Takeover and Handover Requests using Non-Speech Auditory Displays in Semi-Automated Vehicles

Kyle Kutchek
Mind Music Machine Lab
Department of Electrical and Computer Engineering, Michigan Tech
Houghton, MI USA
kakutche@mtu.edu

Myounghoon Jeon
Mind Music Machine Lab
Department of Industrial and Systems Engineering, Virginia Tech
Blacksburg, VA USA
myounghoonjeon@vt.edu

ABSTRACT

Since non-speech sounds can convey urgency well, they have been used as alerts in the vehicle context, including control transitions (handover and takeover) in automated vehicles. However, their potential has not been fully investigated to make use in international standards. To contribute to making authentic standards, the present paper investigated the effects of various non-speech displays to further refine auditory variables. Twenty-four young drivers drove in the

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313078

KEYWORDS
Auditory displays; multimodal displays; autonomous vehicles; take-over request

Figure 1: NADS MiniSim Driving Simulator.

Figure 2: 2048 Game as a secondary task.

driving simulator that had both handover and takeover transitions between manual and automated modes with a secondary task. The reaction times for handover and takeover and other sound user experience questionnaire results are reported with discussions and future work.

1 INTRODUCTION

Level 3 automated vehicles have been using mostly visual and auditory takeover displays. To date, however, there is no international standard for takeover display design. Multiple Resources Theory (MRT) [1] predicts that multiple modalities would be more effective for doing multiple concurrent tasks such as one that would take place while driving. Based on MRT, there has been some auditory display research for control transitions in automated vehicles. McKeown [2] showed that auditory icons and speech had the highest accuracy and response time compared to other auditory cues, but noted that speech was poor at communicating a high-urgency situation. Another study [3] showed that earcons and speech showed comparable takeover time, whereas spearcons were not optimal for this purpose. However, little research has refined the detailed parameters of non-speech sounds (e.g., earcons) in the context of control transition in the automated vehicles. From this background, we explored different types of non-speech auditory cues, including already implemented sounds in commercial vehicles. Also, we were specifically interested in drivers' psychological readiness to transitions so we asked our participants to respond to the auditory cues when they felt they were ready to take over or hand over.

2 METHODS

2.1 Participants

Twenty-four undergraduate students (6 female; *Mean* age = 20, *SD* = 1.1; *Mean* years of driving = 3.7 years, *SD* = 1.4) participated in this study for course credit. All of the participants had a valid driving license and more than two years of driving experience to control over novice effects.

2.2 Stimuli

We designed six auditory displays using non-speech sounds for both takeover and handover. We had three different sound types: existing takeover sounds (Hyundai and Tesla), indicator sounds (starting up and shutting down), and looming soundscapes (2 main frequency and 1 main frequency). In the Hyundai IONIQ EV if the driver removes their hands from the steering wheel for too long, a series of 0.09 second tones at 1498 Hz 0.03 seconds apart are played up to 42 times until the driver touches the steering wheel again. If the driver fails to take control during the 5 second interval, a 43rd tone is played for 0.21 second followed by a 44th tone for 0.09 second. We used the first 1498 Hz tone repetition with a 0.09 second onset and 0.03 second interval for this experiment. In the Tesla firmware 8.0, when the driver's hands are taken off the wheel for too long, the vehicle will alert the driver with a single pair of 0.2 second 553 Hz tones that are 0.06 second apart and flash the dashboard. After a period of no hand contact on the wheel, the same pair of tones will play twice at an increased volume. Finally, automation will be deactivated for the rest of the drive and the tone pairs will play continually at an increased volume and the display will flash until the driver's hands are on the wheel. We used a 553Hz tone repetition with a 0.2 second onset and a 0.06 second interval.

Table 1: Auditory display conditions

Auditory Display 1	Existing Sound1 – Hyundai
Auditory Display 2	Existing Sound2 – Tesla
Auditory Display 3	Indicator Sound1 – Starting up
Auditory Display 4	Indicator Sound2 – Shutting down
Auditory Display 5	Looming Soundscape1 – two frequencies
Auditory Display 6	Looming Soundscape2 – one frequency

Figure 3: Experimental Scene with the simulator.

This Tesla sound satisfies NHTSA and ISO guidelines, but these tones are outside of the frequency boundaries for the lowest response time defined by Lin et al [4] (i.e., 1750 or 3000 Hz). The indicator sounds are composed of a pair – starting up and shutting down, which are similar to the operating system sound on computers. Both sounds include multiple frequencies. The starting up sound includes five dominant frequencies (around 400Hz, 530Hz, 620Hz, 788Hz, and 1050 Hz) with the increasing polarity. The shutting down sound includes the exactly same frequencies with the decreasing polarity. This design was intentional to see the appropriateness of the use of the combination of the increasing and decreasing pair as usually applied in other electronic devices [5]. To keep the similar duration across stimuli, these sounds were repeated twice. Finally, we designed two looming soundscapes. Literature [5] introduces the plausibility of using soundscapes in addition to intermittent warnings in the automated vehicle environment for better user experience. The first looming sound includes two dominant frequencies (around 560Hz and 620Hz) and the second one includes only one dominant frequency (around 560Hz), which are similar to Tesla sound frequency. All sound stimuli were normalized to a similar length (6-7 seconds), a similar volume, and the same file format (.wav).

2.3 Apparatus

For this experiment, a mid-fidelity National Advanced Driving Simulator (NADS) MiniSim (Figure 1) was used. The MiniSim had three 42" plasma displays with a 1280x800 resolution. It included a real steering wheel, adjustable car seat, gearshift, and gas and brake pedals, as well as a TFT LCD monitor with a 1280x800 resolution to display the speedometer, etc. Environmental sound effects were played through two embedded speakers. The auditory displays were also generated through these speakers. For the secondary task, an online game (Figure 2), a Dell laptop was used.

2.4 Design and Procedure

This experiment used a within-subjects design with six auditory display conditions and two task types. The order of the conditions for each participant was randomized prior to testing. The driving scenarios were set in a rural area with the car driving on a two-lane road. There were no oncoming traffics and no intersections. Speed limit was set at 50 miles per hour. Situations on the road that triggered the participants to take over included deer, a parked car, and a service vehicle, all blocking most or all of the driving lanes (Figure 3). After completing the consent form procedure, the participants were brought to the driving simulator, allowed to adjust the seat to get comfortable, and briefed about the experimental procedure. When the participants felt ready, the experiment began with a randomly selected scenario. The participants began to drive manually for about 30 seconds to get familiar with the driving simulator. After a short driving, one of the six auditory cues signaled the participants that the vehicle would hand over driving and then, the participants would press the button *when comfortable to enable the autonomous mode*. During the autonomous mode, the participants played an online game, 2048 (Figure 2) on a laptop placed next to them on the center console (Figure 1). They were instructed not to look at the driving simulator while playing the game. After about one minute, the same auditory cue was presented for the participants to take over control of the vehicle due to an on-road situation (Figure 3).

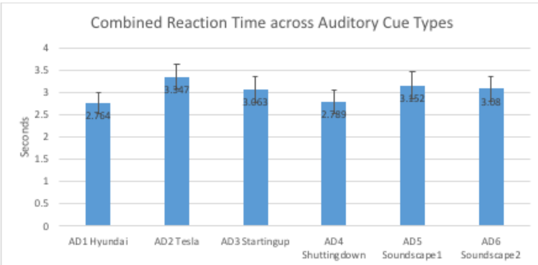

Figure 4: Combined reaction time across different auditory cue types.

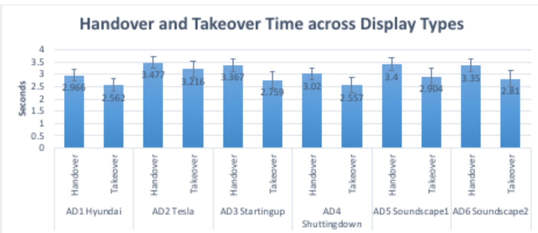

Figure 5: Reaction time for both handover and takeover tasks across different auditory cue types.

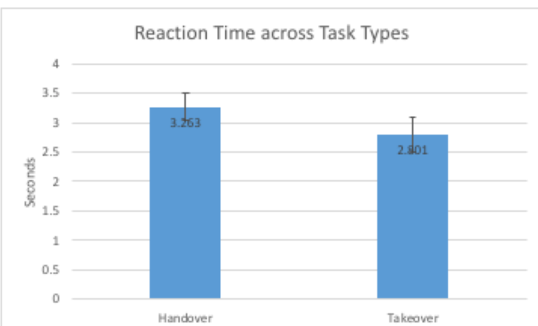

Figure 6: Reaction time across task types.

The participants would then press the button *when comfortable to disable autonomous mode*. The time measured was from the moment the auditory display was presented to the moment the participants disabled the autonomous mode by pressing the button. Shortly after, another auditory display would play indicating a handover event again. The participant would then press the button when comfortable to enable the autonomous mode again. This process continued for three obstacles (takeover and handover) in the scenario. Each driving scenario lasted roughly 4-5 minutes. Once the scenario was completed, participants completed subjective questionnaires. This process was repeated for the six conditions. After all scenarios were completed, the participants completed a post demographic questionnaire. Finally, the participants were debriefed about the experiment.

3 RESULTS

3.1 Take-over Time

Figure 4 shows combined mean reaction time of both handover and takeover for each display type. Results were analyzed with a 6 (Display) x 2 (Task) repeated measures analysis of variance (ANOVA), which revealed a statistically significant difference among display types in mean reaction time, $F(2.80, 61.68) = 3..88$, $p < .05$, $\eta_p^2 = .15$. (using Greenhouse-Geisser due to the Sphericity violation) (Figure 5) and task types, $F(1, 22) = 11.63$, $p < .01$, $\eta_p^2 = .35$ (Figure 6). There was no significant interaction between Display and Task, $F(3.50, 76.77) = 1.10$, $p > .05$, $\eta_p^2 = .17$. For the multiple comparisons among display types, we conducted paired-samples t-tests. 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 = .0033 (0.05/15)).

Participants showed significantly faster handover in AD1 Hyundai ($M = 2.966$, $SD = 1.02$) than AD2 Tesla ($M = 3.477$, $SD = 1.16$), $t(22) = -5.665$, $p < .001$. Participants showed significantly faster handover in AD1 Hyundai than AD5 Soundscape1 ($M = 3.400$, $SD = 1.27$), $t(22) = -3.753$, $p = .001$. Participants also showed significantly faster handover in AD4 ShuttingDown ($M = 3.020$, $SD = 1.19$) than AD2 Tesla, $t(22) = 3.903$, $p = .001$. In a similar way, participants showed significantly faster takeover in AD1 Hyundai ($M = 2.562$, $SD = 1.24$) than AD2 Tesla ($M = 3.216$, $SD = 1.59$), $t(22) = -3.323$, $p = .003$. Participants also tended to show faster takeover in AD4 ShuttingDown ($M = 2.557$, $SD = 1.48$) than AD2 Tesla, $t(22) = 3.064$, $p = .005$, but this did not lead to the statistically significant level due to the conservative alpha. In sum, AD1 Hyundai and AD4 ShuttingDown sounds led to faster reaction time in both handover and takeover tasks (Figure 5).

3.2 Sound User Experience Questionnaire

For sound user experience questionnaire, results were analyzed with a 6 (Display type) repeated measures ANOVA. For preference as handover and takeover displays, there was no statistically significant difference among display types. However, for takeover AD1 Hyundai received the numerically highest score among the display types and AD5 Soundscape1 received the lowest score (Figure 7). For handover AD4 ShuttingDown received the numerically highest score among the display types and again, AD5 Soundscape1 received the lowest score (Figure 8). ANOVA revealed a statistically significant difference among display types in mean “pleasing” rating score, $F(5, 105) = 6.194$, $p < .001$, $\eta_p^2 = .228$. The multiple comparisons using paired-samples t-tests are in

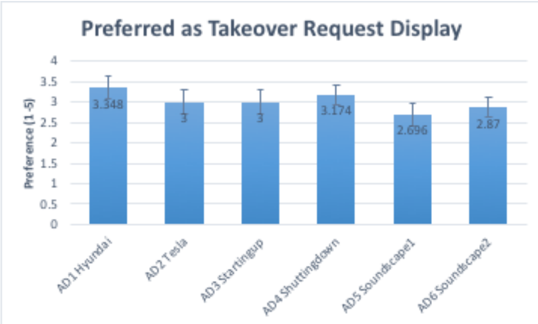

Figure 7. Preference rating as takeover request display

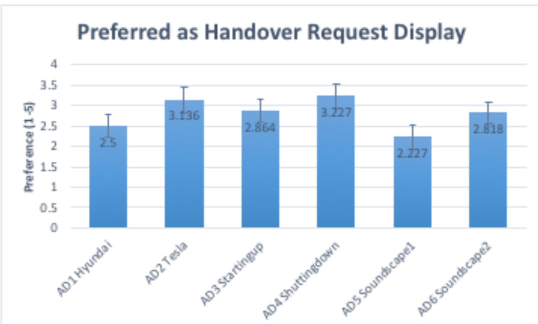

Figure 8. Preference rating as handover request display

Table 2: Pleasing rating results

AD2 Tesla > AD1 Hyundai $t(22) = -3.994$, $p = .001$
AD3 StartingUp > AD1 Hyundai $t(22) = -5.061$, $p < .001$
AD6 Scape2 > AD 1 Hyundai $t(22) = -6.282$, $p < .001$

Table 3: Annoying rating results

AD1 Hyundai > AD3 StartingUp $t(22) = 3.710$, $p = .001$
AD1 Hyundai > AD6 Scape2 $t(22) = 4.542$, $p < .001$

Table 2. In sum, AD1 Hyundai is significantly less pleasing than other auditory displays. ANOVA revealed a statistically significant difference among display types in mean "annoying" rating score, $F(5, 110) = 3.711$, $p = .004$, $\eta_p^2 = .144$. The multiple comparisons are shown in Table 3. In sum, AD1 Hyundai is significantly higher annoying than other auditory displays. ANOVA revealed a statistically significant difference among display types in mean "urgent" rating score, $F(5, 110) = 14.589$, $p < .001$, $\eta_p^2 = .399$. The multiple comparisons are shown in Table 4. In sum, AD 1 Hyundai and AD 4 ShuttingDown show significantly higher urgent than other auditory displays. ANOVA revealed a statistically significant difference among display types in mean "attention capturing" rating score, $F(5, 105) = 6.490$, $p < .001$, $\eta_p^2 = .236$. The multiple comparisons are shown in Table 5. Likewise, AD 1 Hyundai and AD 4 ShuttingDown show significantly higher attention capturing than other auditory displays. Finally, ANOVA revealed a statistically significant difference among display types in mean "intuitive" rating score, $F(5, 105) = 4.419$, $p = .001$, $\eta_p^2 = .174$. The multiple comparisons are shown in Table 6. Only AD1 Hyundai showed significantly higher intuitive score than AD5 Soundscape1. AD4 ShuttingDown also showed the second higher score, but it did not reach a statistically significant level due to the conservative alpha level.

4 DISCUSSION

We evaluated handover and takeover times and subjective measures for six types of auditory displays for the semi-automated vehicle in the presence of an engaging secondary task (online game). Takeover time was significantly faster than handover time. This makes sense because takeover is usually a more urgent situation and so participants responded accordingly. The results indicate that the difference of decision making between takeover and handover is less than 500 ms (around 440 ms). The results clearly showed that the traditional warning sound (AD1 Hyundai sound) with high frequency and fast repetition makes faster reaction time and the participants' subjective responses about sound profiles supported the results. Participants also preferred this sound as a takeover sound (Figure 7). Among the new auditory displays, AD4 ShuttingDown sound showed faster reaction times in both handover and takeover tasks. The participants preferred this design as the handover display (Figure 8). The subjective measures showed consistent results. Participants rated AD1 Hyundai sound as least pleasing, most annoying, most urgent, most attention capturing, and most intuitive. Even though the most annoying sound made fastest reaction time, auditory researchers claim that alerts do not necessarily need to be annoying [5]. Therefore, future research should look at how to design urgent and intuitive auditory displays that are not too annoying. AD4 ShuttingDown sound showed promise to be such a sound. This sound showed the fastest reaction time for takeover and the second fast reaction time for handover. Moreover, this sound received the highest preference score for the handover request and the second highest preference score for the takeover request. AD4 ShuttingDown sound showed at least, less annoying than AD1, but also less urgent. These results may indicate that this sound would fit best in a slightly less urgent scenario, where attention is still required, such as handover. It was also of interest to test how participants perceived a pair of starting up and shutting down style sounds for handover and takeover scenarios. We can cautiously infer that the participants perceived the shutting down style sound as indicative of an end of task period such as handing

Table 4: Urgent rating results

AD1 Hyundai > AD2 Tesla $t(22) = 5.147$, $p < .001$
AD1 Hyundai > AD3 StartingUp $t(22) = 6.160$, $p < .001$
AD1 Hyundai > AD4 ShuttingDown $t(22) = 3.990$, $p = .001$
AD1 Hyundai > AD5 Scape1 $t(22) = 8.290$, $p < .001$
AD1 Hyundai > AD6 Scape2 $t(22) = 7.365$, $p < .001$
AD4 ShuttingDown > AD5 Scape1 $t(22) = 4.526$, $p < .001$
AD4 ShuttingDown > AD6 Scape2, $t(22) = 3.286$, $p = .003$

Table 5: Attention-Capturing rating results

AD1 Hyundai > AD2 Tesla, $t(21) = 4.405$, $p < .001$
AD1 Hyundai > AD3 StartingUp $t(22) = 4.255$, $p < .001$
AD1 Hyundai > AD5 Scape1 $t(22) = 5.127$, $p < .001$
AD1 Hyundai > AD6 Scape2 $t(22) = 4.800$, $p < .001$
AD4 ShuttingDown > AD5 Scape1, $t(23) = 3.715$, $p = .001$

Table 6: Intuitive rating results

AD1 Hyundai > AD5 Sacpe1 $t(21) = 3.697$, $p = .001$

ACKNOWLEDGMENTS

This work was partially supported by a grant (code 17TLRP-B131486-01) from Transportation and Logistics R&D Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

over the driving task back to the vehicle. Note that not all existing auditory displays did perform similarly. AD2 Tesla showed the worst reaction time for both handover and takeover tasks among the six auditory cue types even though our participants rated it as the second highest for a handover request display. Unsurprisingly, soundscapes were rated the most pleasing and least annoying. However, these sounds did not sufficiently provide a perception of urgency and attention-capturing. Thus, these sounds were not preferred as either the takeover or the handover request display. Participants seemed to perceive that this type of soundscape is not appropriate for functional displays in urgent situations.

There were interesting anecdotal observations during the study. Some participants waited for the sound to be completed before pressing the button, resulting in a crash into the obstacle. This is why we should keep the auditory displays short. Likewise, during the takeover event, some participants continued to play the game for several seconds before coming back to the driving situation. This is suggestive of designing auditory displays that are not only attention-capturing, but also annoying to a certain extent. This ensures that drivers can quickly perceive the urgency of a situation and react to stop the sound. Some participants were aversive to handover control during a road turn. This is an interesting observation regarding trust issues. Since our participants were not sure about the quality of the self-driving, they wanted to drive for themselves for the turn. These types of on-road variables and handover/takeover timings should be considered in the semi-automated vehicles so drivers can comfortably change their control status. One of the existing auditory displays showed the best performance and highest preference. However, it does not mean that it is literally the best. To make a better alternative design, more research is required varying other acoustic parameters (interval, loudness, etc.).

This research can also be formalized with more systematic approaches. We are currently modeling drivers' behavior in handover and takeover situations with varying different parameters, including display parameters (lead time, display type, length of display, etc.), road conditions, driver states, etc. Once we build a systematic model, we will minimize running a human subject study, but we can run simulation with updated parameters, thereby reducing time and simplifying future research.

REFERENCES

[1] Christopher D. Wickens. 2002. Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3, 2: 159-177.

[2] Denis McKeown. 2005. Candidates for within-vehicle auditory displays. Georgia Institute of Technology.

[3] Erin Richie, Thomas Offer-Westort, Raghavendran Shankar, and Myounghoon Jeon. 2018. Auditory displays for take-over in semi-automated vehicles. Proceedings of the 20th HCII2018 Conference.

[4] Chin-Teng Lin, Teng-Yi Huang, Wen-Chieh Liang, Tien-Ting Chiu, Chih-Feng Chao, Shang-Hwa Hsu, and Li-Wei Ko. 2009. Assessing effectiveness of various auditory warning signals in maintaining drivers' attention in virtual reality-based driving environments. Perceptual and motor skills 108, 3: 825-835.

[5] Myounghoon Jeon, Maryam FakhrHosseini, Eric Vasey, and Michael A. Nees. 2017. Blueprint of the auditory interactions in automated vehicles: Report on the workshop and tutorial. Adjunct Proceedings of the 9th AutomotiveUI Conference.