Multimodal Takeover Request Displays for Semi-Automated Vehicles: Focused on Spatiality and Lead Time

Harsh Sanghavi [1], Myounghoon Jeon[1], Chihab Nadri [1], Sangjin Ko[1], Jaka Sodnik[2], Kristina Stojmenova[2]

[1] Mind Music Machine Lab, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA
[2] Laboratory of Information Technologies, ICT Department, University of Ljubljana, Faculty of electrical engineering, Ljubljana, Slovenia

{harshks; cnadri; sangjinko; myounghoonjeon}@vt.edu
{jaka.sodnik; kristina.stojmenova}@uni-lj.si

Abstract. To investigate the full potential of non-speech sounds, this study explored the effects of different multimodal takeover request displays in semi-automated vehicles. It used a mixed design - the visual and auditory notification lead time was within-subjects, whereas the auditory notification spatiality was between-subjects. The study was conducted in a motion-based driving simulator with 24 participants. All participants were engaged in four 9-minute driving tasks in level-3 automated vehicle and simultaneously performed a non-driving related task (NDRT, online game). Each driving session contained three hazardous events with takeover request (in total 12 requests per user). The results showed that 3-second lead time evoked the fastest reaction time but caused high perceived workload and resulted in unsafe and non-comfortable maneuver. In terms of workload and maneuver, 7-second lead time showed better results than others. Auditory displays with directional information provided significantly better reaction times and reaction types. Subjective evaluation, on the other hand, did not show any significant differences between non-directional and directional displays. Additionally, the results showed that braking is a more common first reaction than steering, and that the NDRT did not influence the takeover request.

Keywords: Takeover, Multimodal displays, Semi-automated Vehicles; Sound spatiality; Lead time.

1 Introduction

As the autonomous vehicle technology advances through the different levels of automation, it is important to ensure a safe transition from autonomous to manual mode. Much research has been conducted in studying driver performance for takeover requests in highly autonomous vehicles [1–4]. Gold et al. [2], for example, measured driver performance for different levels of lead time.

Studies have also focused on takeover and the effect of non-driving related tasks (NDRT) performed in the vehicle during the autonomous mode [5–7]. The distraction from completing NDRT can impair performance [7]. Moreover, higher levels of automation correspond with higher levels of engagement in non-driving related tasks and lower situation awareness [1]. Therefore, to increase driver situation awareness and improve driver response, it is important for the takeover request to be as salient as possible to the driver. Ho et al. [8] found that multimodal warning signals effectively captured driver attention in demanding situations. Politis et al. [9] showed that multimodal displays performed better in takeover situations than unimodal visual displays. Petermeijer et al. [10] confirmed that multimodal displays are preferable over unimodal displays for takeover requests in autonomous vehicles.

2 Related Work

2.1 Selection of Lead Time

The lead time plays an important role in the effectiveness of takeover requests, especially in collision avoidance situations. The lead time is defined as the time from takeover notification offset until the time of collision into the obstacle if the participant would not perform any action to avoid the collision. Eriksson and Stanton [11] tested takeover requests in non-critical situations and reviewed other studies of takeover requests in critical situations. Different autonomous vehicle takeover related papers were reviewed and their details such as lead time and take over time were tabulated. They found that mean lead time between all 25 papers was found to be 6.37 ± 5.36 seconds having a mean reaction time of about 2.96 ± 1.96 seconds. They also noted that 3, 4, 6 and 7 seconds were the most frequently used lead times with 1.14, 2.05, 2.69 and 3.04 seconds as their corresponding reaction times [11].

Damböck et al. [4] tested lead times of 4,6 and 8 seconds and found that for simpler conditions of takeover, 4 second lead time was good enough, but drivers required longer lead times for complex situations. In another study by Gold et al. [12] two lead times were used, 5 seconds and 7 seconds. The 5 second lead time led to harder stops, swerving and failure to check blind spots while switching lanes. Although 5 seconds had better reaction times, 7 seconds led to better braking. A caveat of this better braking performance for a lead of seconds was that the car was still in accelerating motion, and thus, it was not fully in control, suggesting even 7 seconds lead time being not entirely safe for takeover requests. Zeeb et al. [2] showed that the reaction times (1.14 seconds) were substantially lower than other existing studies but the faster reaction times also coincided with a higher number of collisions.

Wan et al. [13] used different lead times to measure the optimum lead times for collision warnings in connected vehicle settings. Although it is not a highly autonomous vehicle setting, it extensively tested different lead times between 0 and 30 seconds and used measures indicative of the safe avoidance of obstacles. The study used three major dependent measures: kinetic energy reduction, minimum time to collision and collision rate. Kinetic energy reduction was considered most important as it determined if a vehicle can slow down enough to avoid a collision. A major conclusion of this paper was

that a lead time of 4-8 seconds was best for optimal safety benefit and a lead time between 5-8 seconds gave the best reaction time. It was also concluded that below 2 seconds, even at 45 mph the kinetic energy of the vehicle is too high for collision avoidance [13]. Mok et al. [14] used 2, 5 and 8 second lead times for their study and showed that 2 second lead time was not enough for drivers to regain sufficient control. Moreover, drivers rated the 2 second lead time as significantly less trustworthy. This suggests lead times below 2 seconds may be unsafe for takeover from highly autonomous vehicles. Mok et al. [14] showed that the 5 second lead time was found to be sufficiently long enough for drivers to regain control. The 8 second lead time also yielded good performance in takeover with higher levels of trust in drivers [14]. Wan et al. [13] found that lead time above 8 seconds showed a higher collision rate among drivers.

To summarize, the minimum value for lead time has been suggested as 2 seconds and maximum has been suggested as 8 seconds [13]. Many papers have used 7 seconds as their lead time [7, 9, 11,15]. Optimum lead time in a connected vehicle setting has been suggested as between 5 and 8 seconds [13]. Based on this data, we used the following lead times, 3 seconds, 5 seconds, 7 seconds and 9 seconds; two falling within the "optimum range" and two outside. Moreover, with previous studies already testing 5 and 7 seconds, it was of interest to see if the present study could get similar reaction time values for those levels of lead times. Testing the two outside the optimum range of lead times would also help compare the results with the study by Wan et al. [13] and help define the optimum lead times within the autonomous vehicle setting in a collision avoidance setting.

2.2 Spatiality of Auditory Warnings

Research has shown that the presentation of relevant spatial cues can help facilitate responses to target events, whether it be as the driver or pedestrians on the road [8, 16].

Driving research on collision spatial warning signals found out that the recognition of auditory warnings for threats from the front and back of the vehicle was more effective than vibrotactile ones, and provided a meaningful facilitatory effect on response time to events [8]. A closer look to spatial auditory warning cues also indicated how meaningful such warnings can be at providing locational data and capturing attention [17]. Research on auditory perception for pedestrians also indicated significant effects from spatiality, with users detecting the direction of traffic coming at a speed greater than 12mph at around 90% accuracy, indicating the importance and information provided from spatiality [16].

As drivers make similar decisions when evaluating obstacles on the road, research on directional warnings for drivers at intersection has also proven that driver behavior is improved by directional information in auditory warnings, significantly reducing drivers' brake time and deceleration rate [18]. Additionally, research on spatiality and display modality showed that spatially congruent auditory displays significantly improved drivers' performance as a unimodal display, with the hybrid visual-auditory performing better [19].

2.3 Selections of Subjective Characteristics

Marshall et al. [20] stated that it was important to consider the annoyance of an alert as annoyance can undermine the influence of warning systems. They also noted that the alert characteristics of warning signals affect perceived urgency and annoyance in different degrees. They also state that people judge highly urgent results as more appropriate while also being annoying. Auditory alerts having a high level of urgency may help drivers respond more effectively [20]. Studies have also shown that people respond to auditory alerts more quickly if they sound urgent [21, 22]. Therefore, annoyance and urgency were considered important parameters to measure the effectiveness of an auditory alert. Intuitiveness, the immediacy of recognizing cues and their relation to the users' mental model [23], was also considered an important characteristic as it would allow us to measure the effectiveness of the spatiality of the auditory warning.

2.4 Our Research Contribution

The main goal of our research is to revisit and extend some open questions related to suggested lead time in auditory displays for semi-automated driving mainly as previous reported results are not entirely coherent with each other. Additionally, we also studied the directionality of sound in this type of warning systems as it has shown to be very efficient in other HMI domains, but has not been extensively studied for takeover request displays.

Our main research questions are, therefore, the following:

- What is the optimal lead time in visual and auditory displays for takeover request?
- What are the main benefits and drawbacks of directional information in auditory displays for takeover request?

Both properties of auditory displays are studied through a wide combination of dependent variables, which reflect drivers' performance, safety related issues, mental workload and subjective preferences. This enabled us to draw comprehensive conclusions and provide some concrete guidelines for auditory takeover displays for semi-automated driving.

3 Experimental Setup

The experiment was conducted in a motion-based driving simulator by Nervtech [24]. It consisted of a racing car seat, a steering system and sport pedals. The visuals were displayed on a triple-screen configuration, which covered 120° horizontal field of view and consisted of three equal curved 48' HD TVs. The driving scenarios were developed using AV Simulations software [25], which run on a computer with an i7 - 8086K CPU and Nvidia GTX 1080 graphics card. The driving scenarios involved driving on a three-lane highway with 110 km/h (70 miles/h) speed limit. The visibility was lowered to approximately 100 meters using fog (see **Fig. 1**).

Fig. 1. Reduced visibility using fog.

4 Methodology

A simulation of a SAE Level 3 [26] automated vehicle was used for the study. The automated driving system (ADS) was engaged at all times, except during unexpected situations (for example, car accident, deer on the road, construction work, etc.). When the ADS was on, the vehicle would always drive in the middle lane of the road with a fixed speed of 110 km/h. The ADS system could be disengaged and reengaged at any time with a simple pull up of the left lever on the steering wheel. The unexpected situation always resulted in closing two out of the three high-way lanes – the center and either the left or the right lane – so that the driver would have to perform a lane-change in order to avoid an accident.

The experiment involved a test drive and four 9-minute-long trials. The test drive lasted 3 minutes and was intended for the participants to get familiar with the driving simulator, the ADS system, the auditory display and the NDRT. In each of the four trials, the participants were exposed to 3 critical situations in which they had to take over control of the vehicle. The event time and location of the free lane was randomized to try to avoid learning and anticipation effects.

To ensure consistency of the use of the ADS system, approximately 150 meters after the obstacle on the road, a pre-recorded sound notification asked the participant to re-engage the ADS system and continue with the NDRT.

4.1 Participants

Twenty-four participants (20 male), aged from 20 to 42 years old (M = 26.5, SD = 5.13) participated in the study. Only participants with valid driving licenses were asked to participate in the study.

4.2 Tasks

During the experiment, the participants were asked to perform two tasks with the same level of priority:

- Non-driving-related task (NDRT),
- Take control of the vehicle in case of a takeover request.

The non-driving-related task was *2048* - a single-player sliding block puzzle game, whose objective is to slide and combine numbers on a grid with the purpose of achieving a sum of 2048.

Since the vehicle had a high level of automation, participants had to perform the takeover task only in case of disabled automation of the vehicle due to an unexpected situation or obstacle. When an obstacle appeared, the driver was given an auditory and visual takeover request, consisting of text and warning icon, as shown in Figure 2. The visual notification was presented at the same with the auditory notification, on the dashboard display in the simulator (see **Fig. 2**). Upon the takeover request, the autonomous vehicle switched over to the non-autonomous mode. This allowed the driver to take control of the vehicle. When operated manually, the vehicle had an automatic transmission system.

Fig. 2. Visual notification used for takeover request.

4.3 Lead Times and Spatiality

Lead times and spatiality were the two independent variables primarily observed in the study. Lead times were assigned within-subject, whereas spatiality was observed between-subjects.

Based on previous research, lead times were set as 3s, 5s, 7s and 9s (LT3, LT5, LT7 and LT9). The lead time was defined as the time from takeover notification offset until

the time of collision into the obstacle if the participant would not perform any action (brake or steer). Each lead time was observed in a separate trial, where the participant had to take over control of the vehicle three times.

Two types of auditory displays for takeover notifications were used in the study: non-directional and directional sound. The non-directional auditory display played the takeover notification from the front center of the vehicle. For the trials with the directional sound, the notification always came from the direction of the free lane, where the participant should drive in order to avoid collision. Using the Nervtech simulator surround sound system, the sound would be played from direction of the left and right front side of the simulated vehicle.

A non-speech auditory display was designed, including two dominant frequencies (880, 1760 Hz) repeated four times using sine wave, following ISO [27] guidelines. The sound was played at 65 dB, whereas the ambient environment sound was played at 60 dB.

4.4 Dependent variables

We observed three groups of dependent variables:

- takeover reaction time and reaction type (steering or braking),
- takeover success rate and driving performance, and
- self-reported data (perceived workload and a sound questionnaire).

Reaction times and reaction types were recorded for all takeover situations varying in lead time and sound direction. Reaction time was measured from the moment of the auditory-visual takeover request until the driver regained control by breaking or steering the wheel. Breaking was initiated by pressing the brake pedal for at least 10% while steering was detected as turning the steering wheel for more than 2°, based on similar experimental data reported by Gold et al. [2]. As reaction type, we detected which of these two operations the participants would perform first, steering or braking.

Furthermore, the takeover success rate, and logs of speed, acceleration and jerk were observed. Each of these variables was observed from the auditory takeover request until the pre-recorded notification asking the participant to reengage the ADS.
After completing each trial (with a different lead time), participants were asked to evaluate the takeover task using the NASA TLX questionnaire. For each trial, the score of the 2048 (task success rate) was also recorded. At the end of the experiment, participants were further asked to evaluate their user experience with the auditory display using a sound questionnaire.

5 Results

Levene's test was used to assess the equality of variances and Shapiro-Wilk test for exhibiting the normal distribution of each group of data. Based on these results, appropriate parametric and non-parametric tests were used for analyzing the data.

5.1 Reaction Time and Reaction Type

All groups (lead time and sound display type) did not have a normal distribution. The between-subjects design results (sound direction) were analyzed with the independent-samples Mann-Whitney U test 4. The within-subject design results (lead time) were analyzed with the related-samples Friedman test and post-hoc Wilcoxon-signed ranks test. For all pairwise comparisons a Bonferroni adjustment was applied to control for Type-I error, resulting in adjusted alpha levels (critical alpha level = .0083 (0.05/6)).

Reaction time. Fig. 3 shows the mean reaction times for each of the tested lead time for trials with directional and non-directional sound display. Reaction times for the directional sound (M = 1857.24 ms, SD = 1161.94 ms) were significantly faster than the reaction times for the non-directional sound display (M = 1977.83 ms, SD = 1079.06 ms), U = 7843, p = .030.

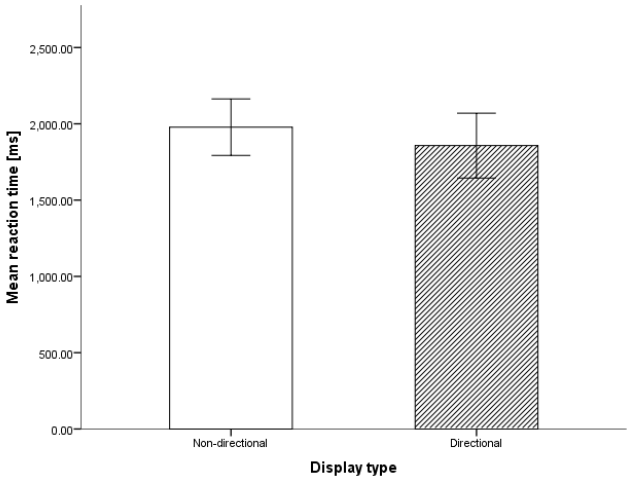

Fig. 3. Mean reaction times and SE to takeover requests with different lead times for non-directional and directional display

The results for different lead times showed that there was a statistically significant difference in reaction times to the takeover request, $\chi 2(3) = 50.961$, $p < .001$ (Figure 4). The reaction times to takeover requests with LT3 were significantly faster than reaction times for takeover requests with LT5 ($p < .001$), LT7 ($p < .001$) and LT9 ($p < .001$). Furthermore, the reaction times to takeover request with lead time of LT5 were significantly faster than LT7 ($p = .002$) and LT9 ($p = .004$). There were no statistical differences between reaction times lead LT7 and LT9 ($p = .305$). **Table 1** reports on the mean reaction times for each display type takeover request for all four observed lead times.

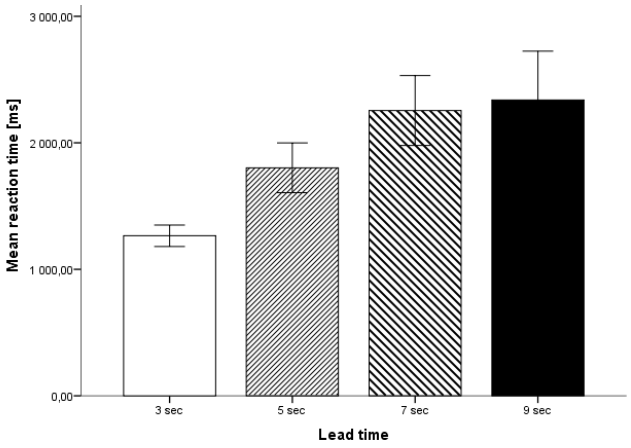

Fig. 4. Mean reaction times and SE to takeover requests for different lead times

Table 1. Mean reaction times to takeover requests for different lead times

Display type	**Lead time**			
	3 sec	**5 sec**	**7 sec**	**9 sec**
Non-directional	1296.71 ms	2006.27 ms	2165.18 ms	2430.71 ms
Directional	1233.82 ms	1604.29 ms	2348.21 ms	2245.66 ms

We also observed the reaction times for different types of reaction, steering or braking. These results showed that the reaction times for breaking (M = 1857.24 ms, SD = 1240.37 ms) were significantly faster compared to steering (M = 1977.84 ms, SD = 1079.06 ms), U = 11 960, p < .001 (see **Fig. 5**).

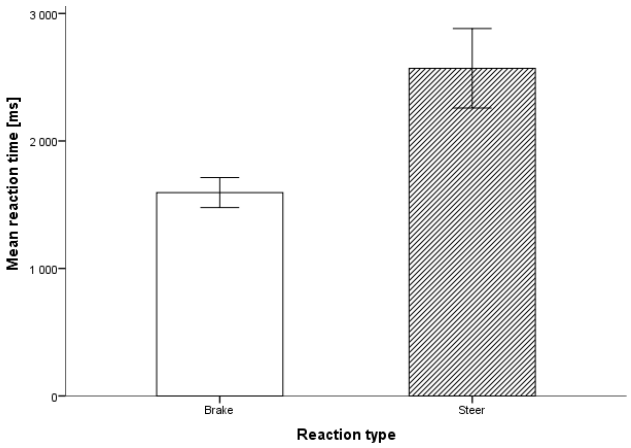

Fig. 5. Mean reaction times and SE for different types of takeover

Reaction Type. In 68.1 % of takeovers, the participants' first reaction to the takeover notification was breaking, and in 31.9 % it was steering. These ratios are similar for all lead times as shown in **Table 2** and **Fig. 6**.

Table 2. Reaction type frequency for different lead times

Reaction type	Lead time			
	3 sec	5 sec	7 sec	9 sec
Brake	66.7%	69.4 %	68.1%	68.1%
Steer	33.3%	30.6%	31.9%	31.9%

The frequency of reaction types to takeover requests with different lead time, however, indicated that there is no statistically significant association between the reaction type and the lead time, $\chi(1) = 0.064$, $p = .800$; that is, the reaction type preference was not influenced by the lead times (see **Fig. 6**).

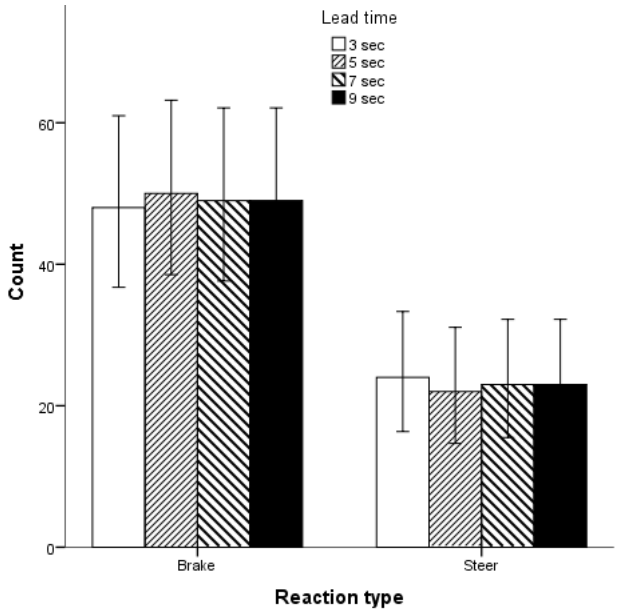

Fig. 6. Reaction type frequency for different lead time

5.2 Takeover Success Rate and Driving Performance

Takeover Success Rate. Each participant performed 12 takeover requests (3 with all four lead times). Out of all of the 288 takeover requests in this study, 99.986% were successful.

Speed. The results on speed showed that the average speed during the takeover (manual driving of the vehicle) was higher in trials with the takeover request with non-

directional sound display (M = 92.26, SD = 18.02) compared to trials with the directional sound display (M = 88.31 km/h, SD = 17.69 km/h), U = 8979, p = .049 (**Fig. 7**). When observing the average speed for trials with different lead times, the analysis did not reveal any statistically significant differences, $\chi 2(3) = 2.050$, p = .562 (**Fig. 8**).

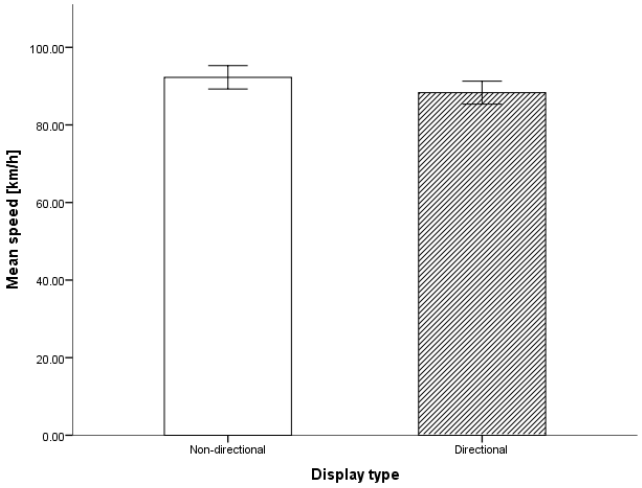

Fig. 7. Mean speed and SE during takeover for trials with non-directional and directional display

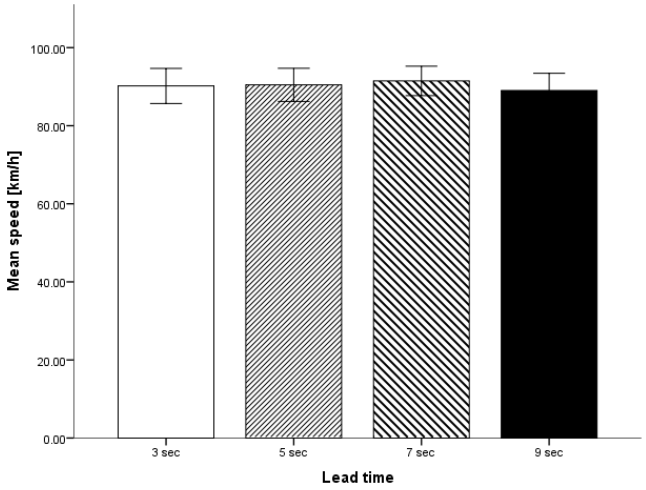

Fig. 8. Mean speed and SE during takeover for trials with different lead times

Deceleration (braking). The results for deceleration (braking) did not show any statistical differences in the takeover situations with directional sound display (M = 0.65

km/h2, SD = 0.59 km/h2) compared to trials with non-directional sound display, U = 10674, p = .665.

Significant differences in deceleration were, however, revealed for different lead times (see **Fig. 9**), $\chi 2(3) = 67.223$, $p < .001$. The pairwise comparisons showed that the average deceleration for trials with LT3 was higher compared to LT5 ($p < .001$), LT7 ($p < .001$) and LT9 ($p < .001$). Furthermore, the average deceleration for trials with LT5 was significantly higher than LT7 ($p < .001$) and LT9 ($p < .001$). There were no statistical differences in deceleration for trials with lead time of LT7 and LT9 ($p = .053$).

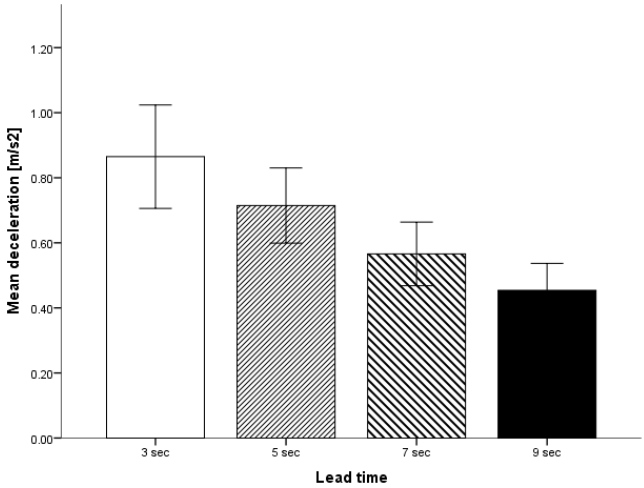

Fig. 9. Mean deceleration during takeover for different lead times

Jerk. Additionally, we observed also jerk – a time derivative of acceleration, which gives us the information of the rate of change of acceleration. Jerk has been used to determine the driving comfort experience, with (absolute) values more than 0.6 m/s3 indicating uncomfortable maneuvers [28].

The results did not reveal any statistical differences in mean jerk in the takeover situations with the directional sound display (M = 0.086, SD = 0.236) compared to trials with the non-directional sound display (M = 0.031, SD = 0.256), U = 11577.5, p = .068. There were, however, significant differences in mean jerk for takeover situations when comparing trials with different lead times, $\chi 2(3) = 8.561$, $p = .036$ (see **Fig. 10**). The post-hoc test showed that the average jerk was statistically higher for LT3 compared to LT5 ($p = .003$).

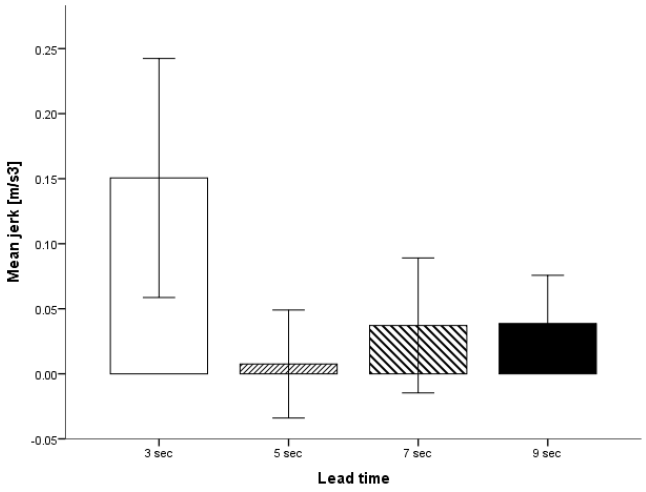

Fig. 10. Mean jerk during takeover for different lead times

Fig. 11 and **Fig. 12** were created to give a visualization of the steering maneuvers performed after the request to takeover. The first steering directions, swerving and faster maneuvers that resulted into higher jerk can be seen for each lead time for both non-directional (see **Fig. 11**) and directional sound displays (see **Fig. 12**) to avoid the road obstacle.

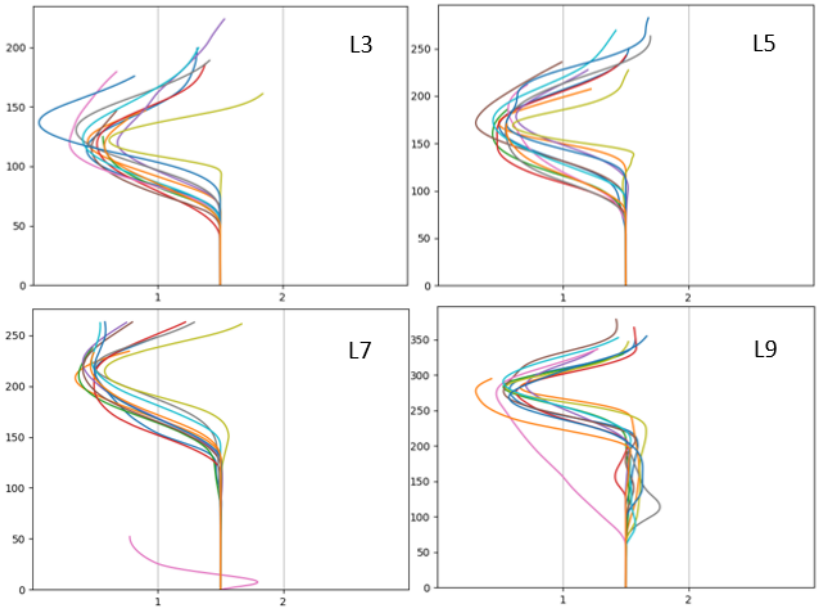

Fig. 11. Vehicle trajectories for takeover situations with non-directional display

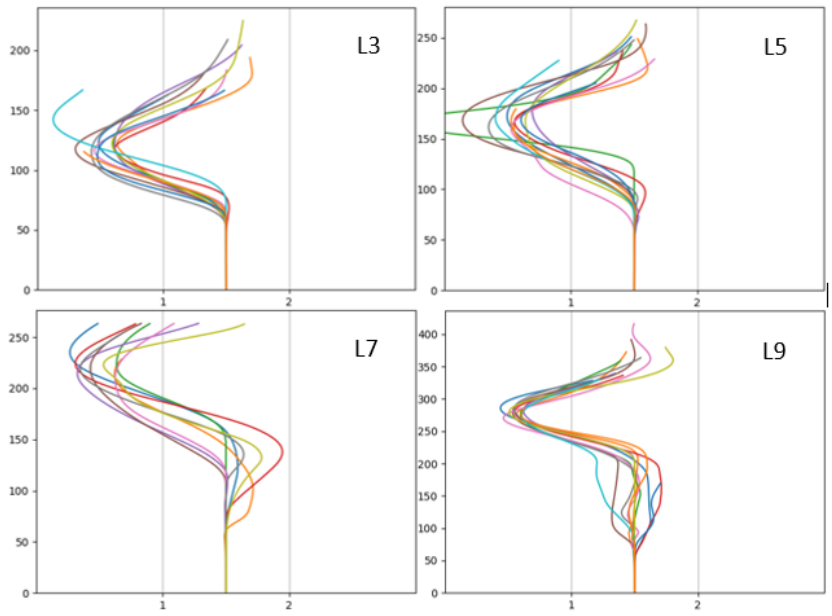

Fig. 12. Vehicle trajectories for takeover situations with directional display

5.3 Questionnaires

Sound questionnaire. Table 3 and **Fig. 13** show the results of subjective rating scores on a 0-100 scale for the non-directional display and the directional sound display. The directional sound display received a numerically higher score in "intuitiveness" and lower scores in both "identification" and "annoying" than the non-directional display. Results were analyzed with an independent-groups t-test for each question showing differences in scores (Questions 2, 3, and 5). None of them revealed statistically significant differences among display types in mean "identification" rating score ($t(22) = -0.890$, $p = .384$), in mean "intuitiveness" rating score ($t(22) = 0.127$, $p = .900$), and in mean "annoying" rating score ($t(22) = -1.538$, $p = .139$).

Table 3. Sound questionnaire rating scores in different display types

Questions	Non-directional	Directional
How helpful was the sound?	6.50	6.50
Did the sound help you identify which direction you are supposed to go?	2.00	1.58
How intuitive was the sound?	5.25	5.33
How urgent was the sound?	6.75	6.75
How annoying was the sound?	2.92	2.00

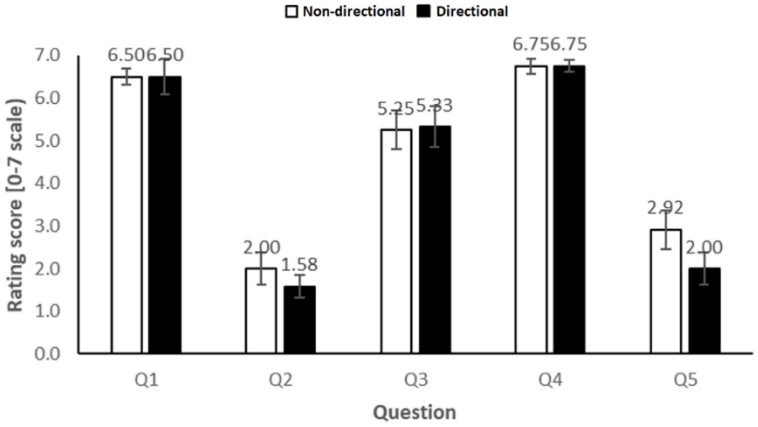

Fig. 13. Sound questionnaire rating scores of post-questionnaires

NASA-TLX Rating Score. The results were analyzed with a 4 (Lead time) x 2 (Display) repeated measures analysis of variance (ANOVA) for each subscale including mental, physical and temporal demand, performance, effort, and frustration.

Mental Demand. The analysis showed that there were no statistically significant differences among display types in mean "mental demand" rating score, $F(1, 22) = 1.309$, $p = .265$ and lead times, $F(1, 22) = 2.368$, $p = .079$. There was no significant interaction between Display and Lead time, $F(3, 66) = 0.080$, $p = .970$.

Physical demand. The analysis showed that there were significant differences among lead times in mean "physical demand" rating score, $F(3, 66) = 3.802$, $p = .014$. However, there were no statistically significant differences found among display types, $F(1, 22) = 1.309$, $p = .191$. There was no significant interaction between Display and Lead time, $F(3, 66) = 0.366$, $p = .778$.

Temporal demand. The analysis showed that there were significant differences among lead times in mean "temporal demand" rating score, $F(3, 66) = 8.255$, $p < .0001$. However, there were no statistically significant differences found among display types, $F(1, 22) = 2.344$, $p = .140$. There was no significant interaction between Display and Lead time, $F(3, 66) = 2.143$, $p = .103$.

Performance. The analysis showed that there were no significant differences among lead times in mean "performance" rating score, $F(3, 66) = 0.334$, $p = .801$ and display types, $F(1, 22) = 0.037$, $p = .849$. There was no significant interaction between Display and Lead time, $F(3, 66) = 1.1820$, $p = .323$.

Effort. The analysis showed that there were significant differences among lead times in mean "effort" rating score, $F(3, 66) = 4.113$, $p = .010$. However, there were no

statistically significant differences found among display types, F(1, 22) = 1.746, p = .200 or between Display and Lead time, F(3, 66) = 0.308, p = .819.

Frustration. The analysis showed that there were no significant differences among lead times in mean "frustration" rating score, F(3, 66) = 2.488, p = .068 and display types, F(1, 22) = 0.035, p = .853. There was no significant interaction between Display and Lead time, F(3, 66) = 1.423, p = .244.

NASA-TLX rating scores vs. Lead time

Fig. 14 and **Table 4.** NASA-TLX Rating scores in different lead times show NASA-TLX rating scores of each six subscales over different lead times. It shows that physical demand and temporal demand decrease as a lead time increases. However, the highest rating score of "performance" and the lowest rating score of "effort" were observed when the lead time was 7 seconds. For the multiple comparisons among lead times (LT3, LT5, LT7, and LT9) in "physical demand", "temporal demand", and "effort" rating scores, a paired-samples t-tests were conducted. All pairwise comparisons in this study applied a Bonferroni adjustment to control for Type-I error, which meant that we used more conservative alpha levels (critical alpha level = .0083 (0.05/6)).

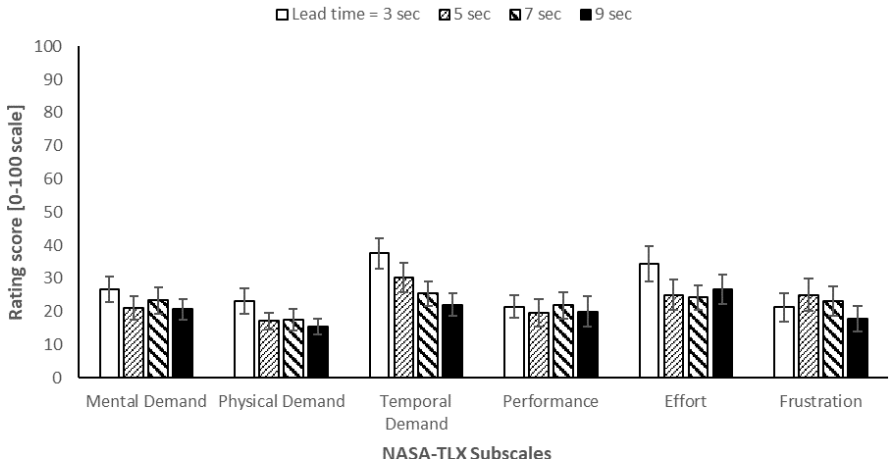

Fig. 14. NASA-TLX Rating scores of subscales

In "physical demand", participants showed significantly higher temporal demand in LT3 (M = 23.123, SD = 18.928) than LT9 (M = 15.417, SD = 11.508), $t(66)$ = -3.170, p = .002. In "temporal demand", participants showed significantly higher temporal demand in LT3 (M = 37.500, SD = 22.554) than both LT7 (M = 25.417, SD = 18.053), $t(66)$ = -3.670, p = .005 and LT9 (M = 22.083, SD = 17.126) , $t(66)$ = -4.682, p < .0001. In "effort", participants showed significantly a higher temporal demand in LT3 (M = 34.375, SD = 26.716) than LT5 (M = 25.000, SD = 22.458), $t(66)$ = -2.881, p = .005. Also, it was significantly higher than LT7 (M = 24.167, SD = 18.278), $t(66)$ = -3.138, p = .003.

Table 4. NASA-TLX Rating scores in different lead times

Subscale	Lead time			
	3 sec	**5 sec**	**7 sec**	**9 sec**
Mental Demand	26.67	21.04	23.33	20.63
Physical Demand	23.13	17.08	17.50	15.42
Temporal Demand	37.50	30.21	25.42	22.08
Performance	21.46	19.58	21.88	20.00
Effort	34.38	25.00	24.17	26.67
Frustration	21.25	25.00	23.13	17.92

5.4 NDRT performance

The results for the average score of 2048 were analyzed with a 4 (Lead time) x 2 (Display) repeated measures analysis of variance (ANOVA). The analysis showed that there were no significant differences among lead times in average score of 2048, $F(3, 66) = 0.346$, $p = .792$ and display types, $F(1, 22) = 2.618$, $p = .120$. There was no significant interaction between display and lead time, $F(3, 66) = 2.313$, $p = .084$. There was numerically higher average score of 2048 with the directional sound display (675M = 7321.396, SD = 4758.245) than the non-directional display type (M = 5068.063, SD = 3143). However, there were no statistically significant differences between display types.

6 Discussion and conclusions

To design more optimized takeover request displays in automated vehicles, we evaluated takeover time and type, driving performance, perceived workload, and subjective experience measures for two spatial sounds with four lead times, while our participants performed a non-driving related task. The results for all of the observed variables did not show a consistent pattern for one specific lead time or display type. However, when observed together, it is possible to differentiate which lead time and display type offer safest takeover and to some extent better takeover experience.

The shortest (3 sec) lead time evoked fastest reaction times, which is in line with previous research [13, 14]. However, based on the jerk results, this lead time resulted in the takeover maneuver surpassing the threshold of 6 m/s3, which suggests unpleasant maneuvers in driving. **Fig. 11** and **Fig. 12** further support these results as it can be seen that the L3 resulted in most sharp swearing when avoiding the road obstacle. Furthermore, the results from the NASA-TLX showed that participants reported high perceived workload in terms of physical demand, temporal demand and effort. By extending the lead time to 5s and 7s, the reaction time decreased but the perceived workload and subjective preferences of users increased significantly. The avoiding maneuver also improved suggesting the 7s lead time as the optimal choice among the tested signals. This is coherent with previous results reported in [13].

The reaction times for braking were significantly faster than steering, and in almost two thirds of all takeover request situations, the participant's first reaction was braking.

This is probably due to the fact that the NDRT was manually operated. However, when explored the frequencies, the results revealed that the length of the lead time did not affect the first reaction type. Although not statistically significant, it is interesting to notice that in the 3-sec lead time condition majority of the participants decide to steer instead of break. Due to the proximity of the road obstacle, which at this point was already visible in comparison to the situations with longer lead times, participants could have found braking as not efficient (not enough time to stop). In real life situations, such decisions for sudden lane changes without reducing the speed can be dangerous and less comfortable.

Furthermore, we observed the influence of spatiality and directionality of auditory signals. Regardless of the reaction type, the directional display evoked faster reaction times. Interestingly, the directional display showed lower average vehicle speed during the takeover. We can cautiously infer that the participants might be slower with the directional sound because they recognized the direction and so, prepared for changing the direction of the vehicle.

When observing the self-reported ratings, there was no significant difference in the sound experience questionnaire. However, at least, the annoyance question favored the directional sound (directional: 2.00 vs. non-direction: 2.92), which bodes well for further investigation of display directionality in this domain. Annoying sounds have been shown to have a positive correlation with alertness and urgency, but research has also suggested that too much annoying and startling sound will discourage users to use auditory displays. Given that our directional sound performed better and received lower annoyance, it seems desirable. Research with more participants will clarify whether there is different preference for either display type.

Finally, when observing the NDRT results, it can be noticed that all of the participants achieved high scores suggesting a high engagement in the task. Although previous research has indicated that higher engagement in NDRT can affect the takeover success rate, we could not find any evidence of such effect in this study. This could be further explored with different NDRT tasks and different durations of the trials.

As for multimodality, in this study we tested only auditory variables with a fixed visual display. Future research can include variations of other modalities (visual and tactile) and spatiality of visual and tactile displays. Planned research includes making a mathematical model and validating it with empirical research outcomes to formalize takeover procedure in automated vehicles.

Acknowledgment

This work was partially supported by a grant (number BI-US/19-21-008) from Slovenian Research Agency and grant (code 17TLRP-B131486-01) from Transportation and Logistics R&D Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

1. Merat N, Jamson AH, Lai FC, Carsten O. Highly automated driving, secondary task performance, and driver state. Human factors;54(5):762-71. (2012)
2. Gold C, Damböck D, Lorenz L, Bengler K. "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 2013 Sep (Vol. 57, No. 1, pp. 1938-1942). Sage CA: Los Angeles, CA: Sage Publications (2013).
3. Zeeb K, Buchner A, Schrauf M. What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident analysis & prevention. 2015 May 1;78:212-21.
4. Damböck D, Bengler K. Übernahmezeiten beim hochautomatisierten Fahren. In5. Tagung Fahrerassistenz 2012.
5. Feldhütter A, Gold C, Schneider S, Bengler K. How the duration of automated driving influences take-over performance and gaze behavior. InAdvances in ergonomic design of systems, products and processes 2017 (pp. 309-318). Springer, Berlin, Heidelberg.
6. Neubauer C, Matthews G, Saxby D. The effects of cell phone use and automation on driver performance and subjective state in simulated driving. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2012 Sep (Vol. 56, No. 1, pp. 1987-1991). Sage CA: Los Angeles, CA: Sage Publications.
7. Radlmayr J, Gold C, Lorenz L, Farid M, Bengler K. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. InProceedings of the human factors and ergonomics society annual meeting 2014 Sep (Vol. 58, No. 1, pp. 2063-2067). Sage CA: Los Angeles, CA: Sage Publications.
8. Ho C, Tan HZ, Spence C. The differential effect of vibrotactile and auditory cues on visual spatial attention. Ergonomics. 2006 Jun 1;49(7):724-38.
9. Politis I, Brewster S, Pollick F. Language-based multimodal displays for the handover of control in autonomous cars. In Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications 2015 Sep 1 (pp. 3-10).
10. Petermeijer S, Bazilinskyy P, Bengler K, De Winter J. Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. Applied ergonomics. 2017 Jul 1;62:204-15.
11. Eriksson A, Stanton NA. Takeover time in highly automated vehicles: noncritical transitions to and from manual control. Human factors. 2017 Jun;59(4):689-705.
12. Gold C, Berisha I, Bengler K. Utilization of drivetime–performing non-driving related tasks while driving highly automated. InProceedings of the Human Factors and Ergonomics Society Annual Meeting 2015 Sep (Vol. 59, No. 1, pp. 1666-1670). Sage CA: Los Angeles, CA: SAGE Publications.
13. Wan J, Wu C, Zhang Y. Effects of lead time of verbal collision warning messages on driving behavior in connected vehicle settings. Journal of safety research. 2016 Sep 1;58:89-98.
14. Mok B, Johns M, Lee KJ, Miller D, Sirkin D, Ive P, Ju W. Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles.

In2015 IEEE 18th international conference on intelligent transportation systems 2015 Sep 15 (pp. 2458-2464). IEEE.
15. Gold C, Körber M, Lechner D, Bengler K. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. Human factors. 2016 Jun;58(4):642-52.
16. Barton BK, Ulrich TA, Lew R. Auditory detection and localization of approaching vehicles. Accident Analysis & Prevention. 2012 Nov 1;49:347-53.
17. Ho C, Spence C. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of experimental psychology: Applied. 2005 Sep;11(3):157.
18. Zhang Y, Yan X, Yang Z. Discrimination of effects between directional and non-directional information of auditory warning on driving behavior. Discrete Dynamics in Nature and Society. 2015 Jan 1;2015.
19. Liu YC, Jhuang JW. Effects of in-vehicle warning information displays with or without spatial compatibility on driving behaviors and response performance. Applied ergonomics. 2012 Jul 1;43(4):679-86.
20. Marshall DC, Lee JD, Austria PA. Alerts for in-vehicle information systems: Annoyance, urgency, and appropriateness. Human factors. 2007 Feb;49(1):145-57.
21. Burt JL, Bartolome DS, Burdette DW, Comstock Jr JR. A psychophysiological evaluation of the perceived urgency of auditory warning signals. Ergonomics. 1995 Nov 1;38(11):2327-40.
22. Haas EC, Casali JG. Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. Ergonomics. 1995 Nov 1;38(11):2313-26.
23. Garzonis S, Jones S, Jay T, O'Neill E. Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. InProceedings of the SIGCHI Conference on human factors in computing systems 2009 Apr 4 (pp. 1513-1522).
24. Vengust M, Kaluža B, Stojmenova K, Sodnik J. NERVteh compact motion based driving simulator. InProceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct 2017 Sep 24 (pp. 242-243).
25. AV Simulations. SCANeR DT. Available online: https://www.avsimulation.fr/
26. SAE (2014) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. SAE Standard J3016, USA
27. Draft, ISO Working. Road vehicles-Ergonomic aspects of transport information and control systems-Specifications for in-vehicle auditory presentation," ISO Standard 15006:2011(E).
28. Kilinç, A.S. and Baybura, T., 2012. Determination of minimum horizontal curve radius used in the design of transportation structures, depending on the limit value of comfort criterion lateral jerk. *TS06G-Engineering Surveying, Machine Control and Guidance, Rome, Italy*.