

# Evaluating take-over requests with directional audio

---

Assessing the effect of ipsilateral and contralateral verbal stimuli on response times and visual behavior

Master of Science Thesis

For the degree of Master of Science in Biomechanical Design  
at Delft University of Technology

*Submitted by*

Jimmy Hu  
4246969

*Supervisors*

Dr. Ir. Joost de Winter  
Ir. Bastiaan Petermeijer

2 October, 2018

Faculty of Mechanical, Maritime and Materials Engineering  
(3mE) - Delft University of Technology

## **Preface**

This thesis is my final research paper for the master degree of biomechanical design on the TU Delft. Prior to this thesis, I have did a literature report on spatially located auditory feedback, which can be found in the appendix. The main objective of this thesis is to investigate the use of directional auditory feedback in take-over requests in the domain of autonomous driving. With this work I hope to add valuable knowledge in the field of automotive human factors, and make autonomous cars safer in the future.

I want to thank my supervisors Joost de Winter and Bastiaan Petermeijer for supervising me during this period and contributing to the level of which my thesis is currently. Both have always been available for me for any questions and feedback when I needed it. Also I want to thank David Abbink for his enthusiasm and the guidance in my earlier stages of this thesis. Finally I want to thank my friends and family who have helped and supported me during my graduation process and my study.

## **Abstract**

Taking over control from an automated vehicle may take a substantial amount of time if the driver is not engaged in the driving task. Take-over requests containing directional information of hazardous surrounding cars could aid the driver in taking over the vehicle faster. However, whether the directional information should be presented ipsilaterally or contralaterally is still inconclusive. In this study, 34 participants were presented with animated video clips of traffic situations on a three-lane road, ending with a near-collision in front after 1,3, or 6 seconds. In each video, one lane was free to maneuver to safely. Participants were instructed to make a safe lane-change by pressing the left or right arrow key. At the start of each video, participants were provided with verbal auditory feedback: (1) 'Go left/right' (ipsilateral), (2) 'Danger left/right' (contralateral), and (3) Non-directional beeps as a baseline. 80% of the trials provided valid auditory feedback (i.e., relevant to the video situation). 20% of the trials provided invalid auditory feedback (i.e., feedback opposite to the video situation, so left instead of right and vice versa). Auditory feedback 'Go/Danger left' was always presented from the left speaker, and 'Go/Danger right' was always presented from the right speaker, whereas the non-directional beeps were presented from both speakers. Participants' keyboard responses and first gazes were recorded in each trial. It was hypothesized that when there was 1-second to respond, ipsilateral feedback ('Go') would lead to fastest responses because little time is available to detect the hazard. For 3 and 6-second-to-respond situations, it was hypothesized that contralateral feedback ('Danger') would lead to a faster detection time of the hazard, because it facilitates visual detection. The results showed that for 1 and 3-second videos, ipsilateral feedback led to significantly faster responses compared to the baseline, and for 3 and 6-second videos contralateral feedback 'Danger' led to significantly faster responses compared to the baseline. 'Go' and 'Danger' did not yield a significant difference in response time between

each other in all videos. First fixations seem to be placed on the most salient visual stimuli, independent of the audio feedback. In 1-second time-to-respond videos this was the center of the road (the location where the potential collision is happening. In 3-seconds, this was the hazard coming from the left or right lane. And for 6 seconds the first fixations were more distributed. In conclusion, verbal auditory feedback 'Go' and 'Danger' can aid in taking over a vehicle, by reducing the take-over response time compared to baseline warnings. However, this may not be the result of facilitation of the visual detection of the hazard, because visual behavior seems to be influenced mainly by visual stimuli, independent of auditory stimuli.

## **1. Introduction**

### **1.1. Autonomous Driving**

Fully autonomous driving is seen as the future of driving, and car companies such as Tesla, Mercedes-Benz, BMW as well as non-car companies such as Google and Amazon are investing in automated driving. Automated driving has the potential to reduce human-related accidents, by keeping the human out of the loop. However, highly automated driving (HAD) is the current state-of-the-art, and according to experts, it could take decades before cars will be able to drive fully automated (Shladover, 2015). In HAD the car is capable of driving automated, but needs to be taken over by the human in case it encounters situations which cannot be solved by the automation. If the automation recognizes such a situation, it will provide the driver with a warning, also called a take-over request (TOR). Taking over the vehicle properly after a TOR can take time up to 7 seconds (Gold et al. 2013), because the driver is not engaged in driving (see review by De Winter et al. 2014) . Lu et al. (2018) also showed that the amount of time to regain reasonable situational awareness is around 7 seconds (Endsley 1988). In emergency situations there may not be sufficient time to regain



complete situational awareness and take control of the car, leading to potentially life-threatening situations. It is therefore very important to provide drivers with a TOR that drivers can interpret quickly, so they can take-over the car as fast as possible.

### 1.2. TOR Modality

A take-over request can be given in different modalities, for example by using vibrotactile feedback, visual feedback, auditory feedback or a combination of these (see Bazilinskyy et al. 2018). In this research is focused on the auditory modality. The auditory modality is practical as warning because it is gaze-free (Sanders and McCormick 1987), composed to visuals, where the human has to look at the specified location. Also the human has the capability to detect the location of which the sound is coming from (Terrence et al. 2005),(Populin 2008), which can be used to communicate locations. Finally, Auditory feedback has the advantage of being able to convey semantics to it, allowing information to be communicated. In take-over situations, communicating directional information, of for example, surrounding cars or steering direction may be useful in aiding the driver to take-over the car faster and more safely.

### 1.3. Directional TORs

Little research has been done on directional TORs, but this topic is gaining attention (Petermeijer & Bazilinskyy 2017; Cohen-Lazry & Katzman 2018). Directional warnings have previously been researched in ADAS such as collision warning systems (CWS) (Wang et al. 2007; Ho & Spence 2005) and lane-keeping assistance (LKA) (Rimini-Doering et al. 2005; Suzuki & Jansson 2002). Studies were performed to compare directional warnings by means of spatially located warnings to non-spatial warnings, to study the effect of directional warnings. For example, Ho & Spence (2005) performed a study where videos of potential collision were shown to the participants from the front or back. The participants were busy

doing a 2<sup>nd</sup>-task , and different audio feedback types were used: Car horn and verbal speech: “Front” and “Back”, both were tested spatially from the back and front and non-spatially. Participants had to respond by either pressing the brake pedal or the gas pedal, depending on the location of the danger. The results of this study showed that spatially located verbal speech led to the fastest reactions of all audio types, and concluded that spatially located audio can facilitate detection of threats. Also, Rosmeier et al. (2005) experimented with drowsy drivers in a driving simulator, and provided drivers with the sound of a wheel driving over a road mark from one direction to alert the driver when crossing a road mark. The time-gap between opening the eyes and steering response was measured. This duration was found below 0.19 seconds in some cases, which is too fast to be a visually driven action. They concluded that the effect of a warning on the steering response is not only to focus the attention on the driving scene, but can also evoke steering reactions. Some studies found no effect of spatial warnings compared to non-spatial warnings. For example in a driving-simulator study by Suzuki et al. (2002) participants were occupied with a secondary task, and small steering disturbances were given to the car. Participants were presented with mono auditory beeps and stereo auditory beeps if they departed from the lane. Stereo beeps were given from the departure side, whereas mono beeps were presented from both sides. The authors concluded that stereo directional information did not affect the response time, because drivers use their visual information to decide, and the audio is used just to grab the attention of the driver. In the case of HAD, Petermeijer et al. (2017) investigated multimodal and directional TORs, by letting drivers take over an automated car in an emergency situation where a stationary car was stranded in the middle lane. Spatially located beeps were presented from either the left or right side, and drivers were free to choose in which direction they avoided the obstacle. The results showed no effect of the spatially located warnings in uninstructed drivers. Most drivers were unaware of the directional audio, and took over the car

at the left side because of German traffic rules. However, directional TORs still have potential when more salient stimuli and higher levels of semantics are provided in the feedback. During TORs drivers are initially disengaged, so may not have the time to regain complete situational awareness. Directional TORs can help during these situations and therefore deserves to be studied thoroughly.

#### 1.4. Ipsilateral or Contralateral

One of the topics of discussion in CWS, LKAs and TORs is whether to provide the driver with feedback on the escape direction (i.e., safe lane or steering direction), also called ipsilateral feedback, or with feedback on the hazard direction (i.e., lane departure side or impending car), also called contralateral feedback. Literature of non-driving studies suggests that for laboratory stimulus-reaction tasks, ipsilateral feedback (or compatible feedback) leads to faster reactions (Fitts & Posner 1967), because the stimulus and response are at the same location. In a driving-related study where participants had to steer a steering wheel to the location of the sound (compatible) or away from the sound (incompatible), was also found that steering to the location of the sound (ipsilateral feedback) led to faster responses (Wang & Proctor 2003). However, Musseler et al. (2012) concluded that there is a reverse effect of compatibility in naturalistic driving scenes, meaning that in natural driving scenes drivers are naturally inclined to move away from stimuli instead of towards stimuli. Moreover, studies on cross-modal attention have shown faster visual detection time of objects using spatially located auditory feedback from the visual location (Driver & Spence 1998). In a driving simulator study, it was found that response times were shorter when the collision warnings were provided at the location of the danger (Wang et al. 2007). They concluded that, drivers use their visual impression to make a steering decision, and that the spatially located side collision warning at the side of the hazard, directs drivers' attention to the location of an

impending threat, before a driver would ordinarily perceive it, therefore decreasing response time. However this is not always the case; Straughn et al. (2009) tested a CWS in a simulator setting where drivers in foggy weather were provided with spatially located ipsilateral and contralateral collision warnings by means of a beep. Early warnings (2 seconds) and late warnings (4 seconds) were given and found that for late warnings, ipsilateral feedback led to a safer maneuver compared to contralateral. For early warnings, contralateral feedback led to a safer maneuver and faster reaction time compared to ipsilateral feedback. It is proposed that this effect occurs, because in late warnings there is not sufficient time to shift the attention to the location of the hazard, so the warning is utilized to generate an adequate response. On the other hand, when an warning is provided early, the driver can utilize it to detect the danger faster and respond accordingly.

From literature was found that contralateral feedback and ipsilateral feedback both can be beneficial in reducing response time. However, it is not certain whether this effect occurred because the directional audio effected the visual behavior and helped in detection of danger, or evoked a steering response directly. Also, it seems that the time-to-respond or the warning timing is important in the effectiveness ipsilateral and contralateral feedback.

### 1.5. Aim of the study

The aim of this study was to investigate drivers' response time and visual behavior when provided with directional auditory TOR in an emergency situation. Furthermore, this study aimed to compare ipsilateral and contralateral TORs, while varying the time-to-respond. The hypothesis was that ipsilateral audio would lead to the fastest responses when there was very little time to respond. When there is more time to respond, we expected no effect, because more time is available to scan the environment. Finally, another hypothesis was that the visual detection time is faster for contralateral TOR, because it can aid in the participants' hazard detection.

## 2. Method

### 2.1. Hardware

The EyeLink 1000 Plus eye-tracker of SR Research with head support was used to track the right eye. The videos were presented on a 24-inch BenQ XL2420T-B monitor with a resolution of  $1920 \times 1080$  pixels placed at a distance of 70 cm. Audio was provided using Logitech speakers at a distance of 70 cm and at a 45 degrees angle from the ears. Participants sat on a height adjustable office chair. In figure 1 a photo of the experimental setup is shown.



*Figure 1. Picture of the experimental setup, in this picture is shown: The chin/head support, the keyboard, the eye-tracker and the computer screen.*

### 2.2. Videos and audio

The videos were programmed using PreScan 7.0 (Tass International, 2015). The videos have been used previously in a study by Lu et al. (2018) in a study on situation awareness (see figure 2). Auditory verbal take-over requests 'Go left/right' (ipsilateral) and 'Danger left/right' (contralateral) were made using a speech generator (Naturalsoft Ltd. 2018), and post-processed using Garageband (Apple Inc., 2018) to equal the audio level peak and set start time to 50 ms. Audio level peaks were 66-68 dB, measured with a decibel meter from the head support, and the duration of 'Go left/right' and 'Danger left/right' was 600ms and 680ms respectively. More details about the audio clips can be found in appendix B. The audio beeps were based on a Tesla collision warning sound and made in Garageband using an Epiano

producing 4 beeps of 1148 Hz with 125 ms in between.

The audio clips were stereo, so spatially presented in the direction congruent to the directional instructions (i.e., left/right). In the audio type “Go” the direction of the audio is compatible with the direction of the response. That is, the participant is presented with the words “Go left” in the left speaker or “Go right” in the right speaker. Hence, the sound is spatially mapped in the ipsilateral direction. For the audio type “Danger”, the instruction direction is incompatible with the response, because the driver has to respond to the opposite side. That is, the participant is presented with the words “Danger left” in the left speaker or “Danger right” in the right speaker. Hence, the sound is spatially mapped in the contralateral direction.

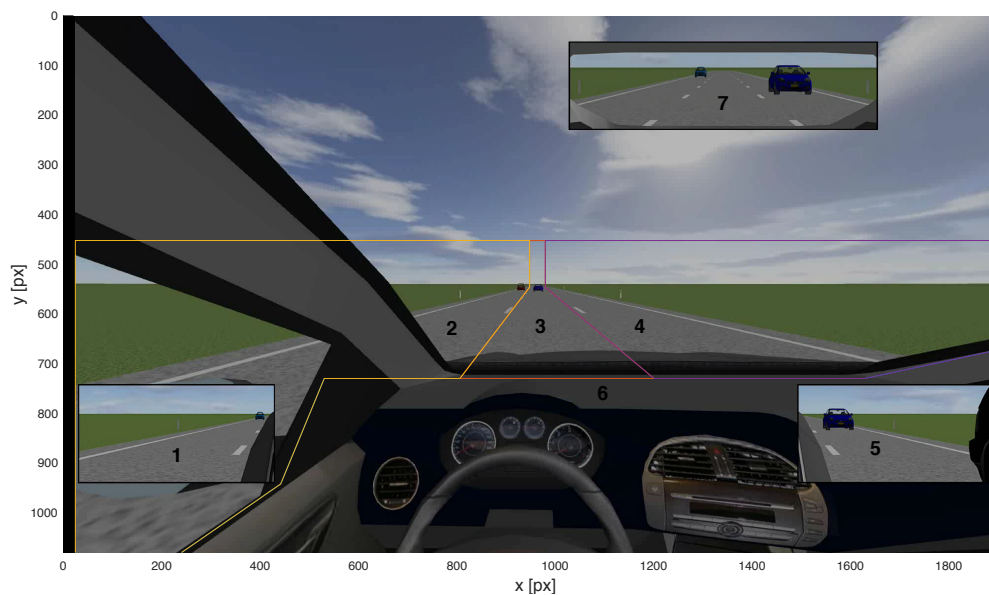


Figure 2. Example of driving scenario where the left lane is the escape direction and a video length of 1 second. Areas of interest indicated with numbers: 1. Left Mirror, 2. Left Road, 3. Centre Road, 4. Right Road, 5. Right Mirror, 6. Dashboard, 7. Centre Mirror

### 2.3. Participants

Thirty-four participants (29 males, 5 females), aged between 18 and 29 years ( $M = 23$ ,  $SD = 2.5$ ) participated in this study. Thirty participants had a driver license; 4 participants did not. On a scale of 1 (almost never), 2 (less than once a month), 3 (less than once a week), 4 (1–3

times a week), 5 (almost every day), the mean answer on “How much do you drive?” was  $M = 3.38$  ( $SD = 1.41$ ). All participants were recruited at the TU Delft from the faculty of 3mE and did not receive compensation for their participation. A within-subject design was used, so all participants did the same trials in a counterbalanced order.

#### 2.4. Video Situations

Participants viewed videos of traffic scenarios on a three-lane highway, on which the ego-car was driving in the middle lane with a constant speed of 28 m/s. In the videos, there were 5 surrounding cars, one of which was stationary in the middle lane, and one was overtaking the ego-car on either the left or right side. The other cars drove at a safe distance relative to the ego-car during the whole video. Videos had a length of 1, 3 or 6 seconds which was also the amount of time the driver had to respond. All videos ended with a 0.6-second time-to-collision with the stationary car in the middle lane. The overtaking car drove either on the left or right lane, and the other lane was free to safely maneuver to (see figure 3 for a example with a left escape direction). Each car was continuously visible, drove at a constant speed, and did not switch lanes during the whole video. For each video, three audio verbal take-over requests were provided: “Go left/right!” (ipsilateral), “Danger left/right!” (contralateral), and non-directional beeps (baseline). This yielded 18 different conditions: six different videos (L/R, 1/3/6s) and three different audio types (“Go”, “Danger”, beeps). Extra conditions were added with invalid auditory directions (i.e. verbal auditory “left” instead of “right” and vice versa), which means that the participants also had to rely on visuals, instead of solely the audio feedback. The audio beep is non-directional and therefore always valid.

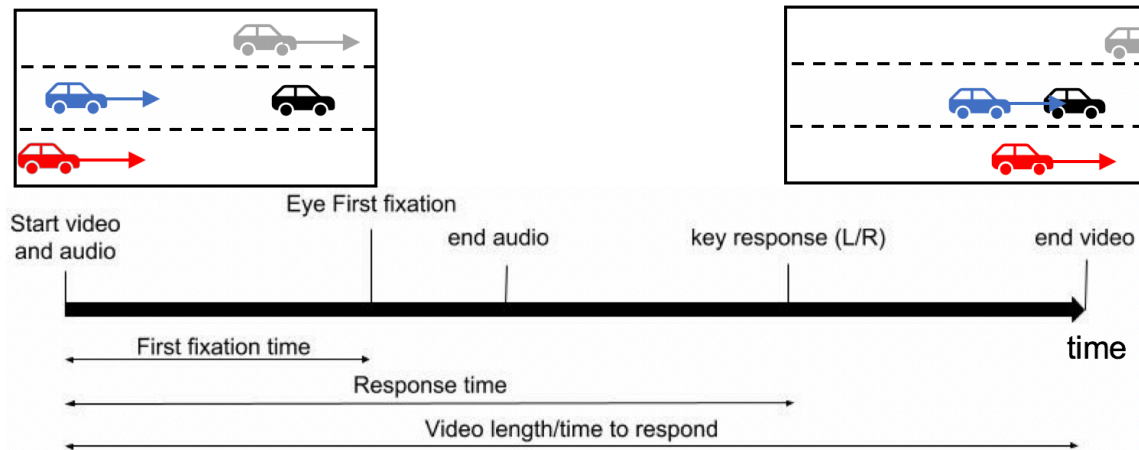


Figure 3. Timeline example of experimental trial where left is the escape direction. Blue car is the ego car, the black car is the stranded car, the red car is the car of the hazard direction and the grey is a car driving at a safe distance. Arrows indicate the approximate velocity of the car.

## 2.5. Experimental Design

A within-subject design was used with three blocks. In each block, one type of auditory take-over requests was provided; the blocks were Latin-square counterbalanced. Each block consisted of 15 trials, which were randomized. Per block, each of the 6 video situations was presented twice, and three invalid take-over requests were provided (see table 1). For each video length, one invalid take-over request is given. 80% of the trials valid feedback is given, which has been shown before to be effective in shortening response time in a similar study on stimulus-reaction responses (Spence, 2005).

Table 1. For each block the number of trials is given for the videos with different video lengths and escape directions. The block order is randomized using Latin-square counterbalancing (\*Invalid trial).

Video length	Block : Danger		Block : Go		Block : Beeps	
	Left	Right	Left	Right	Left	Right
1 second	2	2 + 1*	2 + 1*	2	2 + 1	2
3 second	2 + 1*	2	2	2 + 1*	2 + 1	2
6 second	2 + 1*	2	2	2 + 1*	2	2 + 1

## 2.6. Procedure

Prior to the experiment, participants completed a questionnaire about their age and driving experience. Participants were instructed to adapt the chair height in order for their head and



chin to fit on the head support. Before each block, the right eye was calibrated using the eye-tracking equipment, followed by an instructions screen. The instructions explained the take-over request situation, and that the goal was to switch to a safe lane as soon as possible by pressing the left/right key on the keyboard. When the participant hit a key or when the video was over, the trial ended. Participants were also instructed to place their right hand on the keyboard, eliminating the need to move the arm. Before each video, a drift correction was presented to eliminate small drift on the eye, followed by a screen with a circle located in the lower middle part of the screen. The duration of this screen was either 2, 4, 6, or 8 seconds randomized per trial in order for the participant to not know when the video occurs. Participants were instructed to focus on the circle, until the video started. . After each block, participants completed a questionnaire for assessing the acceptance of new technology (Van der Laan et al., 1997) and NASA task load index (NASA-TLX; Hart and Staveland, 1988) questionnaire.

## 2.7. Independent Variables

The following independent variables were used:

1. **Audio type:** (1) “Danger”, (2) “Go” or (3) Beeps.
2. **Escape direction:** In each video, there was one direction of hazard and one direction of escape (left or right). The escape direction is also the verbal auditory direction (when auditory feedback is valid).
3. **Video length:** The length of the video is equal to the time the driver has to respond (see figure 3).
4. **Audio validity:** 80% of the auditory feedback trials (12 out of 15) were valid in the given direction (left/right), while 20% (3 out of 15) were invalid.

## 2.8. Dependent Variables

1. **First fixation time:** The first fixation time was determined by calculating how much time from the onset of the trial has lapsed until the eye is fixated for at least 100 ms (Velichkovsky et al., 2002) on any area of interest excluding the dashboard (see figure 2). A fixation is when the gaze is maintained on a single location, which can be defined as when the eye is not in a saccade. Saccades are detected by the speed of the eye above a threshold (Eisma et al., 2018), which was set at 26.5 degrees (Salvucci & Goldberg, 2000). The speed of the eye was defined by using Pythagoras theorem for the x and y coordinates and taking the absolute value of the derivative of it.
2. **First fixation location:** The first fixation location was determined by taking the mean of the coordinates during the fixation as described in the first fixation time and splitting the scenes into areas of interest. The area was divided into (1) left mirror (2) left road (3) center road, (4) right road (5) right mirror (6) dashboard, and (7) center mirror (see Figure 1). When no fixations or fixations outside these areas is found, no first fixation is defined.
3. **Escape/hazard detection time:** The escape or danger detection time is defined by correlating the video direction with the first fixation location. The first fixation on the escape side defines the escape detection time, whereas the first fixation on the danger defines the hazard detection time.
4. **Response time (RT):** The RT is the time between the start of the trial and the press of the left or right button. When no button was pressed before the end of the video, no RT was logged.
5. **Response direction (RD):** The RD is which key is pressed by the participant, which can be either left or right; other keys were not logged.

6. **Response correctness:** The correctness is determined by relating the escape direction of the video, with the RD. If the RD is corresponding with the video direction, the correctness is reported as 'Correct'; else the correctness is reported as 'Incorrect'. If no response was given, the correctness is reported as 'Too Late'.
7. **Van der Laan:** A paper questionnaire to determine the usefulness and satisfaction. The mean usefulness score was determined across the following five items: 1. useful-useless; 3. Bad-good; 5. Effective-superfluous; 7. Assisting-worthless; and 9. raising alertness-sleep-inducing. The mean satisfaction score was determined from the following four items: 2. Pleasant-unpleasant; 4. Nice-annoying; 6. Irritating-likeable; and 8. Undesirable-desirable. All items were on a five-point semantic-differential scale. Sign reversals were conducted for items 1, 2, 4, 5, 7, and 9 so that a higher score indicates higher usefulness/satisfaction.
8. **NASA-TLX:** A paper questionnaire for measuring six different types of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. All aspects were filled in with a scale between 1 and 21 with 1 being the lowest workload or demand and 21 the highest workload or demand.

## 2.9. Analyses

Eye data was sampled at 2000 Hz, and filtered using a low-pass filter on 100 Hz to attenuate the high-frequency components, which removed a large part of the high-frequency noise. However, there was still some noise visible in the 50-100 Hz range, which was removed by a moving average filter with a window of 50 frames (i.e., 25 ms). Kolmogorov-Smirnov test for normally distributed data was used to test the response times for normality. For response times and response correctness of valid trials, the mean is taken across repeated trials, and repeated-measures ANOVAs were done for each video length separately. Post-hoc pair-wise

comparisons between audio types were done by means of paired-samples t-test, followed by a correction on the significance level. Using the Bonferroni correction the alpha-value is set to 0.0167 ( $= 0.05/3$ ). Invalid trials only occurred once per audio type and video length, so two sample t-test is used to test the effect on RT. Some trials measured no first fixations time so the data contains missing data points, therefore an unpaired samples t-test was used to test for significance for this metric.

### 3. Results

#### 3.1. Performance measures

In Figure 4 the mean and standard deviation of the RT for each audio type and video length is shown. The ANOVAs yielded a significant difference in RT between audio types for 1 second videos ( $F(2,66) = 5.24$ ,  $p = 0.0077$ ). A post-hoc test shows that audio type 'Go' yielded a significantly faster RT for the 1 second videos situations compared to the baseline ( $M=616.31$ ,  $M=683.65$ ) ( $t(33) = -4.5684$ ,  $p=0.0058$  respectively). For 'Danger' the effect is not significant compared to the audio type 'Go' and the baseline ( $p>0.0167$ ).

For the 3 second videos the ANOVA showed a significant difference in RT ( $F(2,66) = 9.72$ ,  $p = 0.0002$ ). The post-hoc tests showed that the 'Danger' and 'Go' audio types yielded significantly faster RTs than Baseline ( $t(33) = -5.7464$ ,  $p = 0.0004$ ,  $t(33) = -4.9696$ ,  $p = 0.0023$  respectively). There is no significant difference in RT between the audio types 'Go' and 'Danger' in the 3 seconds videos ( $p>0.0167$ ).

For the 6 second video, the ANOVA showed a significant effect on RT ( $F(2,66) = 6.6$ ,  $p = 0.0024$ ). The post-hoc test shows that the audio type 'Danger' yielded a significantly faster RT compared to the baseline group ( $t(33)=4.8562$ ,  $p=0.003$ ). No significant difference was found between other groups ( $p > 0.0167$ ).

In Table 2, the mean and standard deviation of the reaction times are given separately for the left and right direction. There is no significant difference in reaction times between left and right situations of the same video lengths ( $p > 0.05$ ).

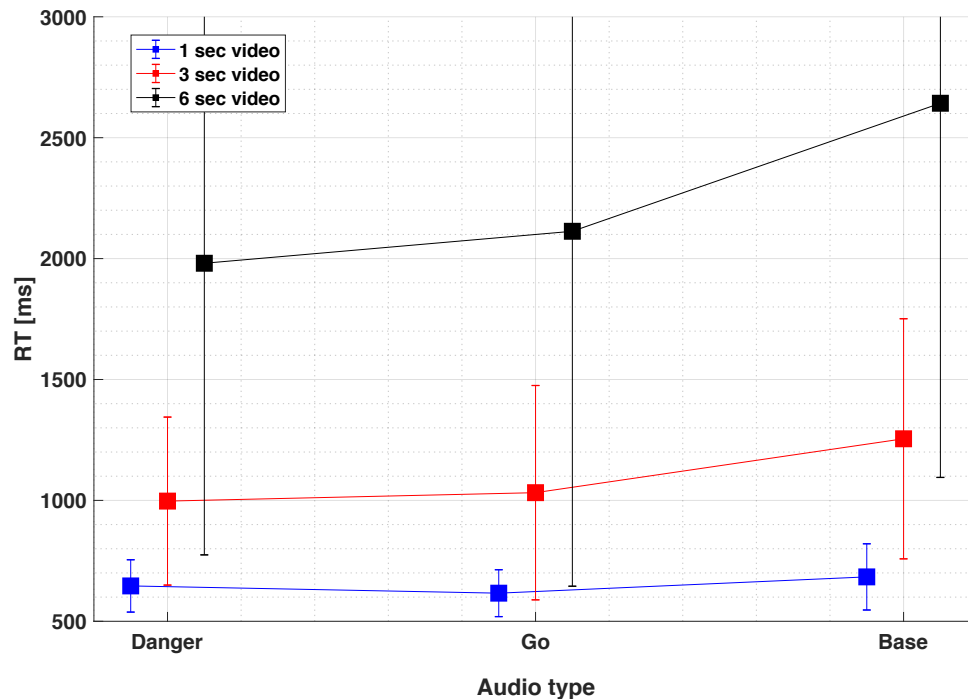


Figure 4. Mean response times and standard deviations across participants grouped per audio type and video length for all valid trials ( $N=34$ ). For each participant the mean is taken over all left and right situations per video length. Mean response times are indicated with the squares and SDs are shown with the error bars from  $-SD$  to  $+SD$

Table 2. Mean and SD of the response times in milliseconds across all participants grouped for each audio and video length and escape directions for valid trials

Video Length	Block: Danger		Block: Go		Block: Beeps	
	Left M(SD)	Right M(SD)	Left M(SD)	Right M(SD)	Left M(SD)	Right M(SD)
1000 ms	447 (105)	418 (123)	397 (107)	419 (141)	547 (157)	512 (237)
3000 ms	672 (290)	679 (278)	711 (344)	681 (308)	1014 (397)	827 (495)
6000 ms	1427 (1080)	1384 (918)	1143 (869)	1486 (1119)	2320 (1671)	1802 (1184)

Figure 5 shows the mean and standard deviation of the percentage of correct responses of each audio and video type. The ANOVA test showed a significant difference in correctness between audio types for 1 second videos ( $F(2,66) = 6.9$ ,  $p = 0.0019$ ). Post hoc tests show that there is a significantly higher percentage of correctness for “Go” and “Danger” compared to

the baseline ( $t(66) = 4.5491$ ,  $p = 0.0056$  and  $t(66) = 4.5491$ ,  $p = 0.0056$  respectively). For 3 and 6 second videos there was no significant difference.

In table 3, the percentage of correctness is reported for left and right situations separately. For the all audio types, more mistakes are made in the R1 situations than the L1 situation, especially for the baseline situation (60.8% compared to 93.1%). Comparing this to the R1 situation for audio types 'Go' and 'Danger', an increase in performance can be seen (88.2%, 85.3% correct responses respectively). In Table 4, the total number of response of all participants to the left, right and no/late responses are given, grouped by audio type and video type. For the baseline, the R1 situations shows a high number of no/late-responses (26) and wrong responses (14). ANOVA tests for 3 and 6-second videos did not show a significant difference in correctness between groups for  $p > 0.0167$ ).

Finally, table 5 shows the mean RT and percentage of correctness responses grouped per audio type and video length trials with invalid auditory feedback. Participants responded approximately twice as slow when provided with invalid auditory feedback compared to valid auditory feedback for the 1 and 3-second videos. For the 6 second videos there was no significant difference in RT between the audio types. Participants also made more mistakes when provided with invalid feedback compared to valid feedback, especially with 6-second videos. 6-second situations seem to be especially confusing when provided with invalid auditory feedback, probably because the danger was not yet visible and the participants felt urgent to respond. Comparing between invalid feedback 'Go' and 'Danger', the mean RTs for 'Go' are slower for each video length, however not significantly different ( $p > 0.05$ ). Participants also made more mistakes when provided with invalid feedback 'Go' compared to invalid feedback 'Danger' for each video length(see table 5).

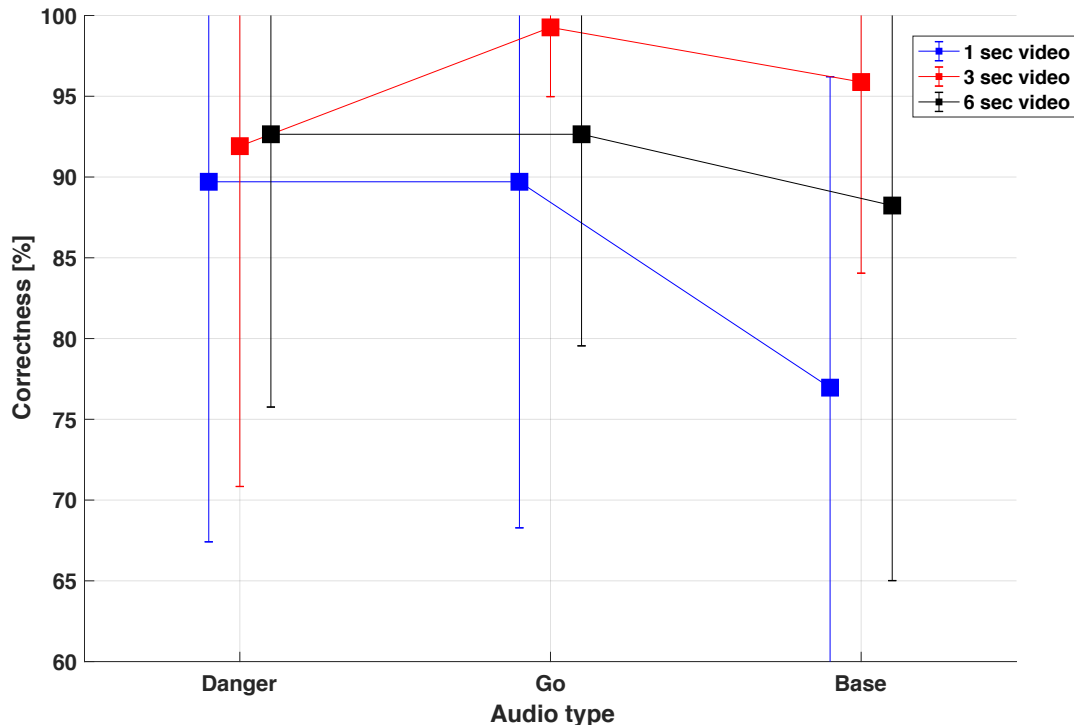


Figure 5. Mean and standard deviations of percentage of correct response across participants grouped per audio type and video length for all valid trials (N=34). Mean percentage of correct responses is indicated with squares; SD is shown with an error bar from -SD to +SD

Table 3. Mean and SD of the percentage of correct responses across all participants grouped per each audio and video length and video direction for all valid trials.

Video Length	Block : Danger		Block : Go		Block : Beeps	
	Left	Right	Left	Right	Left	Right
1000 ms	93.1% (20.5)	85.3% (31.5)	91.2% (22.9)	88.2% (98.5)	93.1% (21.4)	60.8% (29.0)
3000 ms	91.2% (22.9)	92.6% (21.8)	100.0% (0.0)	98.5% (8.6)	95.1% (14.5)	97.1% (11.9)
6000 ms	88.2% (27.7)	97.1% (11.9)	92.6% (18.0)	92.6% (21.8)	86.8% (30.9)	89.7% (26.9)

Table 4. Number of responses in left/right directions for all valid trials grouped by audio type, video length and escape direction. Also no/late responses are indicated by the \*.

Video Length	Audio type									
	Block 1: Danger				Block 2: Go				Block 3: Beeps	
	Left		Right		Left		Right		Left	Right
	L / R		L / R		L / R		L / R		L / R	L / R
1000 ms	64	3	7	58	62	5	6	60	95	2
	1*		3*		1*		2*		5*	
3000 ms	62	5	5	63	68	0	1	67	97	3
	1*		0*		0*		0*		2*	
6000 ms	60	8	2	66	63	4	4	63	59	9
	0*		0*		1*		1*		0*	
Total	186	16	14	187	193	9	11	190	251	14
									20	189

Table 5. Mean and SD of the RT and the percentage correct responses across all participants grouped per audio and video for invalid trials.

Video Length	Block: Danger (invalid)		Block: Go (invalid)	
	RT	CR	RT	CR
1000 ms	857.0 (220.7)	82.4% (39.0)	870.6 (190.0)	79.4% (41.0)
3000 ms	1394.4 (609.0)	82.4% (39.0)	1473.1 (660.0)	64.7% (49.0)
6000 ms	1569.1 (814.29)	50.0% (51.0)	1571.5 (907.7)	44.1% (50.0)

### 3.2. Eye-tracking measures

In Figure 6, the first fixation locations are summed up and divided by the total number of first fixation per group of audio type, video length, and escape direction to get the percentage. In the 1 second videos, participants looked mostly at the center of the road for all audio types (59.0%, 60.0%; 73.9%, 55.0%; 55.9%, 32.8% for L1, R1 and 'Danger', 'Go' and Baseline respectively). At the center the stranded car was approaching the driver, and therefore was the most salient visual stimuli in these videos. Interestingly, for the baseline R1 situation the lowest percentage of first fixations on the center of the road is measured, as well as the lowest percentage of correctness. In this situation it seems that participants fixate more on the right road, than in other audio types.

In the 3 second videos, participants were more likely to place their first fixation on the hazard side compared to the escape side in all audio types (80.5%), while for 1 second and 6 second videos this was more equal (49.5%, 53.8% respectively) (see Figure 7). The overtaking car at the hazard side was the most salient visual stimuli in these 3-second videos, which may explain the focus on this area. For 6-second situations, the visual gaze is more distributed, for all audio types. Comparing the amount of left and right first fixations for all participants per audio type, results show that participants were more likely to first check the right side, compared to the left side (29%/17%, 24%/17%, 22%,16% for 'Danger', 'Go' and baseline respectively). An explanation for this could be the left seat position in which the ego-person is



placed in the car, giving a better view on the right road, or that drivers have the automatism to check the right side first.

In Figure 8, the first fixation times are grouped per fixations on the escape direction and hazard direction. Participants who placed their first fixation on the hazard direction had a significantly faster first fixation time than participants who placed their first fixation on the escape direction for video lengths of 3 seconds and all audio types (Figure 10). No difference in first fixation time is found between same video situations but different audio type ( $p>0.05$ ). Pearson's linear correlation coefficient was computed to find if there is a relationship between the first fixation time and response time. There was a weak positive linear relationship between first fixation time and response time. There was a weak positive linear relationship between first fixation time and response time for 1,3 and 6 second videos ( $r=0.17$ ,  $r=0.25$  and  $r=0.17$  respectively).

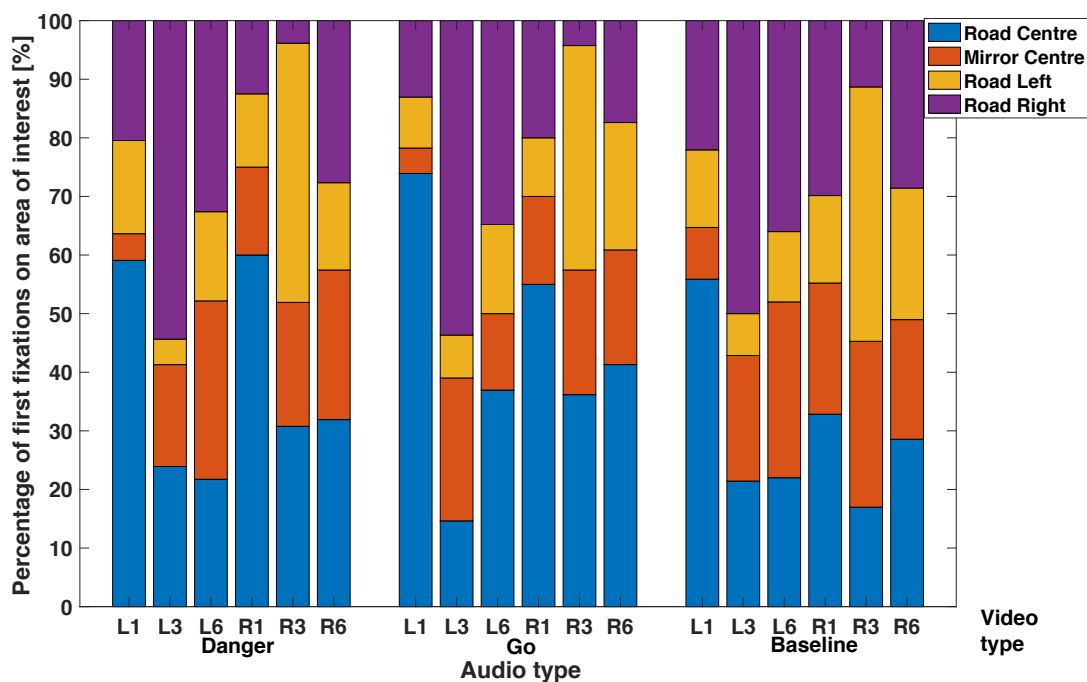


Figure 6. Percentage of first fixations placed on each area of interest grouped per audio type, video length and direction. (L1 = left video of 1 second etc.)

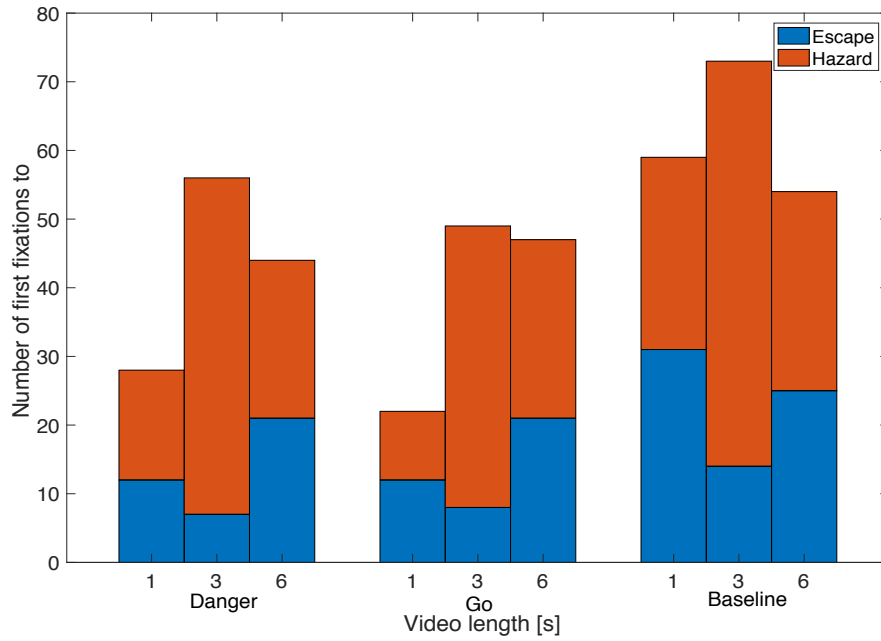


Figure 7. Number of trials for which the first fixation is placed on the hazard or escape side

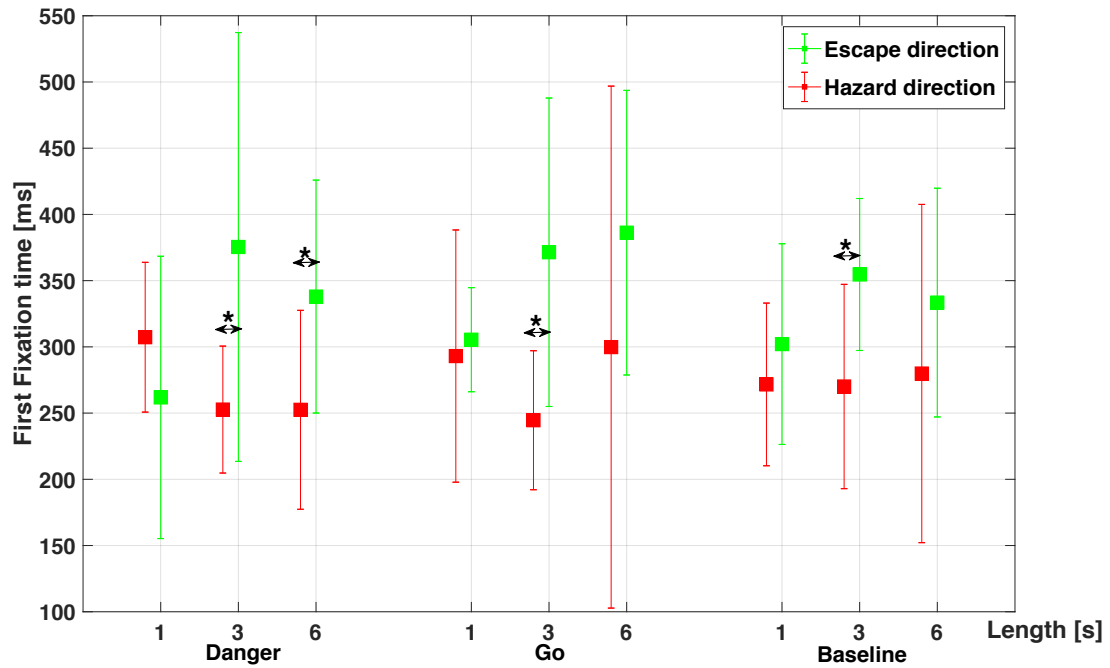


Figure 8. Mean and standard deviation of first fixation times on hazard or escape for all valid trials grouped per audio type and video length. T-tests shows significant differences indicated with the asterisk (\* $p < 0.05$ )

### 3.3. Subjective measures

The NASA-TLX questionnaires showed a lower temporal workload and effort with the audio type “Go” as compared to the other two audio-type conditions. For all other measures the means of the scores do not vary more than 5% between all audio types. Physical workload

was low for all audio types. From the acceptance questionnaire, the baseline audio type was rated as the most useless and least accepted by participants. The audio type 'Danger', 'Go', and the baseline were evenly rated for usefulness (0.53, 0.52, 0.53, respectively). Comparing the satisfaction score of 'Danger', 'Go' and the baseline, 'Go' scored the highest (-0.08, 0.24, -0.15), however not significant ( $p > 0.05$ ). The baseline scored the highest in raising alertness ( $t(66) = 0.98398$ ,  $p = 0.007$  for 'Danger' and  $t(66) = 4.4279$ ,  $p = 0.045$  for 'Go').

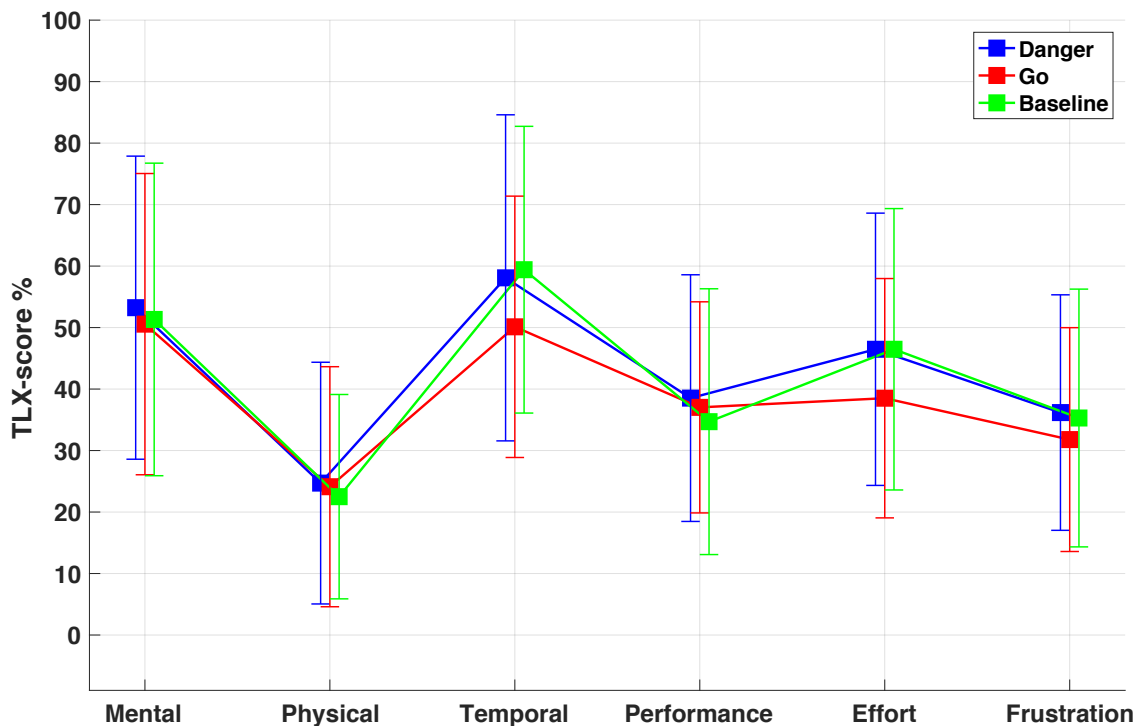


Figure 9. Mean and standard deviation of the NASA-TLX questionnaire scores plotted for each audio type

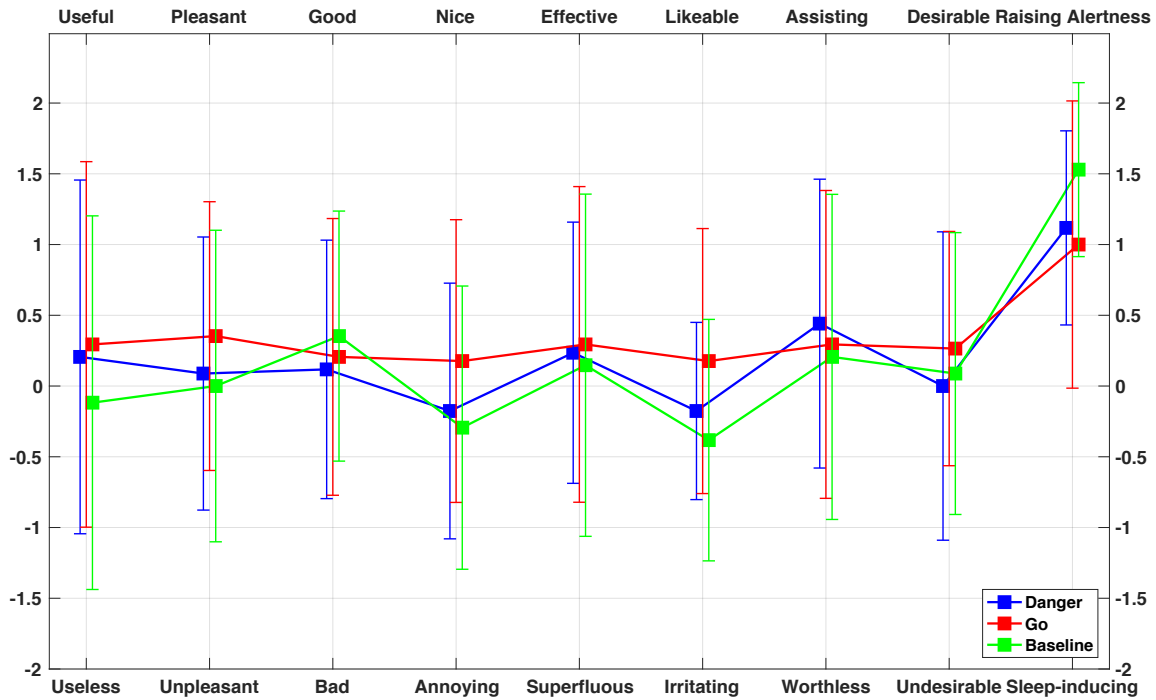


Figure 1. Mean and standard deviation of the van der Laan questionnaire scores plotted for each audio type

## 4. Discussion

The aim of this study was to investigate drivers' response time and visual behavior when provided with directional auditory TOR in an emergency situation. Videos of emergency take-over situations were shown to participants with a time-to-respond of 1,3 and 6 seconds. Participants were provided with ipsilateral and contralateral auditory feedback by means of 'Go left/right', 'Danger left/right' and compared to warning beeps. Participants' performance was evaluated using the response time and the response correctness. Eye-tracking was used to evaluate detection time of the danger based on the first fixations.

### 4.1. Ipsilateral and contralateral feedback take-over performance

The results showed that for 1 and 3 second videos ipsilateral feedback 'Go' led to significantly faster responses compared to the baseline, and for 3 and 6 second videos contralateral feedback 'Danger' led to significantly faster responses compared to the baseline. 'Go' and

'Danger' did not have a significant difference in response time between each other in all video lengths. It seems that there is more effect of ipsilateral feedback when there is less time to respond and contralateral feedback when there is more time to respond. This finding is in line with the study of Straughn et al. (2009), who found that drivers' made a safer maneuver when presented with ipsilateral feedback compared to contralateral feedback when less time is available and vice versa. However, in a similar study of Wang et al. (2007) the results shows a faster response for contralateral auditory beeps. In this study participants drove in a simulator, where a car appears after 0.5 seconds and dangerously merges into the drivers' lane after another 0.5 seconds; It is not clear how long it takes to merge into the lane. Auditory beeps were given before the hazardous car was visible, which might explain the effect of contralateral feedback. We may compare this result to the result of this study in the 6 second situation, where the hazard was also not visible from the start. Also noticed was that participants first looked to assess the situation visually before making a decision, this may be related to a lack of trust in the system, or that drivers prefer to choose visual information instead of an artificially produced sound. From the response correctness results it was found that in the R1 situation participants were especially errorous. This video situation may be more difficult, because the time-to-respond was very short and drivers are not used to overtake a car on the right side, or that the view on the left side is smaller than the right side. Both ipsilateral and contralateral helped the participant in making more correct decisions in this situation, while for other videos this effect was not as large, therefore we can conclude that depending on the situation complexity, drivers can benefit more from ipsilateral or contralateral feedback compared to simple beeps.

Finally, results show that when providing invalid auditory feedback, responses are approximately twice as slow in 1 and 3-second situations compared to valid feedback. Participants need extra time to respond accordingly in these situations, therefore validity

should be considered in practical implementation of directional take-over requests, especially if this feedback is only given very rarely.

#### *4.2. Visual gaze behavior*

The hypothesis was that the audio 'danger' induces faster detection time of hazard, because of facilitation of visual search (McDonald et al., 2000). However, we observed a very similar pattern in the first fixations in the audio types 'Go' and 'Danger' and the baseline for all videos, which tells us that the audio type did not affect the gaze behavior in these situations. It seems that humans respond to salient visual stimuli in the peripheral vision, which explains the visual behavior in the 1 and 3 second situations to the center of the road and the overtaking car respectively, and the more distributed behavior in the 6 second situations. Explanations for why the audio did not influence the visual gaze may be that: (1) The audio and video were played at the same time, and participants' gazes were already focused on the dashboard not far from the road before the video started, it is likely that participants fixated based on the visual scene first. (2) First fixation times happened before the audio ended. (3) Participants use audio and visual stimuli as two independent channels for information processing, one for auditory and one for visuals (Wickens 1991), so the use of audio is not only to attract attention (Rossmeier et al. 2005). It would be interesting to examine how providing audio before the video starts effects the gaze behavior, and how the gaze behavior is in driving simulators or in field studies.

#### *4.3. Transfer to real life*

From previous studies it is shown that simple stimulus-reactions task results do not transfer into natural environments, because a natural scene is much more complex (Musseler et al. 2002). The steering response in our study is obviously simplified, but the videos show

'natural' scenes so the decision making until the response stays the same. Also, in real or simulated driving it is more difficult to define response times and response correctness, since it is a dynamic task and the reaction is made over a period of time. A simple key response, eliminates the steering behavior and other variables from the response time. Eye gaze behavior may not be the same in real-life scenarios. The head was fixed in a chin-rest position, and no head turns can be made, because the head had to stay in position for accurate eye scanning. Also the situation was displayed on a monitor providing a small field of view and low immersion, so participants field of view covered the whole monitor. We expect that, directional auditory TORs may be more effective when hazardous objects do not occur in the field of view, but this should be investigated first. Finally, the frequency of take-over requests was very high, and participants were expecting it. When exposed to a novel sound in case of HAD, the driver might have an spontaneous response, changing the visual behavior (Driver & Spence 1998).

## **5. Conclusion**

The effect of verbal directional auditory take-over requests was studied, by an eye-tracking experiment showing videos of traffic situations. Ipsilateral and contralateral audio by means of "Go left/right" and "Danger left/right" and spatially located in the direction of the instruction direction were compared with the beeps as baseline without a spatial characteristic and were compared with each other. It was hypothesized that the audio type 'Go' would have faster reaction times, because the audio direction is congruent with the response direction, while 'Danger' would help in faster hazard detection, because it can grab visual attention. The following conclusions are made:

- Auditory directional take-over requests 'Go left/right' (ipsilateral) and 'Danger left/right' (contralateral) both can increase the performance of take-over requests, by reducing response time, and increasing the correctness of responses.
- When there is a short time to respond (1-3 seconds) verbal ipsilateral auditory TOR is beneficial in reducing response time, however when there is more time to respond (>3 seconds) verbal contralateral auditory TOR is preferred.
- Eye gaze behavior is mainly the result of the visual stimuli and may be independent of auditory stimuli.
- Drivers have a natural tendency to place their first gaze on the most salient visual stimuli.

In conclusion, drivers can benefit from directional information provided during take-over requests, because it can aid the driver in taking-over the car faster and more safely.

## 6. References

- Apple Inc. 2018. Audio software. <https://www.apple.com/nl/ios/garageband/>
- Cohen-Lazry, G., & Katzman, N. Ipsilateral Versus Contralateral Tactile Alerts for Take-Over Requests in Highly-Automated Driving.
- Eisma, Y. B., Cabrall, C. D., & de Winter, J. C. (2018). Visual Sampling Processes Revisited: Replicating and Extending Senders (1983) Using Modern Eye-Tracking Equipment. *IEEE Transactions on Human-Machine Systems*, (99), 1-15
- Endsley, M. R. (1988, May). Situation awareness global assessment technique (SAGAT). In *Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National* (pp. 789-795). IEEE.
- Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013, September). "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* (Vol. 57, No. 1, pp. 1938-1942). Sage CA: Los Angeles, CA: SAGE



Publications.

- G. Hart, Sandra & E. Stavenland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* 52. 139-. 10.1016/S0166-4115(08)62386-9.
- Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of experimental psychology: Applied*, 11(3), 157.
- D. Van Der Laan, Jinke & Heino, Adriaan & Waard, Dick. (1997). A simple procedure for the assessment of acceptance of advance transport telematics. *Transportation Research Part C: Emerging Technologies*. 5. 1-10. 10.1016/S0968-090X(96)00025-3.
- Lai, F. C. H., & Barnard, Y. (2010). Spotting sheep in Yorkshire: Using eye-tracking for studying situation awareness in a driving simulator.
- Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation-awareness? A laboratory-based study of automated driving. *Applied ergonomics*, 60, 293-304.
- Naturalsoft Ltd. 2018. Natural Reader Online, text-to-text speech web application. Retrieved from: <https://www.naturalreaders.com/webapp.html>.
- Müsseler, J., Aschersleben, G., Arning, K., & Proctor, R. W. (2009). Reversed effects of spatial compatibility in natural scenes. *The American journal of psychology*, 325-336.
- Petermeijer, S., Bazilinskyy, P., Bengler, K., & 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.
- Rimini-Doering, M., Altmueller, T., Ladstaetter, U., & Rossmeier, M. (2005). Effects of lane departure warning on drowsy drivers' performance and state in a simulator.
- Rosmeier, M., Grabsch, H. P., & Rimini-Doring, M. (2005). Blind flight: Do auditory lane departure warnings attract attention or actual guide action?. Georgia Institute of Technology.
- Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*(pp. 71-78). ACM.
- M. Sanders and E. McCormick (1987). *Human Factors in Engineering Design*. 'New York, NY: \ McGrawHill, Reading, Massachusetts.

- Wang, D. Y. D., Pick, D. F., Proctor, R. W., & Ye, Y. (2007). Effect of a side collision-avoidance signal on simulated driving with a navigation system.
- Shladover, S. E. (2015, November). Road vehicle automation history, opportunities and challenges. In *Mini-Seminar 'Developments Selfdriving Vehicles in USA* (Vol. 9).
- Spencer Jr, B. F., Dyke, S. J., & Deoskar, H. S. (1998). Benchmark problems in structural control: Part I—Active mass driver system. *Earthquake Engineering & Structural Dynamics*, 27(11), 1127-1139.
- Straughn, S. M., Gray, R., & Tan, H. Z. (2009). To go or not to go: Stimulus-response compatibility for tactile and auditory pedestrian collision warnings. *IEEE Transactions on Haptics*, 2(2), 111-117.
- Suzuki, K., & Jansson, H. (2003). An analysis of driver's steering behaviour during auditory or haptic warnings for the designing of lane departure warning system. *JSAE review*, 24(1), 65-70.
- Tass International, 2015. PreScan: Simulation of ADAS & Active Safety. Retrieved from. <https://www.tassinternational.com/prescan>.
- Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145-156.
- Wang, D. Y. D., Proctor, R. W., & Pick, D. F. (2003). Stimulus-Response Compatibility Effects for Warning Signals and Steering Responses.
- Wickens, C. D. (1991). Processing resources and attention. Multiple-task performance, 1991, 3-34.
- De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, 27, 196-217.



# Appendices

# Content

## Appendix A – Experimental setup

A.1 Hardware

A.2 Software

## Appendix B – Experiment library

B.1 Audio clips

B.2 Video clips

## Appendix C – Task instructions

## Appendix D – Consent form

## Appendix E – Questionnaires

E.1 Participant questionnaire

E.2 Van der Laan

E.3 NASA TLX

## Appendix F – Data Analysis

F.1 Filters (preprocess)

F.2 Fixation detection

F.3 Areas of interest

F.4 Creating table of all variables

F.5 Statistics

## Appendix G – Literature report

## Appendix A – Experimental setup

### A.1 Hardware



Figure 2. Experimental setup. In this picture is shown the head mount, Eyelink 1000 plus, the screen and input devices

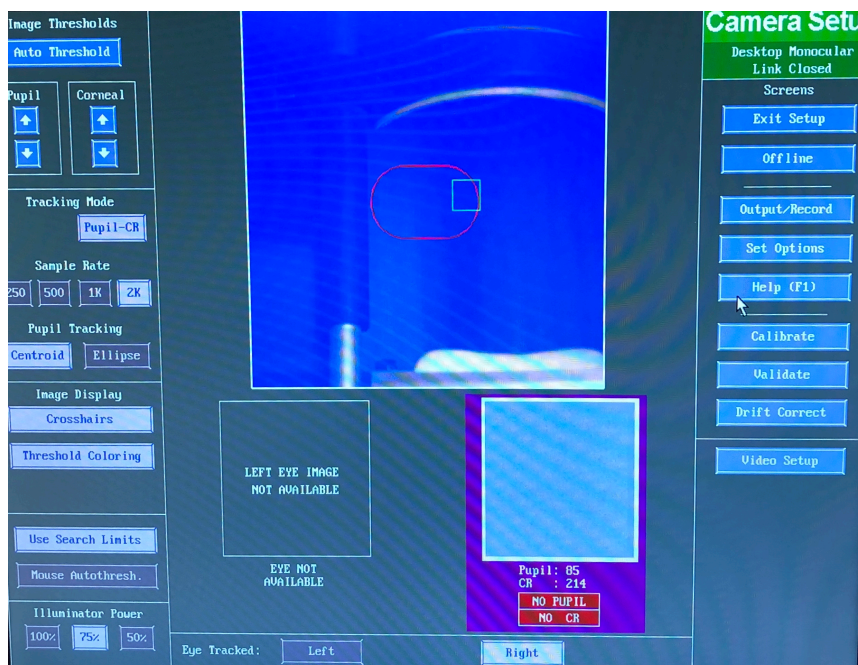


Figure 3. Eye Link experimenter interface, for controlling calibration, validation and eye detection parameters.

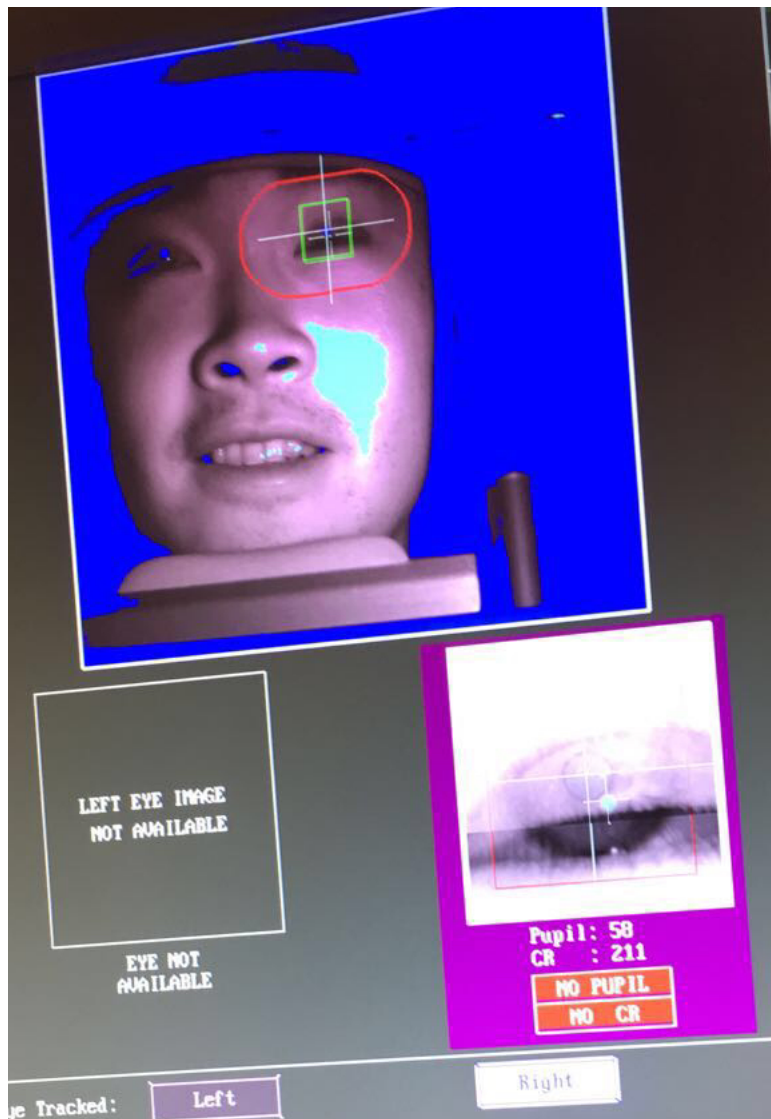


Figure 4. Example of eye-tracking the right pupil

## A.2 Software

Experiment builder (SR Research 2016) was used to create the experiment. The experiment exists of the experiment block with a fixation block and recording block, and calibration blocks. The recording block contained the playing of the audio sound, followed by the displaying of the video. By tracking global time, only a few ms difference was measured between these blocks. The timer measures the elapsed time, and the screen turns black if the video time is reached. The video is prematurely ended if a keyboard key is hit (see figure 12 left). In the right scheme of figure 12 is shown how the answer is updated and checked if the answer is corresponding with the video. In the end the results are saved in the result file.

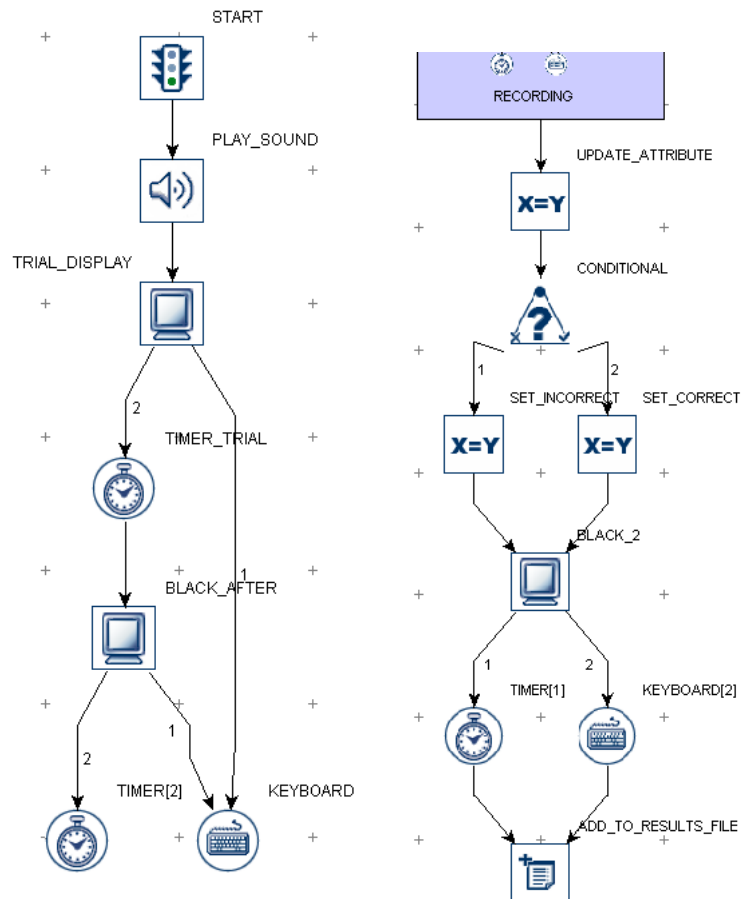


Figure 5. Left: Experiment builder block scheme of the recording trial where the audio and video are played, and the eye gaze is recorded. Right: Experiment builder block scheme after recording trial.



## Appendix B – Experiment library

### B.1 Audio clips

Audio clips that were used in the experiment were made using the a speech-generating website (NaturalSoft Ltd. 2018) The words “Go left”, “Go right”, “Danger left”, “Danger right” were used to create these words and spoken by ‘English (UK) – Selene’ with a speed of 1. The audio clips were edited in Garageband to all start at 50ms. “Go left/right” had a length of 600ms and “Danger left/right” had a length of 680ms (see figures 10)

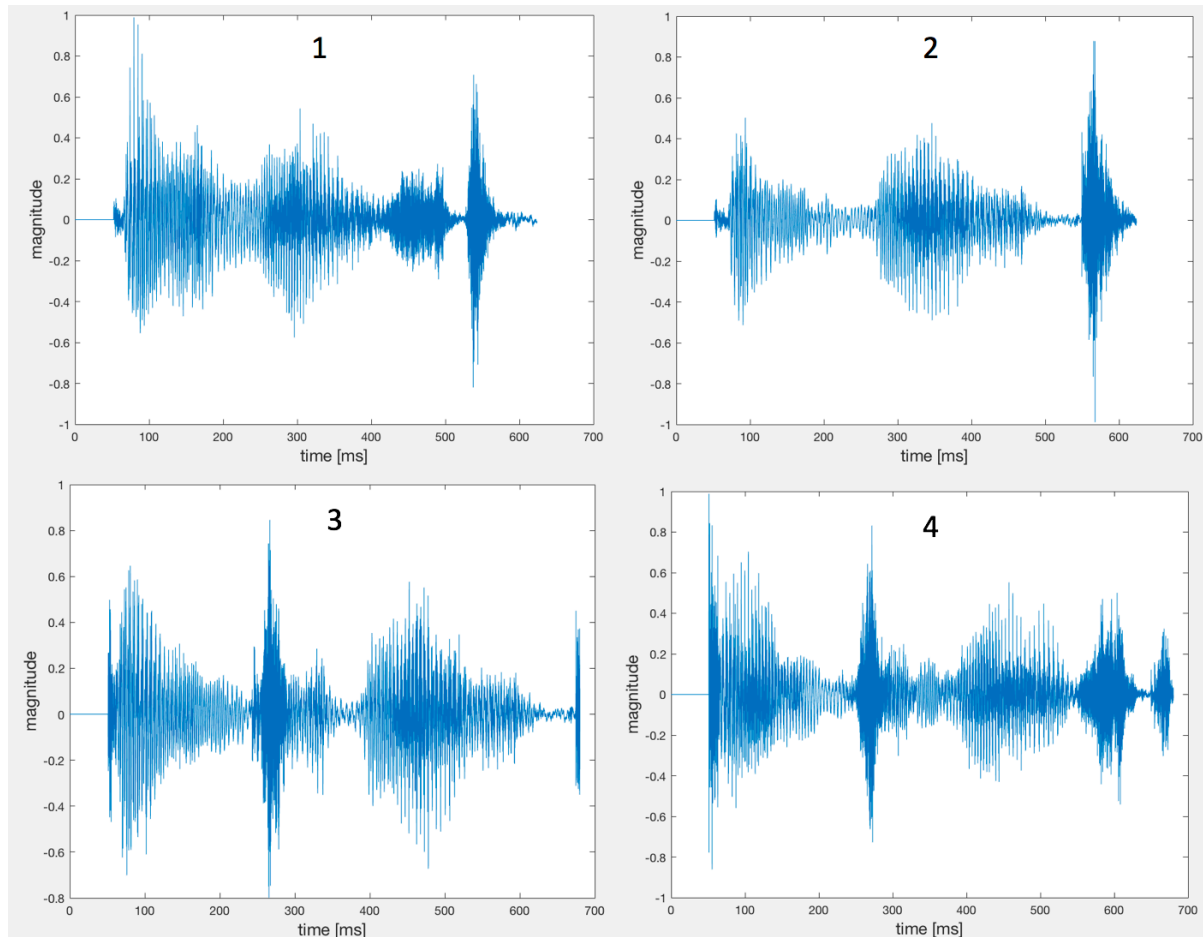


Figure 6. Audio sound waves for (1)"Go left", (2)"Go right", (3)"Danger left", (4)"Danger right" Source: <https://youtu.be/7re9sya4eQY>

### B.2 Video clips

Video clips were made by Zhenji Lu for his study on situational awareness (Lu et al. 2018) in Prescan software (Tass International 2015). In this study six videos were used: 1 second, 3 second and 6 second videos, and left and right situations for each video length. In table 5,6,7 the properties of the surrounding cars are given for the L1, L3 and L6 situations. The right situations are mirrored opposed to the left situations. The speed of the ego car was 27.78m/s and the stranding car in the middle lane (Car1) has an acceleration of  $-5\text{m/s}^2$ . In figure 11-16 the start of each video is shown. Although videos situations are mirrored, the driver is seated in the left position, and therefore has a better view on the road on the right side compared to the left side. The left window is visible, while the right window can only be looked through, through the right mirror.

<b>L,1</b>	Car1	Car2	Car3	Car4	Car5
Speed[m/s]	27,8	27,8	33,3	33,3	22,2
Lane	2	1	3	3	1
Last Position[m]	15	-75	5	-20	50
Initial Position[m]	317,5	225	299,4	274,4	355,6

Table 4. Properties of cars in situation L1. (Lane 1=left, Lane2=middle, Lane3=right)

<b>L,3</b>	Car1	Car2	Car3	Car4	Car5
Speed[m/s]	27,8	33,3	22,2	27,8	33,3
Lane	2	3	1	2	3
Last Position[m]	16	35	-40	-45	-10
Initial Position[m]	338,5	308,3	276,7	255	273,3

Table 5. Properties of cars in situation L3. (Lane 1=left, Lane2=middle, Lane3=right)

<b>L,6</b>	Car1	Car2	Car3	Car4	Car5
Speed[m/s]	27,8	22,2	27,8	33,3	27,8
Lane	2	1	3	3	1
Last Position[m]	17	-50	15	-25	45
Initial Position[m]	406,5	283,3	315	241,7	345

Table 6. Properties of cars in situation L6. (Lane 1=left, Lane2=middle, Lane3=right)



Figure 7. First frame of the 1 second video of left escape situations (L1) Source: <https://youtu.be/LAB9C0EA1dQ>



Figure 8. First frame of 3 second video and left escape situations (L3) Source: <https://youtu.be/nTJVegZR8wl>



Figure 9. First frame of 6 second video and left escape situations (L6) Source: <https://youtu.be/AikQlrVNp-k>



Figure 10. First frame of 1 second video and right escape situations (R1) Source: <https://youtu.be/WcNQizcFkq0>



Figure 11. First frame of 3 second video and right escape situations (R3) Source: <https://youtu.be/-EiqxzIjQCM>



Figure 12. First frame of 6 second video and right escape situations (R6). Source: <https://youtu.be/vs-kEqtGZnY>

In experiment builder, the audio block plays the audio first, and the video follows. The time between the starts of these audio and video files were  $<5\text{ms}$ , therefore we say that the audio and video both start playing at the same time. An example of the trials where audio and videos are played can be found on: <https://youtu.be/-rSZgm0gQ4M>.

## Appendix C – Instruction screen

Before the start of each block, the instructions are given (see text below):

Now the experiment starts.

Preparations:

1. Keep your head in the chin-rest. Only move your head when asked to.
2. Place and keep your right hand on the arrow keys, and your left hand on the space bar.

Instructions:

1. Look at the small circle at the bottom and press the space bar.
2. Focus your attention to the large grey circle at the bottom of the screen.
3. You will be presented with videos of a potential forward collision.
4. Your task is to avoid the car by going left or right, do this by pressing the left or right key.
5. Repeat this for all trials

-

PRESS ANY KEY TO CONTINUE



## Appendix D – Consent form

### Consent form for participants

Research Title: "Effects of auditory verbal directional cues on driver behavior: A desktop based eye-tracking study"

#### Researchers:

Jimmy Hu – Msc Biomechanical Design

E-mail: [j.hu-4@student.tudelft.nl](mailto:j.hu-4@student.tudelft.nl)

Dr.ir. Joost de Winter – Supervisor

Email: [j.c.f.dewinter@tudelft.nl](mailto:j.c.f.dewinter@tudelft.nl)

S.M. Petermeijer – Supervisor

Email: [s.m.petermeijer@tudelft.nl](mailto:s.m.petermeijer@tudelft.nl)

#### Location of the experiment:

Driving simulator lab; room 34 F-2-360

Faculty of Mechanical, Maritime and Materials Engineering

Delft University of Technology

Mekelweg 2, 2628 CD Delft

**Introduction:** Please read this consent document carefully before you decide to participate. This document described the purpose, procedure and potential risks/discomforts. Your signature is required for participation.

**Purpose of the study:** Taking back control from an automated vehicle has been shown to be an important human factors issue. The purpose of this study is to investigate the verbal cues "go left/right" and "danger left/right" as take-over requests, to assess the effectiveness of these cues. During this study your eyes will be tracked and your decision-making will be recorded during the driving scenarios in order to gain insight into the humans' reaction to the verbal directional take-over requests.

**Duration:** Your participation in this experiment will last approximately 20 minutes.



Figure 18. Experimental setup with head support and eye tracker

### **Procedure and instructions:**

**Before the experiment starts:** You will be asked to sign this consent form prior to the experiment. You will be asked to rest your head on the support (see Figure 1) and follow the instructions given on the screen.

**During the experiment:** First a circle is shown at the bottom of the screen; your goal is to focus on this circle. At a random instance between 2 and 8 seconds, you will receive auditory take-over request and you will see a driving scenario on the main screen. In the scenario you are driving in an automated car with traffic around you. You have to make an appropriate reaction by choosing to take-over at the left or right side by pressing the left or right key on the keyboard. Your goal here is to make a safe decision as soon as possible. Each participant has to do 3 blocks of 15 trials per block. In each block a different auditory cue will be presented in a randomized order.

- A. Non-verbal beeps
- B. Verbal: "Go left/right"
- C. Verbal: "Danger left/right"

However, keep in mind that the auditory cue is not always correct.

**After the experiment:** You will be asked to complete a short questionnaire about your gender, age, and driving experience.

**Risks and discomforts:** There are no known risks for you in this study. Some minor eyestrain or discomfort may arise from the task.

**Confidentiality:** All data collected in this study will be kept confidential and will be used for research and/or educational purpose only. You will not be personally identifiable in any future publications based on this work or in any data files shared with other researchers.

**Right to refuse or withdraw:** Your participation in this study is entirely voluntary. You have the right to refuse or withdraw from this experiment at any time, without any negative consequences, and without needing to provide any explanation.

**Questions:** For any questions, you can contact Jimmy Hu ([J-hu.4@tudelft.nl](mailto:J-hu.4@tudelft.nl))

I have read and understood the information provided above. I give permission to process the data for the purpose described above. I voluntarily agree to participate in this study.

Name:

.....

Signature:

.....

Date:

...../...../.....



## Appendix E – Questionnaires

Three questionnaires were given in this experiment. Each participant started off with the participant questionnaire before the start of the experiment. After each block, the van der Laan questionnaire, and NASA TLX were given to the participants.

### E.1 Participant Questionnaire:

**Age:**

.....

**Gender:**

- ☐ Male
- ☐ Female
- ☐ I prefer not to respond

**Driver license:**

- ☐ Yes
- ☐ No
- ☐ I prefer not to respond

**How much do you drive?**

- ☐ Almost Everyday
- ☐ 1-3 times a week
- ☐ Less than once a week
- ☐ Less than once a month
- ☐ Almost Never
- ☐ I prefer not to respond

**Do you have any visual deficiencies?**

- ☐ No
- ☐ Yes
- ☐ I prefer not to respond

**Were you wearing any seeing aids to correct your vision during the experiment?**

- ☐ No
- ☐ Yes, glasses
- ☐ Yes, contact lenses
- ☐ I prefer not to respond

**Do you have any auditory deficiencies?**

- ☐ No
- ☐ Yes
- ☐ I prefer not to respond

**Do you have any problems in separating left and right?**

- ☐ No
- ☐ Sometimes
- ☐ Often
- ☐ I prefer not to respond

## E.2 Van der Laan Questionnaire:

**The audio cue “go left/right” in the trials were:**

- |                     |           |                |
|---------------------|-----------|----------------|
| 1 Useful            | _ _ _ _ _ | Useless        |
| 2 Pleasant          | _ _ _ _ _ | Unpleasant     |
| 3 Bad               | _ _ _ _ _ | Good           |
| 4 Nice              | _ _ _ _ _ | Annoying       |
| 5 Effective         | _ _ _ _ _ | Superfluous    |
| 6 Irritating        | _ _ _ _ _ | Likeable       |
| 7 Assisting         | _ _ _ _ _ | Worthless      |
| 8 Undesirable       | _ _ _ _ _ | Desirable      |
| 9 Raising Alertness | _ _ _ _ _ | Sleep-inducing |

### E.3 Nasa TLX:

Mental Demand

How mentally demanding was the task?


Very Low

Very High

Physical Demand      How physically demanding was the task?

Very Low      Very High


Temporal Demand      How hurried or rushed was the pace of the task?



Very Low      Very High

Performance


How successful were you in accomplishing what you were asked to do?



Perfect Failure

Effort

How hard did you have to work to accomplish your level of performance?



Very Low

Very High

Frustration

How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low

Very High

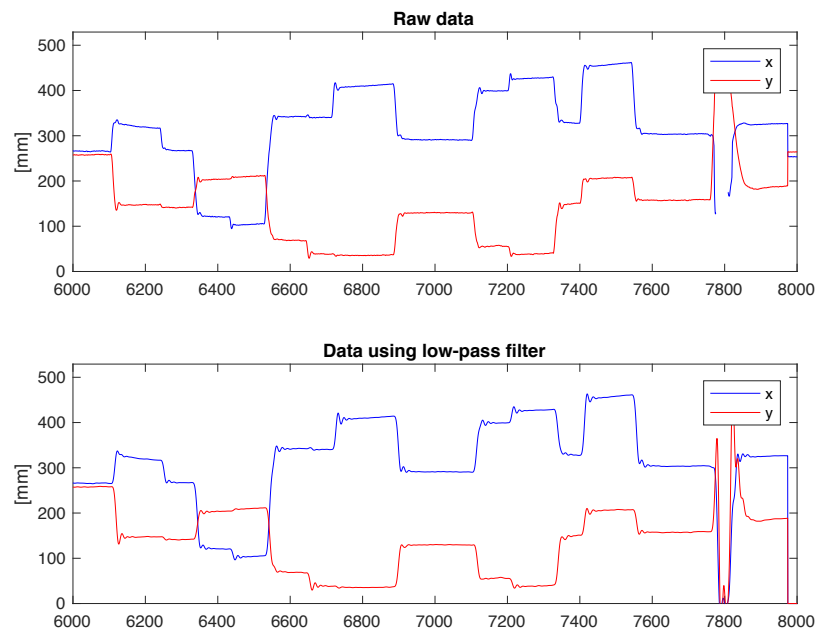
## Appendix F – Data Analysis in Matlab

### F.1 Filtering

A 6<sup>th</sup>-order Chebychev low-pass filter was used (Matlab Script 1). Sampling frequency (fs) was set to the frequency of the eye-scanner, 2000Hz. And the cut-off frequency was set to 100Hz [Nyquist], because we want to gather data up to 50Hz [source]. After filtering, data still contained some overshoot and oscillation (see figure 18). Therefore, a moving average with a window size of 50 was applied over the data, removing the overshoot and oscillation (see figure 19).

```
fs=2000;  
fc=100;  
  
[b,a] = cheby2(6,40,fc/(fs/2));  
x = filter(b,a,x);  
y = filter(b,a,y);  
  
x=movmean(x,50);  
y=movmean(y,50);
```

*Matlab Script 1. Filtering of raw eye data (x,y) by a 6-th order chebychev2 low pass filter, followed by a moving average with a window size of 50 in matlab.*



*Figure 13. X and Y coordinates of the eye-data for participant 6 in trial 1 showing the raw data in the top plot, and the low-pass filtered data in the lower plot.*

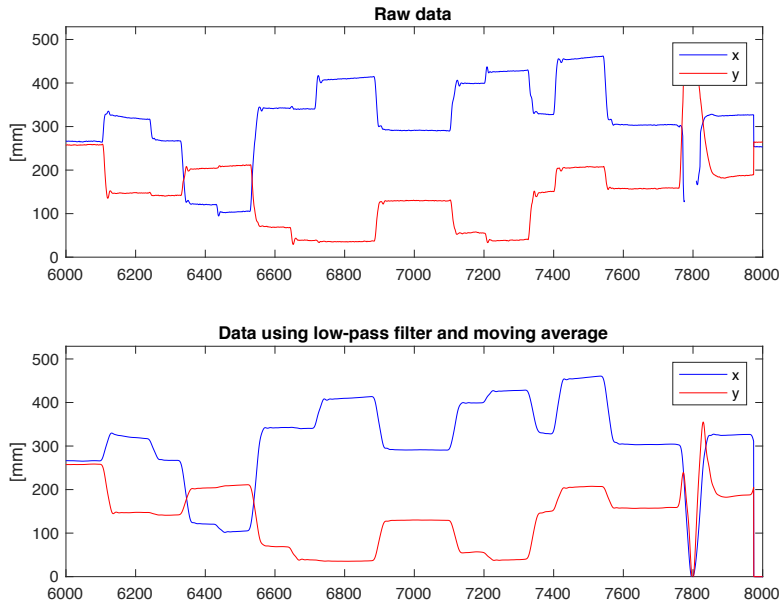


Figure 14. X and Y coordinates of the eye-data for participant 6 in trial 1 showing the raw data in the top plot, and low-pass filtered data with a moving average in the lower plot.

## F.2 Areas of interest

All coordinates were processed through a function, where x and y coordinate were related to an area of interest in the video. Initially, distinctions were made between: Dashboard, left mirror, left window, left road, left grass, left air, center mirror and all the right counterparts (except right window). However, Cars can move along the air and grass from the drivers' viewpoint, and a very low percentage of first fixations were placed on the left window, so the areas of interest were simplified into: Dashboard, left mirror, right mirror, left road, right road, center road and air (figure 1).

All areas of interest were arranged into categorizations to improve performance in matlab. Coordinates were manually selected in figure1 using matlab function 'getpts'. To find if the coordinates are within shapes of a higher order, the matlab function 'inpolygon' was used (Matlab script 2).

## F.3. First fixation direction

Fixation detection was done by first calculating the velocity of the eye at all instances. The velocity was calculated using Pythagoras' theorem on difference between the x and y between two consecutive points in the data. Multiplying these points with the sampling frequency and multiplying with the size of each pixel we obtain the velocity in mm/s. An algorithm which detects cases of saccades was used to detect fixations (Matlab Script 3). Whenever the velocity of the eye was under a defined speed threshold, we consider the eye in fixation or scanning mode. The velocity threshold was set to  $2000\text{mm/s} = 26.5\text{ degrees/s}$  [Salvucci & Goldberg, 2000] which seemed to work for this setup and was. On top of that, the location of the fixation had to be the same for 100ms consecutively [Nystrom]. In figure 20 is an example of a typical velocity plot, and the threshold that was set.

```

category_view = strings(length(x),1);
category_view(:) = 'Else';

for i=1:length(x)
    if x(i)==0
        category_view(i) = {' '};
    elseif isnan(x(i)) && isnan(y(i))
        category_view(i) = {'Blink'};
    elseif y(i)>1080 || x(i)>1920 || y(i)<0 || x(i)<0
        category_view(i)={'Off-screen'};
    elseif x(i)>1492 && x(i)<1740 && y(i)>742 && y(i)<942
        category_view(i) = {'Right Mirror'};
    elseif inpolygon(x(i),y(i),x_dashboard,y_dashboard)==1
        category_view(i) = {'Dashboard'};
    elseif x(i)>34 && x(i)<429 && y(i)>741 && y(i)<939
        category_view(i) = {'Left Mirror'};
    elseif x(i)>1025 && x(i)<1665 && y(i)>51 && y(i)<231
        category_view(i) = {'Centre Mirror'};
    elseif inpolygon(x(i),y(i),x_road_centre,y_road_centre)==1
        category_view(i) = {'Road Centre'};
    elseif inpolygon(x(i),y(i),x_road_left,y_road_left)==1
        category_view(i) = {'Road Left'};
    elseif inpolygon(x(i),y(i),x_road_right,y_road_right)==1
        category_view(i) = {'Road Right'};
    elseif inpolygon(x(i),y(i),x_air,y_air)==1
        category_view(i) = {'Air'};
    end
end

cat_view = categorical(category_view);

```

*Matlab Script 2. Creating a categorical view of all data points (x,y) of participants' eye coordinates*

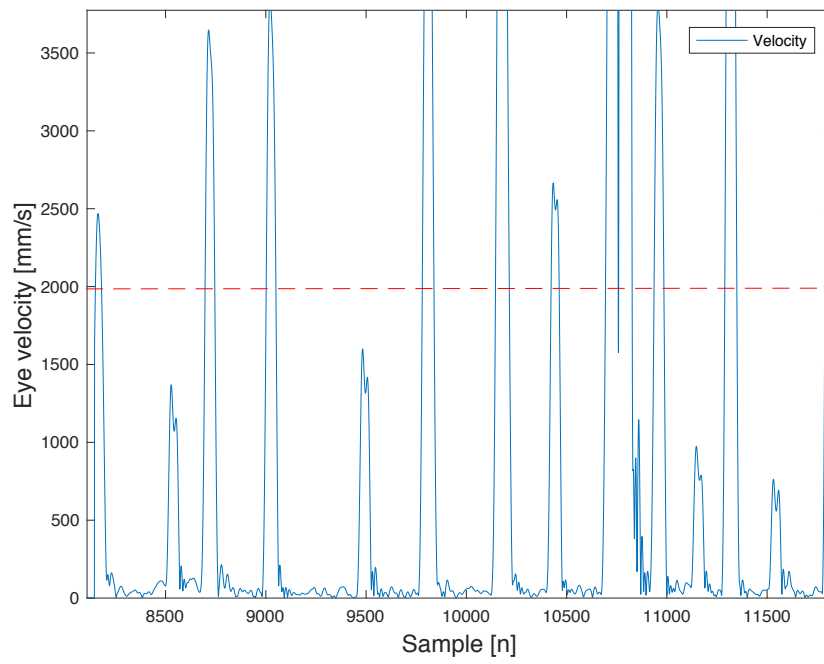


Figure 15. Velocity of the eye for each sample. The red line indicates the velocity threshold set for detecting saccades.

```

%% Find first fixation after visual/audio stimuli
x((x==0))=NaN;
y((y==0))=NaN;

pxtomm = 0.264583333;
fs = 2000;
ddistance = sqrt(xdot.^2 + ydot.^2)*fs*pxtomm;
sac_th=2000; % minimum fixation duration between 80 - 150 ms Nystrom and
Holmqvist, 2010 (I choose 100 ms)

DT=NaN(45,1);
DT_location=strings(45,1);

for i=1:45 %45 trials
    counter=0; %start the counter on zero
    for j=start_video(i,1):(start_video(i,2)-1) %for beginning of trial till
end of trial
        if cat_view(j,1)==cat_view(j+1,1) && cat_view(j,1)~='Dashboard' &&
ddistance(j)<sac_th %find where two elements have the same category excluding
the startn dashboard
            counter = counter+1; %if its true count up
        else
            counter = 0;
        end
        if counter==200 %if the counter hits 200 = 100ms than it is a
fixation in that area
            DT(i,1)= j-start_video(i,1); %Detetion time is the indice j where
we are in the loop minus the beginning of the loop.
            DT_location(i,:) = cat_view(j);
            break
        end
    end
end
end

```

*Matlab Script 3. Determining when a fixation is established by calculating the velocity of the eye and using a threshold.*

#### F.4 Creating table of all variables

The situations and responses and response times are readily saved in result files after the experiment. And in F.3 is shown how the first fixation time, and first fixation location are obtained. For further analysis a loop is made in matlab to loop through all the relevant participants and creates a structure for all participants together (Matlab script 4). Then data was sorted, using parameters such as: (1) Valid/Invalid (2) Removing directions of videos (3) Averaging over repeated trials, and saved in tables for further analysis (Matlab script 5).

```

%% Add all variables in one vector/table for group analysis
clear
clc
close all

direction_total=strings((45*36),1);
video_length_total=NaN((45*36),1);
compat_total =NaN((45*36),1);
validity_total =NaN((45*36),1);
DT_total =NaN((45*36),1);
DT_location_total =strings((45*36),1);
RT_total =NaN((45*36),1);
answer_total =strings((45*36),1);
part_number_total = NaN((45*36),1);

for i=1:36
    part_nr = i
    if i == 21 || i == 22
        continue
    end

    if i==9 || i==15 || i==19 || i==1 || i==5 || i==24
        [compat, video_length, direction, validity, answer, RT]=
get_results(part_nr);
        DT_total((i*45-44):(i*45),1) = NaN(45,1);
        DT_location_total((i*45-44):(i*45),1) = NaN(45,1);
    else
        [direction,video_length,compat,validity,DT,DT_location,RT,answer] =
funcanalyze(part_nr);
        DT_total((i*45-44):(i*45),1) = DT;
        DT_location_total((i*45-44):(i*45),1) = DT_location;
    end

    direction_total((i*45-44):(i*45),1) = direction;
    video_length_total((i*45-44):(i*45),1) = video_length;
    compat_total((i*45-44):(i*45),1) = compat;
    validity_total((i*45-44):(i*45),1) = validity;
    RT_total((i*45-44):(i*45),1) = RT;
    answer_total((i*45-44):(i*45),1) = answer;
    part_number_total((i*45-44):(i*45),1) = ones(45,1)*i;
end

if i==9 || i==15 || i==19 || i==1 || i==5 || i==21 || i==22 || i==24
[compat, video_length, direction, validity, answer, RT]=
get_results(part_number);
end
%% Total
allvars =
struct('direction',direction_total,'length',video_length_total,'compat',comp
at_total,'TTF',DT_total,'TTF_loc',DT_location_total,'RT',RT_total,'validit
y',validity_total,'answer',answer_total);

```

Matlab Script 4. Looping through all participants, extracting the variables and saving in a structure.



```

% This script sorts the allvars.mat to sortrows in an useful order
% Then means values for each condition/trial type
% Each participant has 18 types of trials now excluding invalid cues:
% 'GO', 'Danger', 'beep'
% Left and Right
% 1,3,6 seconds

clear
clc

load('allvars.mat')
removedirections=true; %turn off for full data

alldata =
table(allvars.partnr,allvars.validity,allvars.compat,allvars.direction,allvars.length,allvars.TTFF,allvars.TTFF_loc,allvars.RT,allvars.answer_direction,allvars.answer);
alldata.Properties.VariableNames =
{'Part_nr','Validity','Compatibility','Direction','Length','TTFF','TTFF_location','RT','Answer_Direction','Answer'};
alldata.TTFF = alldata.TTFF/2; %convert to ms

%Remove directions
if removedirections==true
    alldata.Direction = [];
end

alldata2 = sortrows(alldata);

alldata3 = alldata2 ;

alldata3(alldata2.Validity==0,:) = []; %Remove wrong audio trials
alldata3(alldata3.Part_nr==21,:) = []; %Remove part21
alldata3(alldata3.Part_nr==22,:) = []; %remove part22

alldata3.RT(alldata3.RT<0) = NaN; %Set no-reactions to NaN

sumTTFF = 0;
sumRT = 0;
sumCR = 0;
sumIR = 0;
len = 0;
counter = 1;
nancounter_RT = 0;
nancounter_TTFF = 0;

% Loop to find same trials and add them together to 1 mean
for i=1:length(alldata3.Length)
    newLen = alldata3.Length(i);
    RT_add = alldata3.RT(i);
    TTFF_add = alldata3.TTFF(i);

    if isnan(alldata3.RT(i)) %If datapoint in RT(i) = NaN means no reaction:
        nancounter_RT = nancounter_RT+1; %Add to counter for calculating mean
        RT_add=0; % Set alldata3.RT(i) to zero for adding zero to sum
    end
end

```

```

if isnan(alldata3.TTFF(i)) %Same for TTFF
    nancounter_TTFF = nancounter_TTFF+1;
    TTFF_add=0;
end

if i==1 %First step is always the same
    sumRT = sumRT + RT_add;
    sumTTFF = sumTTFF + TTFF_add;
    if alldata3.Answer(i) == "'CORRECT'"
        sumCR = sumCR +1;
    end

    len = newLen;
    continue
end

if newLen == len % If the next length == the current length :
    sumRT = sumRT + RT_add; %Add to the sum RT
    sumTTFF = sumTTFF + TTFF_add; %Add to the sum TTFF
    counter = counter+1; %Counter for calculating mean
    if alldata3.Answer(i) == "'CORRECT'"
        sumCR = sumCR +1;
    end

else
    meanRT(i,:) = sumRT./(counter-nancounter_RT); %If next element is not
equal calculate mean RT
    if meanRT(i,)==0 || meanRT(i,)==Inf
        meanRT(i,)=NaN;
    end
    meanTTFF(i,:) = sumTTFF./(counter-nancounter_TTFF);
    if meanTTFF(i,)==0 || meanTTFF(i,)==Inf
        meanTTFF(i,)=NaN;
    end
    meanCR(i,:) = sumCR/counter;
    sumRT = RT_add; %Start sum over with value of new element
    sumTTFF = TTFF_add;
    sumCR = 1;
    if alldata3.Answer(i) == "'CORRECT'"
        sumCR = 1;
    else
        sumCR = 0;
    end

    len = newLen; % New length
    counter = 1; % Reset counter to 1
    nancounter_RT = 0;
    nancounter_TTFF = 0;
end
if i==length(alldata3.Length(:)) %Last elements
    meanRT(i,:) = sumRT./counter;
    meanTTFF(i,:) = sumTTFF./counter;
    meanCR(i,:) = sumCR./counter;
end
end

% Remove dummy indices
%RT_nans(RT_nans==0)=[ ];
%meanRT(RT_nans) = NaN;
remove_dummys = meanRT==0;
meanRT(remove_dummys)= [ ];
meanCR(remove_dummys)= [ ];

```

```

meanTTFF(remove_dummys)=[ ];

%% Make new table with each conditions once
partnr_temp = ones(18,1);
compat_temp = [zeros(6,1);ones(6,1);2*ones(6,1)];
leftright_temp = {'left' ; 'left' ; 'left' ; 'right';'right';'right'};
direction_temp =
[string(leftright_temp);string(leftright_temp);string(leftright_temp)];
length_temp = [1000;3000;6000];
length_temp = [length_temp;length_temp;length_temp];
length_temp = [length_temp; length_temp];

partnr = [];
compat = [];
direction = [];
length_trial = [];

for i=1:34
    partnr= [partnr; i*partnr_temp];
    compat = [compat; compat_temp];
    direction = [direction; direction_temp];
    length_trial = [length_trial; length_temp];
end

if removedirections==false
T2 = table(partnr,compat,direction,length_trial,meanTTFF,meanRT);
T2.Properties.VariableNames =
{'Part_nr','Compatibility','Direction','Length','TTFF','RT'};
T2.CR = meanCR;
end

if removedirections==true
partnr_temp = ones(9,1);
compat_temp = [zeros(3,1);ones(3,1);2*ones(3,1)];
length_temp = [1000;3000;6000];
length_temp = [length_temp;length_temp;length_temp];

partnr = [];
compat = [];
direction = [];
length_trial = [];

    for i=1:34
        partnr= [partnr; i*partnr_temp];
        compat = [compat; compat_temp];
        direction = [direction; direction_temp];
        length_trial = [length_trial; length_temp];
    end

T3 = table(partnr,compat,length_trial,meanTTFF,meanRT);
T3.Properties.VariableNames =
{'Part_nr','Compatibility','Length','TTFF','RT'};

end

T3.CR = meanCR;

```

## **F.6 Statistical tools**

Matlab statistical tools were used to determine significance of effects. Function 'anova2', which is a repeated measure anova was used for the response times, and correctness. Post-hoc tests were done using the matlab function 'multcompare', multcompare uses 'Tukey-Kramer' as a default for pair to pair comparisons. Significance levels were adjusted according to the Bonferroni correction method.

For fixation time a repeated measure anova did not work, because there were too many missing fixations in the data. Therefore, the function 'anovan' was used to compare between groups.

```

%% RT
%Anova's for RT for audio conditions and video lengths
% format short g
for i=1:3
    %[p,~,stats]=anova2(Xr,1,'off'); % two-way ANOVA
    [p_RT,~,stats]=anova2(RTs(:,[i,i+3,i+6]),1,'on'); % two-way ANOVA
    cp=multcompare(stats,'display','off'); % post-hoc comparison of ranks
    disp('RTs comparison for 1,3,6 second')
    stats
    cp
end

%% CR
%Anova's for RT for audio conditions and video lengths
format short g

for i=1:3
    %[p,~,stats]=anova2(Xr,1,'off'); % two-way ANOVA
    [p_CR,~,stats]=anova2(CRs(:,[i,i+3,i+6]),1,'on'); % two-way ANOVA
    cp=multcompare(stats,'display','off'); % post-hoc comparison of ranks
    disp('CRs comparison for 1,3,6 second')
    cp
end

%% FT

for i=1:3
    %[p,~,stats]=anova1(Xr,1,'off'); % two-way ANOVA
    [p_FT,~,stats]=anovan(FTs(1:26,[i,i+3,i+6]),1,'on'); % two-way ANOVA
    cp=multcompare(stats,'display','off'); % post-hoc comparison of ranks
    disp('FTs comparison for 1,3,6 second')
    cp
end

```

## Appendix G - Literature report

# What is the effect of spatially located auditory feedback on driver's attention?

Jimmy Hu, Joost C.F. de Winter, David A. Abbink

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

**Abstract**—Advanced driver assistance systems are introduced to the market to assist drivers in driving safely. A recommended feedback modality for these systems is the auditory modality, because drivers are already heavily engaged in processing visual information. It is well established that auditory warnings are useful for redirecting attention. However, it is unclear whether spatially located auditory feedback could prove useful in driving. The goal of this literature study is to examine the effects of spatially located auditory feedback on driver's attention.

Results of the literature review show that spatially located audio during driving is not unequivocally useful cue for drivers in a lane-keeping task. However, for urgent situations such as impending collisions, spatially located audio has been found to improve reaction times. Tracking of localized sound in the absence of visuals is generally poor and inaccurate; for frequency tracking the performance is significantly better. Overall, more research is needed in realistic scenarios. Furthermore, how spatial auditory feedback affects visual attention effects in a topic that deserves further investigation.

**Keywords**—auditory feedback, advanced driver support system, spatial sound, cross-modality

## I. INTRODUCTION

Traveling by car has become the primary way of transport. In 2015 approximately 50% of the people in the Netherlands above 18 own a car (Centraal Bureau van Statistiek 2015), and worldwide billions of people are engaged in driving. Driving is a visually demanding task (Groeger 2000), and considering the large number of road users, the driving task is also mentally demanding. Over a million people a year are victim of a fatal crash, and millions more are injured [World Health Organization, 2016].

During driving a high SA (situational awareness) is required in order to drive safely. SA can be described in terms of 3 levels: 1. perception, 2. comprehension and 3. projection (Endsley 1988). In driving this is: taking in visual information, understanding this visual information, and projecting one's state into the future. Perception is the basis of SA and therefore is a critical component of safe driving.

The main cause of most accidents comes down to insufficient driver SA. This is evidenced from accidents, many of which are related to driver distraction, but also driver fatigue and blind spots. Especially with the rise of technology, drivers get bombarded with information unrelated to driving,

such as in entertainment systems, mobile phones and navigation systems. And also driving-related systems such as ACC (Automatic Cruise Control), CWS (Collision Warning Systems), LDS (Lane departure systems), SWS (Speed Warning Systems) provide information to the driver. These ADAS can operate in different modalities (audio, tactile or visual).

ADAS are often evaluated in driving simulators by measures such as: reaction times, lane-keeping performance, safe-driving performance, and self-reported usefulness and satisfaction. While for visual display the effect on visual attention is predictable (van Leeuwen et al. 2011), the role of auditory and tactile feedback on visual attention is not clear. Since the visual modality is the main information source for driving in traffic, the auditory modality is a recommended modality for providing feedback to drivers.

The auditory channel has the benefit of always being 'open', which means that sounds are perceived regardless of the orientation of the head, body, or eyes (Sanders and McCormick 1987). Also, the human has the ability to focus attention on one auditory source and filtering out other sources (Cocktail party effect, Shinn-Cunningham 2008). And the bilateral alignment of the ears allows the human to judge the location of the sound. Principles of cross-modality theory (Driver and Spence 1998) and auditory selective attention (Woods et al. 2001) suggest that multiple sensory sources can either improve or deteriorate information processing of the human. In the future we expect to see more, and more complex ADASs providing feedback in the auditory modality with the aim to grab the attention of the driver (Nees and Walker 2011); therefore the goal of this literature study is to answer the question: What is the effect of spatially located auditory feedback on drivers attention? This paper discusses human ability and limits in perceiving spatial audio in a non-driving scenarios first.

Secondly, the effects of spatial auditory feedback in driving situations are examined. Finally, the implications of the findings are reviewed and recommendations for future work are given.

## II. METHOD

A global search was done in Google Scholar using the combination of search terms and search operators. The search terms were combinations of: audio, auditory, feedback, display, interface, driving, driver. A more in-depth literature search was done on the spatial audio and driver visual attention by

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

using the keywords: directional audio, spatial audio, spatial cuing combined with driving. All available documents that were published in scientific journals or conference proceedings were screened based on the title. The selected literature abstract had to comply with the criteria of being driving related, and feedback had to be delivered through the auditory modality. Additional literature was retrieved from relevant references from the reference lists of the included articles.

### III. RESULTS

#### A. Human perception of spatial audio

1) *Physiology of auditory perception*: The capacity of the human sound channel is relatively limited compared to other animals as dogs and bats. The frequency range of humans is approximately 20Hz–20KHz, which decreases with age. The loudness of a sound is determined by its frequency and intensity (Equal-loudness contour). The perception of the characteristics of sound is dependent on the shape of the head, ears and even position of the body (Terrence et al. 2005). How humans rate pleasantness of a particular sound can differ between persons (Verbist et al. 2009; Jenkins et al. 2007).

Sound arrives at difference moments at each of the ears. This time difference provides the human information of the spatial location of the sound source. On top of that, the character of the sound changes while travelling through space, and because of the expanding character of sound the intensity decreases in space. These two cues are known as ITD (inter-aural time difference) and ILD (inter-aural level difference), and are the most prominent cues in sound localization (Doll et al. 1986).

2) *HRTF (Head Related Transfer Function)*: Binaural audio delivered through headphones are heard as if it were inside the head, resulting in that the sound was localized somewhere between ears, hence binaural audio solely is sufficient in creating a spatial audio illusion (see Mills 1972). Therefore, Bateau et al. (1965) recorded the sound at the ear canal entrance using a artificial ear and created the HRTF for different angles. Nowadays, spatially located sounds can be synthesized using simulated HRTF's (NASA Slab software).

#### B. Sound localization accuracy

Literature points out that participants' localization accuracy is highly dependent on the type of sound, temporal characteristics, and how the sound is delivered (speaker and headphone quality). Verbist et al. (2009) states that broadband signals are better located than single sinusoids, because the former contain more information about its apparent position. Also a better localization performance can be obtained for discontinuous sounds (see Bellotti et al. 2002). A series of experiments are performed were performed by Bellotti et al. (2002). In the first experiment, 12 different sound directions were tested on participants sitting in a car using 4 speakers. The participants had two seconds to choose one of twelve locations given on a paper. The results showed an accuracy of 51% correct localization for untrained subjects, and 75% accuracy for trained subjects. The second experiment tested the accuracy for a smaller resolution of 22.5° of one quadrant in the azimuth. A 48% correctness for

untrained subjects, and 67% correctness for trained subjects were observed. Of the 33% errors of trained subjects, 27% were errors of only 22.5° off. Fitch et al. (2007) performed a similar experiment in an in-traffic study. Participants had to verbally and quickly report the alert direction. The reported correctness was 30%, while a haptic seat scored a correctness of 84%. Similar results were obtained by Terrence et al. (2005). A more extensive study on sound localization has been done (Populin 2008). In this study eye-gaze was used as pointer, an approach which preserves the natural link between perception and action. 32 speakers were installed in a sound-proof chamber. Participants had to stare to the location of 3D sound without visual reference, and the error was measured between the two positions. The average angular error for the straight ahead target was 15°, while for (Carlile et al. 1997) this was only 5-6%. The difference is that Carlile used visual references and trained subjects.

#### C. Audiovisual cross-modality

Fundamental sound localization tasks are concerned with the measurement of performance of a single alone. In real life situations such as driving, there is always presence of visual information in addition to the auditory information. The additional effect of spatial auditory information on a visual search task is substantial. Reported improvements were 175-1200 ms in detection time in a search and forced-choice paradigm (Perrot et al. 1990), and an increase in detection time under a complex visually distracted environment (Perrot et al. 1991). Moreover, Spence (1998) experimented with cross-modality and suggested that preattentive cross-modal integration can, in some cases, produce helpful illusions that increase the efficiency of selective attention in complex scenes. On top of that, the neural circuitry underlying spatial auditory attention has been identified using fMRI's (see Wu et al. 2007).

#### D. Spatial warnings for lane departure events

Lane departures are common errors caused by driver distraction or drowsiness. Many roads contain rumble strips on the sides, which service the purpose to alert the driver through a combination of auditory and tactile feedback. For lane departure warnings the sound of the rumble strip has been artificially reproduced as a warning signals on the side of departure, which may be an effective approach because it is a familiar cue for many drivers (Rossmeier et al. 2005; Rimin-Doering et al. 2005). In one study the effectiveness of this warning signal was tested on drowsy drivers in a driving simulator (Rimin-Doering et al. 2005), the warnings avoided 85% of the potential lane departures. In another study the same warning signal was used. Participants were not informed about the meaning of the directional audio initially. After participants learned the meaning of the directional audio feedback, they produced a faster steering reaction response compared to the naive participants. (Rossmeier et al. 2005). Additionally, a one-level warning (only rumble strip), was compared to a two-level warning (rumble strip and bell noise). Surprisingly, the one-level warning resulted in significantly faster reactions presumably as a result of its simplicity (Rossmeier et al. 2005). However, in another study

where mono non-directional sounds and stereo directional sounds were tested (Suzuki and Jansson 2005), results showed no statistically significant difference in reaction times for the sounds. An explanation could be that the participants first look for visual confirmation before actually implanting a response action (Wang et al. 2007).

#### *E. Spatial warning for collision events*

Collision warnings systems (CWSs) are nowadays mostly used in longitudinal control, and often combined with adaptive cruise control (ACC; Lee et al. 2006; Tivesten et al. 2015; Muhrer et al. 2012). An important question in automated driving is how much the involvement in non-driving tasks is, and how this impacts safe driving. Muhrer et al. (2012) concluded that Forward collision warning systems (FCWSs) are effective in reducing reaction times, and that a combination of FCW and ACC causes even shorter reaction times. Furthermore, the systems were not found to increase the involvement in non-driving tasks. Furthermore, Tivesten et al. (2015) showed that most of the drivers already had their eyes on the road before the warning set in, and that the ACC's deceleration provided the driver with gaze-orienting cue (Morando et al. 2016).

Future CWSs may integrate warnings that convey threat direction. For example, CWSs are tested for providing side collision compatible warnings (Ho and Spence 2009; Wang and Proctor 2003; Wang et al. 2007), front- rear compatible warnings (Ho and Spence 2005), and 360° warnings (Jenkins et al. 2007) (Fitch et al. 2007). Stimulus reaction tasks that are spatially compatible are known to improve reaction speed (Fitts and Posner 1967), which suggests that direction compatible warnings reduces steering reaction time. Wang et al. (2003) tested this hypothesis in a driving context by using a steering wheel as reaction input, finding that spatially compatible auditory feedback (i.e. steering in the direction of the tone) led to faster reaction times than steering away from it (collision warning). The same test has been done in a real driving simulator for collisions (Wang et al. 2007). However, this study found no significant difference in the compatibility options, because the participants withheld their response until they perceived the car visually.

Ho & Spence (2005) tested the effectiveness in a more complex stimulus-response task for front and rear collisions. Participants had to react to the signal, and discriminate a collision from a false alarm by braking or accelerating. In the first experiment a non-spatial audio cue was tested to set the baseline for the upcoming experiments. In the second and third experiments a spatial auditory cue with 50 % and 80 % validity were tested. The 50% valid cue resulted in larger reaction times, presumably because the cue contained redundant information causing extra process time (Ho and Spence 2005). For the 80 % valid cue, significant benefits were found for reaction time and correct responses to compatible cues, showing a cross-modal link facilitation. In Experiments 4 and 5 of the research by Ho and Spence, (spatial) verbal auditory cues were tested. The words 'front' and 'back' were used and presented from the center or from the front and back. Non-spatial verbal cues provided significant benefits in reaction time, but also yielded more errors. Spatial verbal cues had again significant benefits over the non-spatial cues, reducing both reaction time and errors.

Another spatial property is depth. Cognitive neuroscience research suggests that the presentation of peripersonal warning signals, that is, stimuli presented from close to the body of a driver, may be a particularly effective method to alert the driver (Ho and Spence 2009). In an experiment the effectiveness of these warnings were tested in a setting of close speakers. Participants were presented with 'close rear' auditory warnings, 'far front' auditory warnings, as well as vibrotactile warnings (Ho and Spence 2009). Results showed the shortest reaction time (i.e., head turning responses in a visual discrimination task and a braking response task) for the close peripersonal warnings.

Jenkins (2007) proposed an interface that provides users with information of position of other road users around the car. The system delivered a spatial compatible warning according to impending collisions. Different visual and auditory displays were tested, and reaction times were measured for impending collisions. Their results showed a faster reaction time for auditory warnings over visual warnings. For the auditory warnings the smoke alarm sound and car horn sound were subjectively rated as being the most suitable as a warning.

#### *F. Spatial auditory guidance*

Guiding the driver using auditory feedback may be useful in cases where visual feedback is lacking, for example in fog, rain, or darkness. A system for lane-keeping which provided spatial auditory feedback of the position on the road had been tested with and without absence of visuals in a driving simulator (Verbist et al. 2009). The spatial audio combined with visuals did not provide significant benefit to the lane-keeping performance. Furthermore, frequency domain analyses of the human controller showed that human behave like a pure gain with a time delay, suggesting that tracking spatial audio is limited for the human ear. Frequency based mapping seemed to be better trackable for the human in a guided non-visual lane keeping task (Verbist et al. 2009; Woods et al. 2001; Bazilinskyy and de Winter 2016)

#### *G. Spatial auditory for Situational Awareness*

An audio-visual display providing real-time sonification and visualization of the speed and direction of an approaching car on intersections was examined (Houtenbos et al. 2016). The display contained an auditory part, which provided beeps from left and right with beeping frequency depending on vehicle speed. The visual part comprised flashing lights on the dashboard of the car, showing the positions. The participants rated auditory feedback as more useful and pleasant than the visual feedback, but equally satisfactory. The reason for this superiority of auditory above visual display is the placement on the dashboard of the car. Overall the display led to greater traffic efficiency, while not reducing safety.

## IV. DISCUSSION

The aim of this literature search was to answer the question: what is the effect of spatially located auditory feedback on driver attention? To answer this question a literature search



was performed on the topic of (spatial) auditory feedback in driving. First, the capabilities of the human ear to localize sound was examined. Humans are capable of determining sound location, but the accuracy varies between studies. The difference is possibly caused by the difference in experimental design and external factors. Sound characteristics which seem to influence the accuracy are: the frequency content of the sound, which should be broadband (Rayleigh and Press 1907; Verbist et al. 2009), and the temporal discontinuity of the sound (Bellotti et al., 2002). Furthermore, the number of speakers seem to be of influence on the localization accuracy (Populin 2008). While most studies used a static environment, poorer results can be expected for a realistic driving task. For example in a study done by Fitch et al. (2007); however the results of this study for auditory feedback may be affected by the task, since subjects had to verbally report the direction, which is in the same modality of the stimulus.

Spatial warnings for lane departure system may not be useful, because the drivers first look up in order to visually assess the situation (Wang et al. 2007). Moreover, a lane departure is most commonly a result of distraction or drowsiness (Rimin-Doering et al. 2005). Therefore the purpose of the warning is mainly to redirect the attention to the road in general (Jenkins et al. 2007). The side of departure can then easily be seen from environmental cues, and so does not need a specific gaze location for proper reaction.

Spatial warnings can be beneficial for collision warning systems. In both driving and non-driving situations, participants' reaction times and discrimination times were shorter with spatial auditory feedback compared to no feedback, visual feedback, or unidirectional feedback (Ho and Spence 2005; Wang and Proctor 2003; Jenkins et al. 2007).

Many studies have used reaction times as a measurement for effectiveness. Reaction times may be important in urgent situations, but fast reactions may not be relevant to many realistic scenarios, because selecting an appropriate action is more important than a fast reaction. In driving simulator studies (Suzuki and Jansson 2005; Wang et al. 2007) the reaction times were slower compared to stimulus-response studies (Ho and Spence 2005; Ho and Spence 2009; Jenkins et al. 2007). Therefore it is important to consider whether the goal is to provoke a fast steering reaction, or provide the user with an auditory cue to direct visual attention (Rossmeier et al. 2005). In order to make this distinction, a discrimination should be made between evoking exogenous reaction, endogenous reaction, or both (Driver and Spence 1998) (Ho and Spence 2005). Exogenous reactions are stimulus driven, such as reflexes. Endogenous reactions are informative, for example verbal cues. Therefore, exogenous reactions are more useful for fast reaction. However, exogenous ... should be evoked less frequently because of potential nuisance and loss of effectiveness. Endogenous reactions are less intense, but may be harder to implement, because of masking and salience.

In order to study the effectiveness of spatial auditory feedback, it is recommended to also investigate the visual gaze besides reaction time. The visual gaze indicates whether there is a facilitation in visual search, how the user uses the spatial feedback, and also whether the reaction was endogenous or

exogenous. Additional points for consideration for practical effectiveness of the feedback are the trade-off between nuisance and urgency (Baldwin and Lewis 2003), frequency of warnings, and false alarm rates (Maltz and Shinar 2004).

## REFERENCES

- C.L. Baldwin and B. A. Lewis. Perceived urgency mapping across modalities within a driving context. *Applied ergonomics*, 2003.
- D.W. Batteau, R. Plante, R. Spencer, and W. Lyle. Localization of sound: Part 5. auditory perception. *China Lake, CA: U.S. Naval Ordinance Test Station*, 1965.
- P. Bazilinskyy and J. de Winter. Blind driving by means of auditory feedback. 2016.
- F. Bellotti, R. Berta, A. D. Gloria, and M. Margarone. *Using 3D Sound to Improve the Effectiveness of the Advanced Driver Assistance Systems*. Springer-Verlag London Ltd, 2002.
- S. Carlile, P. Leong, and S. Hyams. The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, pages 179–196, 1997.
- T.J. Doll, J. M. Gerth, W. R. Engelman, and D. J. Folds. *Development of simulated directional audio for cockpit applications*. 'Wright-Patterson air force base, Ohio 45433-6573, 5285 Port Royal Road, Springfield, Virginia, 1986.
- J. Driver and C. Spence. Cross-modal links in spatial attention. *Phil. Trans. R. Soc. Lond. B*, pages 1320–1331, 1998.
- M. Endsley. Situation awareness global assessment technique (sagat). *Paper Presented at the Aerospace and Electronics Conference, NAECON 1988., Proceedings of the IEEE 1988 National*, 1988.
- G. M. Fitch, R. J. Kiefer, J. M. Hankey, and B. M. Kleiner. Toward developing an approach for alerting drivers to the direction of a crash threat. *Human Factors*, Vol. 49 No. 4, pages 710–720, 2007. doi: 10.1518/001872007X215782.
- P. M. Fitts and M. I. Posner. *Human Performance*. Brooks and Cole Publishing, A division of Wadsworth Publishing Company Inc., 1967.
- J. Groeger. Understanding driving: Applying cognitive psychology to a complex everyday task. *Hove, East Sussex: Psychology Press*, 2000.
- C. Ho and C. Spence. Assessing the effectiveness of various auditory cues in capturing a drivers visual attention. *Journal of Experimental Psychology: Applied*, Vol 11. No. 3, pages 157–174, 2005. doi: 10.1037/1076-898X.11.3.157.
- C. Ho and C. Spence. Using peripersonal warning signal to orient a driver's gaze. *HUMAN FACTORS*, Vol. 51, No. 4, pages 539–556, 2009. doi: 10.1177/0018720809341735.
- M. Houtenbos, J. de Winter, A. H. c, P. Wieringa, and M. Hagenzieker. Concurrent audio-visual feedback for supporting drivers at intersections: A study using two linked driving simulators. *Applied Ergonomics* 60, pages 30–42, 2016.
- D. P. Jenkins, N. Stanton, G. H. Walker, and M. S. Young. A new approach to designing lateral collision warning systems. *Int. J. Vehicle Design* Vol. 45, No. 3, pages 379–396, 2007.
- J. D. Lee, D. V. McGehee, T. L. Brown, and D. Marshall. Effects of adaptive cruise control and alert modality

- on driver performance. *Transportation Research Record: Journal of the Transportation Research Board*, No 1980, *Transportation Research Board of the National Academies, Washington, D.C.*, pages 49–56, 2006.
- M. Maltz and D. Shinar. Imperfect in-vehicle collision avoidance warning systems can aid drivers. *Transportation Research Part F Traffic Psychology and Behaviour*, 2004.
- A. W. Mills. Auditory localization. In J. V. Tobias (Ed.) *Foundations of modern auditory Theory*, Vol. II, pages 303–348, 1972.
- A. Morando, T. Victor, and M. Dozza. Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses. *Accident Analysis and Prevention* 97, 2016.
- E. Muhrer, K. Reinprecht, and M. Vollrath. Driving with a partially autonomous forward collision warning system: How do drivers react? *SPECIAL SECTION: Human Factors and Automation in Vehicles*, 2012.
- M. A. Nees and B. N. Walker. Auditory displays for in-vehicle technologies. *Reviews of Human Factors and Ergonomics*, pages 58–100, 2011.
- D. R. Perrot, K. Saberi, K. Brown, and T. Z. Strybel. Auditory psychomotor coordination and visual search performance. *Perception Psychophysics*, pages 214–226, 1990.
- D. R. Perrot, T. Sadralodabai, K. Saberi, and T. Z. Strybel. Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors*, pages 389–400, 1991.
- L. C. Populin. Human sound localization: measurements in untrained, head-unrestrained subjects using gaze as a pointer. *Exp Brain Res*, pages 190: 11–30, 2008. doi: 10.1007/s00221-008-1445-2.
- L. Rayleigh and O. Press. On our perception of sound direction. 1907.
- M. Rimin-Doering, T. Altmueller, U. Ladstaetter, and M. Rossmeier. Effects of lane departure warning on drowsy drivers performance and state in a simulator. *PROCEEDINGS of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2005.
- M. Rossmeier, H.-P. Grabsch, and M. Rimini-Doering. Blind flight: Do auditory lane departure warnings attract attention or actually guide action? *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland*, 2005.
- M. Sanders and E. McCormick. *Human Factors in Engineering Design*. New York, NY: McGrawHill, Reading, Massachusetts, 1987.
- B. G. Shinn-Cunningham. Object-based auditory and visual attention. *Trends Cognitive Science*. 2008 May, pages 182–186, 2008. doi: 10.1016/j.tics.2008.02.003 =.
- K. Suzuki and H. Jansson. An analysis of drivers steering behaviour during auditory or haptic warnings for the designing of lane departure warning system. *Society of Automotive Engineers of Japan, JSAE Review* 24, pages 65–70, 2005.
- P. I. Terrence, J. C. Brill, and R. D. Gilson. Body orientation and the perception of spatial auditory and tactile cues. *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 49th ANNUAL MEETING*, pages 1663–1667, 2005.
- E. Tivesten, A. Morando, and T. Victor. The timecourse of driver visual attention in naturalistic driving with adaptive cruise control and forward collision warning. *4th International Driver Distraction and Inattention Conference, Sydney, New South Wales*, 2015.
- P. van Leeuwen, S. de Groot, R. Happee, and J. de Winter. Effects of concurrent continuous visual feedback on learning the lane keeping task. *PROCEEDINGS of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2011.
- K. Verbist, E. R. Boer, M. Mulder, and M. van Paassen. Car lane-keeping using auditory feedback. 2009.
- D.-Y. D. Wang and R. W. Proctor. Stimulus-response compatibility effects for warning signals and steering responses. *PROCEEDINGS of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2003.
- D.-Y. D. Wang, D. F. Pick, R. W. Proctor, and Y. Ye. Stimulus-response compatibility effects for warning signals and steering responses. *PROCEEDINGS of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2007.
- D. L. Woods, C. Alain, R. Diaz, D. Rhodes, and K. Ogawa. Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, pages 65–74, 2001.
- C.-T. Wu, K. R. D.H. Weissmand, and M. Woldorff. The neural circuitry underlying the executive control of auditory spatial attention. *Brain Research* 1134, pages 187–198, 2007.