engaged, even if these response were incorrect, they may also have been less cognitively engaged when responding. The lower rate of correctly detected targets may have led to less interest and consequently less cognitive engagement in the auditory task. This has been suggested in previous research which reported that brake responses can be longer when completing in-vehicle tasks that are interesting for drivers (Horrey et al., 2017).

In answer to the second research question, the multimodal warning was rated as more noticeable and more motivating to respond than the auditory and tactile only warnings. Pjetermeijer et al. (2017) reported auditory-tactile warnings were rated as most effective compared to auditory and to tactile only warnings. However, in their study, this increase in effectiveness was combined with a rating of being more startling. In this current study auditory-only and auditory-tactile warning were also rated more startling compared to the tactile-only warning. Positively, similarly to previous research, our results suggest that the multimodal warning was not perceived as being more annoying compared to the unimodal traditional auditory warning (Biondi et al., 2016). Similarly, in Pjetermeijer et al. (2017) tactile warnings were rated as least annoying compared to auditory and auditory-tactile warnings. However, annoyance might be less important when we consider a safety related warning, as its main purpose is alerting the driver.

A potential limitation of this study, is that the RT did not include the movement of the foot to the pedal. The participants kept their foot on the pedal for the duration of the scenario, which is not a common driving behaviour. In a more realistic setting RTs are expected to increase by an additional time that is required to move the foot to the brake pedal. In addition, within real driving scenarios the time to press the brake pedal might not be the optimal performance measure, as steering around or accelerating through an obstruction might be a more suitable response. Future work should consider a variety of driver interventions in order to eliminate the response bias, and ensure perceptual (or performance) enhancements for multimodal warnings are still observed. The sample selected for this study included inexperienced, or unregistered drivers, as interaction with the control of the vehicle (barring pressing the brake pedal) was not part of this study. Whenever a study would investigate the perception of the road scenery or interaction with other road users, participants without driving license would need to be excluded.

4.1. Implications for research about distraction in automated vehicles

Previous research has shown that automation can lead to an attentional underload situation in which drivers are tempted to engage in non-driving related tasks, which may result in these tasks actually becoming the primary task (i.e., self-regulation; Strayer and Fisher, 2015). In this study we utilised three highly attention-capturing, set-paced tasks as primary tasks to compare the effectiveness of warnings in different modalities. The benefits of these ‘artificial’ tasks, as utilised in this study, were that they could simulate a similar level of demand over various sensory modalities, and were a first step towards an evaluation of distraction specifically in an automated vehicle setting. In addition, each warning types was presented a total of eight times, but randomly distributed throughout each driving scenario. Meaning that the participants were expecting multiple warnings to be presented – which is a key difference to real driving. Within the current study, the self-driving scenario (and indirectly the driving simulator) was there to provide additional context to the study, making it an enhanced, or hybrid, lab study rather than an evaluation of real driving. This is where self-driving or automated driving scenarios – either completed in high fidelity, driver-in-the-loop simulators, or conducted in the real-world – could utilise more tasks from the rich setting of perceptual studies. This approach is supported by Spence and Ho (2015) who advocate that fundamental, lab-based research provides relevant ideas for the design of warning signals and information system.

More research is needed to understand what tasks drivers would conduct in an automated vehicle, to understand the spectrum and effects of those tasks. It is particularly important how drivers switch between tasks as this can influence the time required for a take-over request, as specific goals can extend the transition of attention from one task to another (Lee, 2017). The tasks in this study were a continuous input stream that did not have any option for interruption, beyond simply disengaging from the task itself. Research into distraction in an automated vehicle scenario should consider ‘interruptability’ of tasks. Additionally, naturalistic tasks may involve unforeseen components with a safety-critical impact which can only be understood by observing behaviour. For example, some tasks might require the shutting down or putting away of a device, before the action to the take-over signal can be initiated, and so require additional time before the task can be interrupted – time that is not accounted for in an artificial task setting. A simple example of this is drinking a cup of coffee, where there is a need to put it away before performing a response to a warning in an automated vehicle (Banks, Eriksson, O’Donoghue and Stanton, 2018). Future studies may investigate behaviour in an automated vehicle with artificial and naturalistic tasks.

4.2. Implications for warnings automated vehicles

This study showed that the auditory-tactile warnings were as effective in terms of RT as in-vehicle auditory warnings over the course of the three set pace tasks with a similar level of workload. Compared to unimodal warnings, the multimodal warning resulted in a reduced number of missed and false responses. In future research, this advantage of multimodality on missed and false reactions to warnings could be explored in more detail, over a variation in a highly automated driving context with tasks that reflect the worst cases of user behaviour, e.g., hard to interrupt and highly attention-capturing, such as the set pace tasks in our study. Specifically, a low percentage of missed alarms would shorten the time required for an emergency brake or the take-over manoeuvre. Besides speed and effectiveness, quality of response is equally as important (Gold, Dambök, Lorenz, and Bengler, 2013; Radlmayer et al. 2014). In this scenario we utilised an emergency brake event; however, a take-over of control scenario utilises a wider range of variables, such as steering, speed control and an extended understanding of the surrounding traffic.

An auditory-tactile warning, a tactile warning and an auditory warning were compared in this study. A comparison to a visual-tactile warning could be interesting in a future study. Previously it has been shown that a visual-tactile take-over warning is beneficial over a visual warning (Naujoks et al., 2014; Blanco et al., 2015). Compared to an auditory-tactile warning, a visual multimodal warning could contain more information – for example, information regarding other nearby vehicles. However, it is not yet shown how such more detailed information would be traded of in quality and speed of the take-over manoeuvre.

This study primarily focused on the noticeability of the three warnings, which has been shown to be just the first stage of the human-machine interaction process (Norman, 2002). Other stages such as conveying a meaning and adequateness of response were not considered, but would be relevant for evaluation for future design of warnings. In order to be able to generalise results from this study to the effectiveness of the warnings, they were presented more frequently than in a real driving scenario, which can make participants more vigilant to respond to such warnings. Future studies should consider, in combination with the evaluation of meaning and adequateness, presenting the warnings to drivers a limited number of times in each driving scenario.

Highlights

- Multimodal and auditory warnings had faster reaction times to tactile only warnings
- Multimodal warning resulted in fewer missed and false responses in all task conditions
- Multimodal warning rated as more noticeable and motivating to respond over unimodal
- Three set paced attention capturing tasks of similar workload were developed and used

Acknowledgements

The authors would like to thank all volunteers for participating in the study. Support for this work was provided by Engineering and Physical Sciences Research Council (EPSRC), Jaguar Land Rover and WMG centre HVM Catapult.

References

Auvray, M., Gallace, A., Hartcher-O'Brien, J., Tan, H., and Spence, C., 2008. Tactile and visual distractors induce change blindness for tactile stimuli presented on the fingertips. *Brain Research*, 2008(5), 111-119.

Banks, V., Eriksson, A., O'Donoghue, J., and Stanton, N., 2018. Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, 68, 138-145.

Biondi, F., Leo, M., Gastaldi, M., Rossi, R., and Mulatti, C., 2016. How to drive drivers nuts: Effects of auditory, vibrotactile, and multimodal warnings on perceived urgency, annoyance and acceptability. *Transportation Research Board 96th Annual Meeting*, February 2017.

Biondi, F., Strayer, D., Rossi, R., Gastaldi, M., and Mulatti, C., 2017. Advanced driver assistance systems: Using multimodal redundant warnings to enhance road safety. *Applied Ergonomics*, 58, 238-244.

Birrell, S., Young, M., and Weldon, A., 2013. Vibrotactile pedals: provision of haptic feedback to support economical driving. *Ergonomics*, 56(2), 282-292.

Blanco, M., Atwood, J., Vasquez, H. M., Trimble, T. E., Fitchett, V. L., Radlbeck, J., and Morgan, J. F., 2015, *Human factors evaluation of level 2 and level 3 automated driving concepts*. (Report No. DOT HS 812 182). Washington, DC: National Highway Traffic Safety Administration.

Brown, S., 2005. *Effects of haptic and auditory warnings on driver intersection behavior and perception* (MSc Thesis). Virginia Polytechnic Institute and State University, USA. http://vtechworks.lib.vt.edu/handle/10919/31844 (Accessed 05 December 2017)

Campbell, J., Richard, C., Brown, J., and McCallum, M., 2000. *Crash warning system interfaces: Human factors insights and lessons learned*. Final report for the National Highway Traffic Safety Administration (NHTSA): DOT HS 810 697.

Carsten, O., Lai, F., Barnard, Y., Jamson, A. H., and Merat, N., 2012. Control task substitution in semi-automated driving: does it matter what aspects are automated? *Human Factors*, 54, 747-761.

Desmond, P. A. and Hancock, P. A., 2001. Active and passive fatigue states. *In: P. A. Hancock and P. A. Desmond, Eds.; Stress, Workload, and Fatigue*, 455–465. Mahwah, NJ: Lawrence Erlbaum.

Eriksson, A. and Stanton, N. A., 2017. Take-over time in highly automated vehicles: non-critical transitions to and from manual control. *Human Factors,* 59(4), 689–705.

Geitner, C., Birrell, S., Krehl, C., and Jennings, P. (2018). Haptic foot pedal: influence of shoe type, age, and gender on subjective pulse perception. *Human Factors*, 60(4), 496-509.

Gold, C., Dambök, D., Lorenz, L., and Bengler, K., 2013. "Take over!" How long does it take to get the driver back into the loop?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 57(1), 1938–1942.

Hill, S. G., Lavecchia, H. P., Bittner, A. C., Zakled, A. L., and Christ, R. E., 1992. Comparison of four subjective workload rating scales. *Human Factors - Special Issue: Measurement in Human Factors Archive*, 34(4), 429-439.

Ho, C., Reed, N., and Spence, C., 2007. Multisensory in-car warning signals for collision avoidance. *Human Factors*, 49(6), 1107-1114.

Ho, C. and Spence, C. (2005) Assessing the effectiveness of various auditory cues in capturing a driver's visual attention, Journal of Experimental Psychology: Applied, 11, 157–174.

Ho, C., Tan, H., and Spence, C., 2005. Using spatial cues to direct visual attention in driving scenes. *Transportation Research Part F*, 8, 397–412.

Horrey, W., Lesch, M., Garabet, A., Simmons, L., and Maikala, R., 2017. Distraction and task engagement: How interesting and boring information impact driving performance and subjective and physiological responses. *Applied Ergonomics*, 58, 342-348.

Khastgir S, Birrell S, Dhadyalla G and Jennings P. 2018. Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. Transportation Research Part C: Emerging Technologies, 96, 290-303.

King, A., and Calvert G., 2001. Multisensory integration: perceptual grouping by eye and ear. *Current Biology*, 11, 322–325.

Körber, M., Cingel, A., Zimmermann, M., and Bengler, K., 2015. Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, 3, 2403-2409.

Lee, J. D., 2014. Dynamics of driver distraction: The process of engaging and disengaging. *Annals of Advances in Automotive Medicine*, 58, 24–32.

Lees, M. and Lee, J., 2007. The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics*, 50(8), 1264-1286.

Llaneras, R., Sallinger, J., and Green, C., 2013. Human factors issues associated with limited ability autonomous driving systems: Driver's allocation of visual attention to the forward roadway. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design*, 92-98.

Merat, N., Jamson, A. H., and Carsten, O., 2012. Highly automated driving, secondary task performance, and driver state. *Human Factors*, 54(5), 762-771.

Murphy, S. and Dalton, P., 2016. Out of touch? Visual load induces inattentional numbness. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 761-765.

Naujoks, F., Mai, C., and Neukum, A., 2014. The effect of urgency of take-over requests during highly automated driving under distraction conditions. In *5th International Conference on Applied Human Factors and Ergonomics AHFE*, Krakow.

Naujoks, F., Purucker, C., and Neukum, A., 2016. Secondary task engagement and vehicle automation – Comparing the effects of different automation levels in an on-road experiment. *Transportation Research Part F,* 38, 67-82.

Norman, D., 2002. *The design of everyday things*. Basic Books, Inc. New York, NY, USA, 2002.

Pjetermeijer, S., Bazilinskyy, P., Bengler, K., and de Winter, J., 2017. Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics*, 62, 204-215.

R Core Team, 2014. R*: A language and environment for statistical computing*. Vienna, Austria. http://www.R-project.org/ (Accessed 04th December 2017)

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., and Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 2063-2067.

SAE, 2016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *J3016_201609*: SAE International.

Soto-Faraco, S. and Spence, C., 2002. Modality-specific auditory and visual temporal processing deficits. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55(A), 23-40.

Spence, C. and Driver, J., 1997. Cross-modal links in attention between audition, vision, and touch: Implications for interface design. *International Journal of Cognitive Ergonomics*, 1, 351-373.

Spence, C., & Ho, C. (2015). Crossmodal attention: From the laboratory to the real world (and back again). In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), The handbook of attention (pp. 119-138). Cambridge, MA, US: MIT Press.

Spence, C. and Ho, C., 2017. *The multisensory driver: Implications for ergonomic car interface design*. Aldershot, England, CRC Press.

Strayer, D. L. and Fisher, D. L., 2015. SPIDER: A framework for understanding driver distraction. *Human Factors*, 58(1), 5-12.

Travieso, D. and Lederman, S., 2007. Assessing subclinical tactual deficits in the hand function of diabetic blind persons at risk for peripheral neuropathy. *Archives of Physical Medicine and Rehabilitation*, 88(12), 1662-1672.

Wickens, C., 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159-177.

Wilson, E., Reed, C., and Braide, L., 2009. Integration of auditory and vibrotactile stimuli: Effects of phase and stimulus-onset asynchrony. *Journal of the Acoustical Society of America*. 126, 1960–1974.

Wogalter, M., Conzola, V., and Smith-Jackson, T., 2002. Research-based guidelines for warning design and evaluation. *Applied Ergonomics*, 33, 219-230.

Woodward, M., Dixon, S., Hamilton, J., and Goodall, W., 2017. *Deloitte – Driving connectivity. Global automotive consumer study: Future of automotive technologies*. United Kingdom Insights, March 2017. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/manufacturing/deloitte-uk-driving-connectivity.pdf (Accessed 15th of December)

Young, M. and Stanton, N., 1997. Automotive automation: Investigating the impact on driver mental workload. *International Journal of Cognitive Ergonomics*, 1(4), 325-336.