AN AUDITORY DISPLAY FOR REMOTE ROAD VEHICLE OPERATION THAT INCREASES AWARENESS AND PRESENCE

Pontus Larsson, Johanna Bergfelt Ramos de Souza, Joel Begnert

Ictech AB,
Lindholmspiren 5A,
417 56 Göteborg, Sweden
pontus.larsson@ictech.se

ABSTRACT

Remote operation, which allows a human operator to support a connected and automated vehicle from a distance, has been proposed as a solution to overcome vehicle automation challenges. In the current paper, the use of sound is being explored as a potential improvement to the current predominantly visual remote operator interfaces. An experiment was conducted aimed to test whether an auditory display providing propulsion sound from the ego-vehicle and spatial, augmented sound of surrounding road users would improve the perception of speed and ego motion, as well as situational awareness and presence. The experiment used a within-group 2x2 full factorial design with propulsion sound and spatial augmented sounds on/off as independent variables. 28 participants took part and drove in a simulated environment with the different types of auditory feedback. It was found that the auditory displays' propulsion sound improved participants' speed regulation performance. Moreover, both the propulsion sound and the augmented sound contributed to the sensation of ego-motion, presence and situational awareness.

1. INTRODUCTION

The activities related to the development of automated road vehicles have been ever-increasing during the past few years. Automated Driving (AD) may bring positive societal effects in terms of reduced environmental impact, improved traffic safety, and more efficient mobility systems [1]. Despite the rapid progress in this field, there are still many challenges to be solved before high level driving automation (SAE's Levels 4-5 [2]) can be realized. Some of the most difficult challenges are related to the handling of scenarios that the AD system is unprepared for. Such scenarios could be approaching construction sites, public events, or traffic accidents. Other difficult AD-related challenges include e.g. sensor or system failure or when the vehicle approaches the limits of its Operational Design Domain (ODD) - e.g. conditions or situations that the AD system is not designed to handle. It has been proposed that remote operation enabling a human remote operator to support a connected and automated vehicle from a distance when unexpected events or failures occur could be a way to overcome these challenges [3]. Furthermore, this type of technology could also allow a remote operator to act as a safety driver during the testing phases of AD vehicles that are not equipped with regular driver controls [3]. Hence, remote operation has the potential to overcome some of the challenges that currently are slowing down the commercialization of automated vehicles.

Remote operation can be divided into the following categories: 1) remote driving, where an operator acts as a normal driver and conducts dynamic driving tasks, 2) remote assistance, where the operator assists automated driving systems with commands such as approving the system to maneuver around an obstacle, and 3) remote supervision, where the operator supervises automated vehicles without giving specific commands to the vehicles [1]. In the current article, we focus primarily on category 1, remote driving, although the concepts and results are more or less applicable to all categories of operation.

However, remote operation of road vehicles comes with its own set of challenges, not only related to the technology involved but also related to the human factors aspects of the operation [4]. For example, lack of situational awareness, increased cognitive load and difficulty to estimate speed and acceleration is something which has been identified as issues when driving remotely [4]. Many problems relate also to the quality and field-of-view of the video feed and other visual aspects [4]. Remote operator interfaces for commercially available road vehicle teleoperation systems are predominantly relying on visual cues, which is not surprising as driving is a highly visual task [5]. Nonetheless, the fact that we use all our senses when experiencing and navigating in the world should encourage the exploration of including other sensory modalities in a remote operating station. The current work has focused primarily on how the use of sound may improve the situation for a remote operator of a road vehicle. Several different reasons for using sound in a remote operator station have been identified within the context of the current research and are discussed below.

Information and warning. Today's cars, trucks and other road vehicles employ a number of different interface sounds that guide or warn the driver on several different levels of urgency to the driver. Examples of sounds that can be presented through a driver-vehicle interface (DVI) are turn signal, parking assistance, lane departure warning and drowsiness warning. Some of these sounds are required to be presented by the DVI to fulfill safety ratings such as Euro NCAP and/or legal requirements [6, 7]. While the legal framework for remote operator stations is yet to be fully defined, it is reasonable to assume that some of the sounds required in normally driven vehicles will also become mandatory in remote operation. Other sounds, such as the turn indicator sound or parking assistance sounds, may be included to enhance the understanding of the vehicle's state and/or improve the operator's performance.

Sense of speed and ego-motion. Ensuring accurate perception of a remote road vehicle's speed is essential for safe operation and avoiding unwanted conditions such as discomfort for passengers, strain on the vehicle, and energy inefficiency. In a remote operator station, where visual cues are limited as they are usually monoscopically mediated through a camera-display system and there is a lack of physical motion, the sound resulting from the ego vehicle's movements may become extra important for the operator to be able to regulate the speed properly. Vision is a critical factor in speed perception, but studies have also shown the influence of in-car noise on perceived speed [8-10]. For example, studies have demonstrated that noise reduction can lead to slower speed judgments [8,9]. In another study, participants were found to drive faster and with a more variable speed profile in the absence of sound feedback [10]. Their ability to maintain correct speed was also much worse at higher speeds [10]. These findings suggest that remote operation stations lacking propulsion sound feedback could lead to operators driving too fast, particularly when no speedometer is visible.

Apart from understanding what speed the remote car is currently traveling at, it may be advantageous for the operator to get a sense of actually moving to further increase his/her awareness of the remote vehicle state. Visual stimuli alone can create such sensations [11] but as previous research has shown, auditory stimuli alone can also create ego-motion sensations as well as contribute to an increased sensation in an auditory-visual display [12, 13]. Also, the use of propulsion sound can increase the sensation of actually moving [14].

Situational awareness. Situational awareness (SA) is a broad term that may be defined in various ways but which is generally believed to be important for operators of vehicles and systems. In teleoperation, it is believed that SA is one of the key aspects to consider in operator interface design. In this context we will use the model by Endsley [15] which states that SA consists of awareness on three levels, each stage being a necessary (but not sufficient) precursor to the next, higher, level: 1) Perception of elements in the environment (e.g. perception of other objects around the car, current speed, etc.), 2) Comprehension of the current situation (understanding of the different elements' significance to the driving mission, e.g. if an object is on a collision path with the vehicle) and 3) Prediction of future status (being able to predict the future status of the elements, e.g. knowing where the car or other objects will be within a relevant timeframe). We have hypothesized within the context of the current research that an auditory display may enhance SA in several different ways. Better speed perception through ego sound presentation is one such SA aspect that we already touched upon that likely will enhance all three levels of SA. Another possibility is that being able to hear the remote environment will allow operators to both identify, localize and predict the trajectory of surrounding road users which most likely also will enhance SA on all levels. Also, being able to hear error states and the general acoustic behavior from/of the ego vehicle is believed to be very beneficial, especially when the error state cannot be reported in other ways by means of sensors or similar mechanisms [4]. Appropriate reproduction of the soundscape, both in terms of its spectral, temporal, and spatial qualities, is however likely crucial to gain this SA enhancement.

Sense of presence. The sensation of presence - the sensation of "being there" - is an often-discussed aspect of the experience of virtual environments [16]. Presence and telepresence - the sense of being in a remote environment - is important to study in relation to remote road vehicle operation since it may be linked to task performance [16]. It is also natural to assume that an operator feeling more present in the remote environment would make him/her act more as if controlling the vehicle directly [17] - which would likely mean a safer operation. The importance of auditory displays in achieving a sense of presence has been emphasized by several researchers [18-20]. Several aspects of sound have been shown to influence presence, such as the inclusion of background sounds which may be particularly important for the sensation of "being part of the environment" [19]. While spatialized sound is in general seem to be important for creating presence [17], the sensation can be further enhanced by means of individualized Head Related Transfer Functions (HRFTs) [20].

In sum, auditory displays have the potential to improve the situation for remote operators of road vehicles in several ways, and in several types of scenarios and use cases. The research presented in the current paper evaluates an auditory display designed to improve situational awareness, speed- and ego motion perception and at the same time improve the sensation of presence. The experiment presented below focuses on what we believe are important components of an auditory display for remote driving, namely propulsion sound and the sound of other road users in the remote environment – hereafter referred to as augmented sound.

2. EXPERIMENT

In the current experiment the aim was to test two main hypotheses regarding the auditory display prototype (which is explained in detail further on):

Hypothesis 1: Propulsion sound and augmented sound both contribute to the sensation of ego motion and the perception of speed.

Hypothesis 2: Augmented sound and propulsion sound contribute positively to situational awareness and presence (but augmented sound to a higher degree).

Situational awareness is likely to be increased on all three SA levels perception, comprehension, and projection: Augmented sound helps the user to understand what the surrounding objects are, where they are (perception) and where they are going (comprehension). Propulsion sounds help the user to understand how fast the remote car is traveling and thus allows for projection, knowing how the surrounding objects are moving in relation to the remote car through the Augmented sounds.

2.1. Design

A within-group 2x2 full factorial design was used with spatial augmented sounds (on/off) and propulsion sound (on/off) as independent variables.

2.2. Participants

28 participants (21 male, 8 female) took part in the experiment. Their mean age was M = 32.5 years (SD = 7.24), and all held driver's licenses for passenger car ("B"). Fourteen of the participants stated that they drove less than 1000 km / year, while twelve drove between 1000 and 3000 km/year and two drove more than 3000 km / year. All but four participants stated that they were playing computer games a few times a month or more often. They all received

a gift certificate for cinema tickets (value 300 SEK) as compensation.

2.3. Materials

An in-house developed driving simulator based on the CARLA OpenSource software [23] was used. The visual display consisted of three curved 27” screens (Samsung Odyssey G5) mounted on a standard desktop as shown in Figure 1. Each screen had a curvature radius of 1 m and the resulting total Field-Of-View (FOV) for all three screens was around 122 degrees. Participants were controlling the virtual car using a Logictech G29 steering wheel and pedals mounted on the desk and on a custom-made footrest respectively. Participants were seated on a regular office chair.

An auditory display prototype software intended for remote operated driving and monitoring developed within the context of the current research called LAVA (Layered Augmented Vehicle Audio), was used to render the sound. The LAVA software is developed in Max8 [21] and uses the Max package Spat developed by Ircam [22] to render spatialized sound over headphones (so-called binaural presentation). It has an Open Sound Control (OSC) interface that allows for receiving positions, speeds and types of surrounding road users, as well as ego vehicle speed. The software can also receive head orientation angles as input via OSC to enable head-tracked binaural rendering in Spat. Each road user - a source - is rendered using the [spat5.binaural~] object (square brackets are commonly used to denote Max 8 objects in general and the “~” denotes objects which perform audio generation or processing). Distance attenuation factor as well as high frequency distance roll-off and Doppler effect can be adjusted. The LAVA software has the capability of synthesizing both realistic ego-sound (propulsion sound) and the sound of surrounding road users such as cars, bicyclists and pedestrians using a combination of additive synthesis and recorded sound files (for more information, see section 2.5 below). The software also allows for receiving microphone feeds from the remote vehicle and downmixing these to binaural sound. Two microphone array types are currently supported: Ambisonics (1st order) and Spaced Array, where the Spaced Array mode can be configured to match different types of array configurations.

The LAVA software was used to render spatial sounds as well as propulsion sound, which was presented to the participants over semi-open headphones (Beyerdynamic DT-880). A head tracker (Supperware Head Tracker 1) was used to track the yaw angle of the participant’s head. The yaw information was used as input to the binaural rendering of the spatial sounds so that sources would appear to be stable when the participant would move his/her head (and also to improve localization). The spatialization was using the Head Related Transfer Functions (HRTFs) of the KEMAR manikin. To enhance the low frequency sensation of the propulsion sound, which was somewhat limited when presented only over the headphones, some vibrotactile feedback was also included. The vibrations were presented via two Buttkicker LFE Mini electrodynamic shakers mounted underneath the chair and on the footrest in front of the pedals. An audio signal based on the low frequency part of the propulsion sound from the LAVA software was used to drive the shakers and were amplified by a Behringer KM750 power amplifier. The experiment setup is shown in Figure 1.

2.4. Scenarios and task

The map Town01 provided by the CARLA software [23] was used as a base for the scenarios. This model consists of a square city with a two-lane street running around the city as well as two lane streets within the city. During the experiment trials, participants drove both on the streets around and within the city, following a predefined route. Navigation arrows were presented in intersections along the route to guide the participants from start to finish in each trial (See Figure 1).

Figure 1: Experiment setup. The arrow seen in the top of the middle screen is an example of one of the navigation instructions provided along the route. The hood of the virtual ego vehicle can be seen in the bottom of the middle- and right screens.

The participants encountered four different scenarios along the route:

1. Car 1: An adversary car would turn out just in front of the participant’s car in an intersection and onto the opposite lane
2. Car 2: Similar to 1 but in another intersection
3. Pedestrian: A pedestrian would run across the road ahead of the participant’s car. The pedestrian was partially occluded by a bus stop when it started running.
4. Cyclist: A cyclist would start biking across the street ahead of the participant’s car. The cyclist was partially occluded by a bus stop when starting cycling.

Participants’ task was to follow the navigation instructions, keep 50 km/h on straights and in general drive as if it was a real car they were controlling.

2.5. Stimuli

Visual stimulus consisted of the Town01 model [23]. A nighttime setting was used, mainly since this made the surrounding road users a bit less visually detectable (which possibly would make the auditory cues more important). The viewpoint was set so that the ego vehicle’s hood could be partially seen (see Figure 1); the reason for this was to increase the ease of handling, especially in the corners where it was otherwise tricky to maneuver the car. In pre-trial training sessions a digital-style speedometer was presented on the mid screen in a head-up display fashion, to allow the participants to learn how 50 km/h would be perceived in each condition. In each following real trial, the speedometer was not shown.

Auditory stimuli consisted of propulsion sound (PS) (including low frequency vibrational feedback as described in section 2.3) of the ego vehicle and augmented sound (AS) representing surrounding road users rendered using the LAVA application. The PS was created by cross fading and pitch shifting between recorded sounds from the interior of an electric car played by seven different [groove~] objects (corresponding to constant speeds between 30-90 km/h). The resulting sound had a fairly noticeable change in character around 50 kph where tonal components of the sound became more audible. The idea behind this design was to give the participants a realistic propulsion sound that still had a noticeable auditory reference point when regulating the car's speed. While the current design only had one such reference point, one could extend the design to include several distinct cues representing different speed thresholds. Ego speed was sent as OSC messages from the driving simulator software to the LAVA application.

The AS consisted of three different source types: other cars, pedestrians, and cyclists. The car sounds were created by combining four sinusoidal tones and filtered pink noise ([cycle~] and [pink~] objects) to simulate engine and road/wind noise respectively. The tonal components were varied in level and frequency depending on car speed and the noise was varied in level and low pass cut off frequency depending on car speed. The pedestrian and bicycle sounds were created by [groove~] objects playing recorded sound files of footsteps and a bicycle drivetrain and bell. These sounds were varied with means of time stretching/ compression with regards to the source speed so that the step and pedal frequency would sound realistic in relation to the source speed. Doppler shift and distance attenuation was added to all sounds before spatializing them using [spat5.binaural~] with KEMAR HRTFs. The signal from the headtracker was used as input to the spatialization to increase realism, localization and reduce "in-head" effects. In total, six sources could be presented at the same time and their location, speed and type (car/bike/pedestrian) were sent as OSC messages from the CARLA simulator to the LAVA software. The six sources currently closest to the observer were selected by the driving simulator software. It should be noted that the levels of all AS and the PS were intentionally adjusted in a way so that all sounds would be equally audible. In other words, e.g. pedestrians (footstep sounds) and bicyclists were much louder and PS much quieter compared to what they would have been in a real situation. The rationale behind this auditory display design was to create a sensation of having "supernormal" hearing, which potentially could allow for detecting hazardous situations much earlier than in a corresponding real situation. Should the auditory display have aimed at completely emulating a real situation, the sound of pedestrians and bicyclists would likely be masked by propulsion noise to great extent, even if the ego car would be driving with its windows rolled down.

2.6. Dependent variables

Dependent variables included speed regulation performance (based on the measures mean and standard deviation of speed as well as the absolute value of the difference between mean speed and the target speed of 50 kph - all derived from log files from the simulator), self-rated ego motion, speed perception, driving performance, situational awareness, and presence. Ego-motion, speed perception and driving performance were rated using three different single-item scales. Regarding presence, the Slater-Usoh-Steed (SUS) presence questionnaire was used [24]. Situational awareness was measured using a set of context-related scales representing how well the participant could localize the position and traveling direction of other road users. The subjective rating scales are further explained in the results section below. In addition, participants' general experience of the experiment and their perception of the sound was acquired through semi-structured interviews after the test.

2.7. Procedure

Participants arrived individually to the lab. First, a thorough introduction to the study, the simulator, and participants' task (keep 50 km/h on straights and drive as if it was a real car they were remotely controlling) was given. The experiment leader then made sure that the participants had understood the instructions and participants were informed that their participation was voluntary and that they were free to withdraw at any time, without giving a reason. Demographic data and the participants' driving- and computer game experience were then collected. The experiment leader then gave a brief repetition of the task and the instructions related to the simulator. After this, four blocks of driving sessions, each containing a practice trial and a real trial were performed by the participants. Each block corresponded to a certain auditory display condition: either no sound, PS on, AS on or both sound types on. The practice part lasted for around 2 minutes and in this part, the speedometer was visible. After the practice trial, the real trial was driven, in which the speedometer was hidden. The real trial lasted for about 5 minutes and was followed by a short break where the participants filled in a questionnaire containing the rating scales. Participants were assigned randomly to different block orders (i.e. orders of auditory display conditions) to avoid order effects. In addition, the order of scenarios was also varied across participants to be able to avoid order effects and compare first-exposure reactions to scenarios across participants. After all blocks and questionnaires had been completed, an interview was conducted with the participants focusing on positive and negative aspects of the general experience, the auditory display in particular and potential improvements. Participants were then debriefed and thanked for their participation.

3. RESULTS

Statistical analyses of data from subjective ratings and objective measurements were performed using the software IBM SPSS Statistics 29.

3.1. Presence

Ratings from the SUS Presence questionnaire scale were submitted to individual 2x2 (Propulsion sound, on/off x Augmented sound, on/off) repeated measures analyses of variance (ANOVA).

For the first item of the scale, "I had a sense of "being there" in the virtual environment", a main effect of propulsion sound was found: $F(1,\ 27) = 39.82$, $p < .001$. Post-hoc test using Bonferroni's adjustment for multiple comparisons showed that the propulsion conditions resulted in statistically significantly higher ratings on this item ($M =$

4.357, *SE* = 0.293) compared to conditions without propulsion sound (*M* = 3.268, *SE* = 0.258), *p* < .001. Similar effects of propulsion sound were found for all other presence scales, indicating that all items were rated higher with the propulsion sound on compared to when it was off. The results are summarized in Table 1 below.

Table 1: Results from analysis of effect of propulsion sound on presence (adjustment for multiple comparisons: Bonferroni)

Item	**Main effect**		**PS off**		**PS on**	
	F (1,27)	*p*	*M*	*SE*	*M*	*SE*
1. I had a sense of “being there” in the virtual environment	39.82	<.001	3.27	0.26	4.36	0.29
2. There were times during the experience when the virtual environment was the reality for me...	41.66	<.001	2.93	0.25	3.96	0.30
3. The virtual environment seems to me to be more like... (images that I saw / somewhere I visited)	24.47	<.001	2.84	0.22	3.79	0.30
4. I had a stronger sense of... (being elsewhere / being in the virtual environment)	22.65	<.001	3.48	0.30	4.70	0.28
5. I think of the virtual environment as a place in a way similar to other places that I've been today...	18.30	<.001	2.66	0.24	3.46	0.28
6. During the experience I often thought that I was really driving in the city...	34.40	<.001	2.68	0.23	3.80	0.31

Table 2: Results from analysis of effect of augmented sound on presence (adjustment for multiple comparisons: Bonferroni)

Item	**Main effect**		**AS off**		**AS on**	
	F (1,27)	*p*	*M*	*SE*	*M*	*SE*
1. I had a sense of “being there” in the virtual environment	9.71	0.004	3.52	0.27	4.11	0.29
2. There were times during the experience when the virtual environment was the reality for me...	17.36	<.001	3.09	0.25	3.80	0.31
3. The virtual environment seems to me to be more like... (images that I saw / somewhere I visited)	*3.34*	*ns*	*3.13*	*0.24*	*3.50*	*0.29*
4. I had a stronger sense of... (being elsewhere / being in the virtual environment)	*2.88*	*ns*	*3.82*	*0.31*	*4.36*	*0.30*
5. I think of the virtual environment as a place in a way similar to other places that I've been today...	7.56	0.011	2.82	0.23	3.30	0.28
6. During the experience I often thought that I was really driving in the city...	17.44	<.001	2.88	0.26	3.61	0.28

The analysis also showed a main effect of augmented sound on the first presence item ratings, albeit not as strong as for the propulsion sound: *F*(1,27) = 9.714, *p* = .004. Post-hoc test using Bonferroni’s adjustment for multiple comparisons showed that the conditions with augmented sound on resulted in statistically significantly higher ratings on this item (*M* = 4.11, *SE* = 0.29) compared to conditions without augmented sound (*M* = 3.52, *SE* = 0.27), *p* = .004. The results, including also the other items are summarized in Table 2. As can be seen, statistically significant main effects of augmented sound were found for scale items 2, 5, 6, while no effects were found for items 3 and 4. Looking at the means however, all items are rated higher for the conditions where the augmented sound was on compared to when it was off.

3.2. Situational awareness

As mentioned previously, we have adopted the SA definition from [15] in the current experiment, and while there is a range of different methods on how to measure SA [25], none of the available methods seemed appropriate for the current experiment. For example, the commonly used SAGAT method [25] would involve pausing the simulation at certain points in time which would affect sensed presence and speed regulation performance. Instead, we chose to use a set of context-adapted SA scales, based on a similar idea for measuring trust in automated vehicles [26, 27], and are intended to measure the levels of SA as defined by [15]. The items that were used were:

1. I was able to localize pedestrians around me
2. I understood in which direction pedestrians were walking
3. I was able to localize cyclists around me
4. I understood in which direction cyclists were going
5. I was able to localize cars around me
6. I understood in which direction cars were driving
7. I knew what was going to happen within the next few seconds

All items were rated on a 7-step scale with “not at all” and “very much” as endpoints.

Ratings from the items relating to SA were submitted to individual 2x2 (Propulsion sound, on/off x Augmented sound, on/off) repeated measures ANOVAs.

Main effects of Augmented sound were found for SA items 1, 3, 5 (relating to the participant’s ability to localize pedestrians, cyclists and cars respectively): Item 1, *F*(1,27) = 13.57, *p* = .001, Item 3: *F*(1,27) = 12.06, *p* = .002. and Item 5, *F*(1,27) = 12.60, *p* = .001. Post-hoc tests using Bonferron’s method to adjust for multiple comparisons gave that for SA Item 1, (localize pedestrians), the augmented sound led to a higher rating of *M* = 5.77 (*SE* = 0.21) compared to conditions without augmented sound where *M* = 4.75 (*SE* = 0.29), *p* = .001. For Item 3 (localize cyclists), a similar effect was found in that augmented sound resulted in higher ratings compared to without augmented sound (*M* = 5.43, *SE* = 0.26, vs. 4.30, *SE* = 0.34, *p* = .002). Finally, for item 5 (localize cars) - the effect was again similar with *M* = 5.97, *SE* = 0.17 for augmented sound conditions vs. *M* = 5.21, *SE* = 0.28, *p* = 0.001).

No effects for the items relating to understanding in which directions other road users were going were found. The closest one to being statistically significant was for item 6 relating to the understanding of other car’s directions; *F*(1,27) = 3.19, *M* = 5.96, *SE* = 0.17 (augmented sound on) vs. *M* =5.52, *SE* = 0.27 (augmented sound off), *p* = .085. Thus, while the augmented sound helped participants to localize other road users, it seemed it did not help as much in informing the road users' direction of travel.

Nonetheless, for SA Item 7, "I knew what was going to happen within the next few seconds", a main effect of augmented sound was found: $F(1,27) = 10.44$, $p = .003$. Post-hoc tests showed that the augmented sound gave rise to higher ratings on this item than when no augmented sound was present: $M = 4.93$, $SE = 0.25$ vs. $M = 4.18$, $SE = 0.28$, $p = .003$ (Bonferroni adjusted).

The analysis also showed a main effect of Propulsion sound on SA items 1 (ability to localize pedestrians) and 5 (ability to localize cars): $F(1,27) = 4.47$, $p = 0.044$ and $F(1,27) = 5.09$, $p = .032$. Post-hoc tests showed that propulsion sound on led to higher ability to localize pedestrians and cars compared to no sound conditions (Item 1: $M = 5.45$, $SE = 0.22$ vs. $M = 5.07$, $SE = .24$, $p = .044$; Item 5: $M = 5.77$, $SE = 0.20$, vs. $M = 5.41$, $SE = 0.2$, $p = .032$).

Effects were also found for SA item 6 ("I understood in which direction cars were driving") and SA item 7 ("I knew what was going to happen within the next few seconds"): $F(1,27) = 7.70$, $p = .010$ and $F(1,27) = 10.40$, $p = .003$. According to the post hoc test, both these items were rated higher when propulsion was on compared to when it was off: $M = 5.92$, $SE = 0.18$, vs. $M = 5.55$, $SE = 0.21$, $p = .010$ and $M = 4.77$, $SE = 0.25$ vs. $M = 4.34$, $SE = 0.24$, $p = .003$ respectively.

3.3. Speed and ego-motion perception, driving performance

Situational awareness can also be assumed to relate to knowing how fast the ego vehicle is going. In the current experiment, this was measured by the subjective scale "While driving, how certain were you of which speed you were keeping?". A main effect of propulsion was found on this item: $F(1,27) = 130.72$, $p < .001$. Post-hoc tests showed that participants thought that they were more certain of which speed they were keeping with the propulsion sound on compared to when it was off; $M = 4.41$, $SE = 0.22$ vs. $M = 2.23$, $SE = 0.19$, $p < .001$ (Bonferroni's method was used to adjust for multiple comparisons).

The analysis showed moreover that the sensation of ego-motion (as measured by the scale "While driving, how much did it feel as if you were actually moving?", rated on a 7-step scale with "not at all" and "very much" as endpoints) was affected by both the propulsion sound as well the augmented sound. The effect of propulsion sound was $F(1,27) = 36.72$, $p < .001$ and the effect of the augmented sound was $F(1,27) = 8.85$, $p = .006$. Post-hoc test showed that the ego motion sensation was significantly higher with propulsion sound on compared to when it was off; $M = 4.43$, $SE = 0.33$ vs. $M = 2.82$, $SE = .29$, $p < .001$. Similarly, augmented sound on led to that the ratings of ego motion were statistically significantly higher compared to when it was off; $M = 3.82$, $SE = 0.29$ vs. $M = 3.43$, $SE = 0.28$, $p = .006$ (Bonferroni's method was used to adjust for multiple comparisons).

Finally, participants also rated their driving performance after each condition on the scale "During the last minutes, I think my driving performance was… extremely poor - extremely good". The analysis showed a main effect of propulsion; $F(1,27) = 48.34$, $p < .001$. Post hoc tests showed that propulsion sound led to higher self-rated driving performance than when no propulsion sound was on; $M = 4.39$, $SE = 0.22$, vs. $M = 3.25$, $SE = 0.22$, $p < .001$ (Adjustment for multiple comparisons: Bonferroni).

3.4. Speed regulation performance (objective)

Participants' speed profiles on two straights where no specific scenarios occurred were extracted from logfiles from the simulator. The means and standard deviations of the extracted speed profiles were then calculated, separately for all participants and conditions. In addition, the deviation between mean speed and the target speed of 50 kph, was also calculated as $deviation = |50 - avg.(speed)|$, separately for all participants and conditions. These values were then submitted to individual 2x2 (Propulsion sound, on/off x Augmented sound, on/off) repeated measures ANOVAs.

A main effect of propulsion sound on mean speed was found: $F(1,27) = 26.31$, $p < .001$. Post-hoc test using Bonferroni's method to adjust for multiple comparison showed that with propulsion sound on, the mean speed was statistically significantly lower than when the sound was off: $M = 50.50$; $SE = 1.04$ vs. $M = 57.08$; $SE = 1.48$, $p > .001$.

A main effect of propulsion sound on deviation was also found: $F(1,27) = 21.67$, $p < .001$. Post-hoc tests showed that propulsion sound on led to a lower deviation from 50 kph compared to when the propulsion sound was off: $M = 4.22$; $SE = 0.71$ vs. $M = 8.96$; $SE = 1.07$ (Bonferroni's method used to adjust for multiple comparisons).

Finally, a main effect of propulsion sound on standard deviation of speed was also found: $F(1,27) = 49.46$, $p < .001$, with post hoc test showing that the standard deviation was significantly lower with propulsion sound on compared to when it was off: $M = 4.83$; $SE = 0.34$, vs. $= 9.24$; $SE = 0.57$, $p < .001$ (Adjustment for multiple comparisons: Bonferroni).

No effects of the augmented sound were found for any of the measures related to speed regulation.

3.5. Participant interviews

To analyze the qualitative data from the post-experiment interviews, a KJ analysis [28] was conducted. This is a method which can be used to create an overview of a large set of complex data. The KJ analysis is conducted by segmenting the data set into smaller entities, and these smaller entities are later grouped based on themes. The groups will then expand into clusters and at a last stage the clusters are given a title. The groups do not exist from the start; they form during the analysis. The KJ analysis uses a bottom-up approach, where details are initially studied and from there, the analysis evolves towards a greater understanding of the collected data. From the interviews performed in the current study, 362 unique quotes were collected and analyzed in this manner. The digital platform Miro was used to sort the data entities. Below a selection of the findings are presented.

The propulsion sound generally received an overall positive response from the participants. It created a more realistic experience of being inside a vehicle, helped with understanding changes in speed and gave a sense of acceleration and deceleration. Moreover, the propulsion sound affected decision making in the sense that it made it easier for the participants to know when to brake when approaching junctions, when to stop accelerating and also helped them in keeping the steady target speed. Participants mentioned that when hearing the propulsion sound, they listened for a certain pitch to know when they had reached 50 kph. One participant said, *"Without sound it felt lonely, made it much harder to know how fast to drive; I was shocked at the junctions that I was driving so fast"*. The propulsion sound was perceived as contributing to a better driving flow and that the sound made some participants drive with more caution. Propulsion sound also had an effect on the spatial experience, and it enhanced the feeling of "being there". One participant said that *"When it was quiet, it felt more like driving on ice, that everything became more fluid, there was nothing to touch"*. Thus, this type of feedback can create a natural connection to the environment.

The participants furthermore stated that the augmented sounds influenced the feeling of awareness regarding the surroundings. The sounds provided hints of what was happening around the car, and anchored participants in the virtual environment. This had a positive effect on the feeling of presence, and it helped to create a sense of space in the virtual environment. One participant noted *"Having no outside sound pretty much made you stuck to the center screen; The augmented sounds opened up the visuals"*.

Nonetheless, the augmented sounds seemed to have an ambiguous effect on comprehension of the current situation and the ability to fully create a projection of future status of surrounding objects. Some of the responses to the questions: "What are your impressions of the sound of other road users? Did it help you in any way?" highlighted some downsides of the augmented sounds and how they were presented. One recurring theme amongst these responses was that it at times was difficult interpreting the information provided by the auditory display. One participant said *"[It] Gives me awareness, but not much other information. [I] Stay attentive but not much more"*. Another participant responded: *"[It] Changed the experience in such a way that it became easier to discover or identify, but also a bit difficult / stressful because I have to sift through the information"*. Participants were able to hear and be aware of their surroundings but had difficulties locating the sound sources. Sounds were sometimes perceived as overlapping or hard to distinguish, and this affected participants' ability to fully comprehend how other road users were moving. The difficulty was not only dependent on the sounds, but also on the limited FOV. An outcome of perceiving sounds as difficult to distinguish and not being able to see the sound's source, is that this could create a need to search the interface for information. One participant stated, *"When there were only pedestrian sounds, you were just looking for pedestrians all the time"*. This suggests that invisible, but audible, information creates a split in attention, and a larger cognitive load for the operator.

Another outcome of having trouble interpreting auditory information, in combination with limited FOV, is that users may ignore information provided. A participant who chose to ignore the augmented sounds said *"I hear the pinging but I didn't see a bike. [...] Took it as a secondary, not important sound, unlike the engine (which helped me)"*. Providing information that is not acknowledged is not ideal. The risk that important information will be missed increases if a user deems part of the interface redundant, which is counterproductive when designing for SA.

How the augmented sounds were presented created a sense of incongruence for some participants. One stated: *"the bicycle sounds helped the most, the cars a little less and the pedestrians sounded too much in how big they are and where they are in space"* and another said, *"The pedestrian noise was so loud, it made me worry that I would hit someone, even if they were just walking beside me."*

However, despite the confusion, driving in the virtual environment with any of the auditory feedback still felt better for the participants than driving in a silent world. Participants mentioned that the preferred sound interface was the one with both propulsion and augmented sound. It created a feeling of safety, made it easier to drive safely and keep a good tempo, and increased the sensation of being a part of the virtual environment.

4. CONCLUSION

The current experiment showed that sound can be beneficial for the remote operator in many ways, as it can convey several aspects of the remote environment that may be crucial for the safe and efficient operation of the vehicle. The experiment carried out showed that both subjective and objective measures can be improved by the auditory display's enhanced propulsion sound/vibrations and augmented sound (surrounding road users): Presence, situational awareness, ego-motion, and speed regulation. However, based on the post-experiment interviews it seems as if the currently proposed auditory display concept should be refined in terms of its way of representing surrounding road users. In its current state, the auditory display seems to provide too much or irrelevant information regarding surrounding road users, which may increase the cognitive load for the driver and create a risk that the operator ignores important information.

One way of improving the auditory display could be to simply decrease the radius at which sources can be clearly heard so that only the surrounding road users in the very close vicinity to the ego-vehicle are audible. Another improvement could be to increase the relative level of those sources which are of importance in a specific situation. However, this might be difficult to achieve in a real system implementation since this would require input from automated driving functionality (target detection, threat assessment etc.), and the lack of this information might be the very reason for the remote operator being called in.

Future research should also test whether the effects observed in this study can be found also in more realistic settings. It would furthermore be interesting to look more into potential safety benefits over longer times of use and how operators prefer using these types of auditory displays in daily operations. Another sound-related aspect that has not been explored experimentally to great extent in the current project is the use/role of warning and information sounds in remote operation.

5. ACKNOWLEDGMENT

This work was supported by the Swedish Innovation Agency (VINNOVA) through the project REmote Driving Operations – REDO (Ref. No. 2019-03068).

6. REFERENCES

[1] D. Watzenig, M. Horn, *Automated Driving: Safer and More Efficient Future Driving*, Switzerland: Springer, 2017.

[2] SAE International, "J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 04 2021. [Online]. Available: https://www.sae.org/standards/content/j3016_202104/. [Accessed 9 March 2023].

[3] O. Amador, M. Aramrattana and A. Vinel, "A Survey on Remote Operation of Road Vehicles," in IEEE Access, vol. 10, pp. 130135-130154, 2022, doi: 10.1109/ACCESS.2022.3229168.

[4] F. Tener and J. Lanir, "Driving from a Distance: Challenges and Guidelines for Autonomous Vehicle

Teleoperation Interfaces," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), New York, NY, USA, 2022, pp. 250:1-250:13, doi: 10.1145/3491102.3501827.

[5] M. Sivak, "The information that drivers use: is it indeed 90% visual?," Perception, vol. 25, no. 9, pp. 1081-1089, 1996.

[6] Euro NCAP, "Lane Support," Euro NCAP, 2021. [Online]. Available: https://www.euroncap.com/en/vehicle-safety/the-ratings-explained/safety-assist/lane-support/. [Accessed: Mar. 9, 2023]

[7] European Union, “Commission delegated regulation (eu) 2021/1341 - EUR-Lex” [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R1341. [Accessed: Mar. 9, 2023].

[8] L. Evans, "Speed estimation from a moving automobile," Ergonomics, vol. 13, pp. 219-230, 1970.

[9] M. S. Horswill and A. M. Plooy, "Auditory Feedback Influences Perceived Driving Speeds," Perception, vol. 37, no. 7, pp. 1037-1043, 2008, doi: 10.1068/p5736.

[10] N. Merat and H. Jamson, "A Driving Simulator Study to Examine the Role of Vehicle Acoustics on Drivers' Speed Perception," in Driving Assessment Conference 6(2011), 2011, pp. 226-232, doi: 10.17077/drivingassessment.1401.

[11] L. J. Hettinger, "Illusory self-motion in virtual environments," in K. M. Stanney (Ed.), Handbook of Virtual Environments, Lawrence Erlbaum, 2002, pp. 471-492.

[12] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner, "Auditory Landmarks Enhance Circular Vection in Multimodal Virtual Reality," J. Audio Eng. Soc., vol. 57, no. 3, pp. 111-120, (2009 March.).

[13] B. E. Riecke, A. Väljamäe, and J. Schulte-Pelkum, "Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality," ACM Trans. Appl. Percept., vol. 6, no. 2, pp. 1-27, Feb. 2009, doi: 10.1145/1498700.1498701.

[14] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner, "Sound Representing Self-Motion in Virtual Environments Enhances Linear Vection," Presence: Teleoperators and Virtual Environments, vol. 17, no. 1, pp. 43-56, Feb. 2008, doi: 10.1162/pres.17.1.43.

[15] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," Human Factors, vol. 37, no. 1, pp. 32-64, Mar. 1995, doi: 10.1518/001872095779049543

[16] E. B. Nash, G. W. Edwards, J. A. Thompson, and W. Barfield, "A Review of Presence and Performance in Virtual Environments," Int. J. Hum.-Comput. Interact., vol. 12, no. 1, pp. 1-41, Jan. 2000, doi: 10.1207/S15327590IJHC1201_1.

[17] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments," Presence-Teleop. Virt. environ., vol. 6, pp. 603-616, 1997, doi: 10.1162/pres.1997.6.6.603.

[18] C. Hendrix and W. Barfield, "The Sense of Presence within Auditory Virtual Environments," Presence: Teleoperators and Virtual Environments, vol. 5, no. 3, pp. 290-301, Jun. 1996, doi: 10.1162/pres.1996.5.3.290.

[19] R. H. Gilkey and J. M. Weisenberger, "The Sense of Presence for the Suddenly Deafened Adult: Implications for Virtual Environments," Presence: Teleoperators and Virtual Environments, vol. 4, no. 4, pp. 357-363, Sep. 1995, doi: 10.1162/pres.1995.4.4.357.

[20] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner, "Auditory presence, individualized head-related transfer functions, and illusory ego-motion in virtual environments," in Proceedings of Presence 2004, Valencia, Spain, 2004, pp. 141-147.

[21] Cycling ‘74, "What is Max?" [Online]. Available: https://cycling74.com/products/max. [Accessed: March 12, 2023]

[22] IRCAM, "SPAT - Spatialization toolbox for Max/MSP," [Online]. Available: https://forum.ircam.fr/projects/detail/spat/. [Accessed: Mar. 12, 2023].

[23] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in Proceedings of the 1st Annual Conference on Robot Learning, 2017, pp. 1-16.

[24] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using Presence Questionnaires in Reality," Presence, vol. 9, no. 5, pp. 497-503, 2000.

[25] V. Linkov and M. Vanžura, "Situation Awareness Measurement in Remotely Controlled Cars," Front. Psychol., vol. 12, pp. 1-11, 2021, doi: 10.3389/fpsyg.2021.592930.

[26] S. A. Jessup, T. R. Schneider, G. M. Alarcon, T. J. Ryan, and A. Capiola, "The Measurement of the Propensity to Trust Automation," in International Conference on Human-Computer Interaction, Cham, Switzerland: Springer, 2019, pp. 476-489.

[27] J. Fagerlönn, P. Larsson, and J. Maculewicz, "The sound of trust: sonification of car intentions and perception in a context of autonomous drive," Int. J. Human Factors and Ergonomics, vol. 7, no. 4, pp. 339-355, 2020.

[28] R. Scupin, "The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology," Human Organization, vol. 56, no. 2, pp. 233-237, 1997. DOI: 10.17730/humo.56.2.x335923511444655