Multimodal Displays for Take-over in Level 3 Automated Vehicles while Playing a Game

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

ABSTRACT

Take-over is one of the most crucial user interactions in semi-automated vehicles. To make better communication between driver and vehicle, research has been conducted on various take-over request displays, yet the potential has not been fully investigated. The present paper explored the effects of adding auditory displays to visual text. Earcon and speech showed the best performance

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

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313056

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

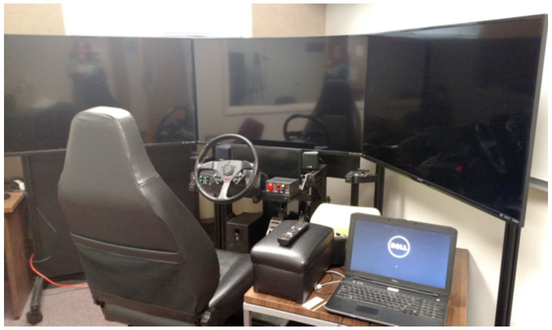
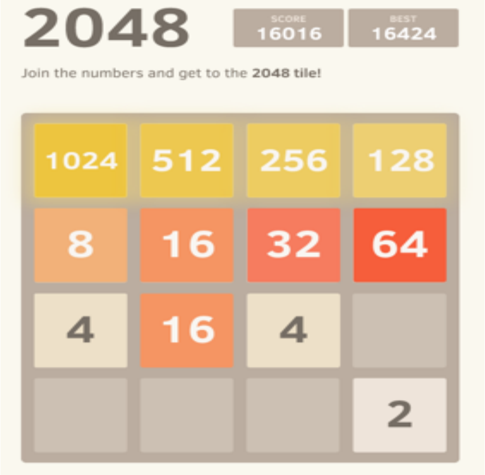


Figure 1: NADS MiniSim Driving Simulator.



2048

SCORE 16016

BEST 16424

Join the numbers and get to the **2048 tile!**

1024	512	256	128
8	16	32	64
4	16	4	
			2

Figure 2: 2048 Game as a secondary task.

and acceptance with spearcon the least. This study is expected to provide the basic data and guidelines for future research and design practice.

1 INTRODUCTION

Automated vehicles have been appearing more on the road. Advanced sensing and computing technologies mainly drive this momentum, but human factors also need to be considered to design better interaction between a driver and a vehicle. One of the most critical points for human-automation collaboration in automated vehicles is a take-over scenario. Level 3 automated vehicles have been already equipped with take-over displays and there has been some preliminary research [1,2], but more research is still required to validate and optimize the display design and make an international standard. Multiple Resources Theory [3] predicts using different modalities other than a single modality would be more effective for doing multiple tasks concurrently as in driving. Thus, auditory displays are widely used in the vehicle context. Researchers have made novel auditory cues that can be used for this type of urgent alert, but those cues have not been fully applied and tested. The present paper evaluated the effects of adding different auditory displays – speech, spearcon [4], and earcon [5], to visual text for the take-over request while a driver is playing an online game, compared to text-only. Both take-over time and subjective experiences were measured.

2 METHODS

2.1 Participants

Forty-four undergraduate students (14 female; *Mean* age = 20, *SD* = 1.7; *Mean* years of driving = 4.0 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 four take-over displays. Male voice, “take over” was recorded for speech. For spearcon, the wave file of the speech clip went through the spearcon generation program using the SOLA algorithm [6]. Earcon was designed, including two dominant frequencies (880, 1760 Hz) repeated four times using sine wave, following NHTSA [7] and ISO [8] guidelines. The volume and length of the sounds were controlled to be equivalent (*Mean* = 70dB, 311ms). Text, “Please take over” was displayed on the center monitor of the driving simulator in white.

2.3 Apparatus

The simulator used was a mid-fidelity National Advanced Driving Simulator (NADS) MiniSim (Figure 1). 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 and the auditory displays were played through two embedded speakers. For playing the 2048 game, a Dell laptop was used with four-arrow keys on a keyboard to slide numbered tiles to create a tile (Figure 2).

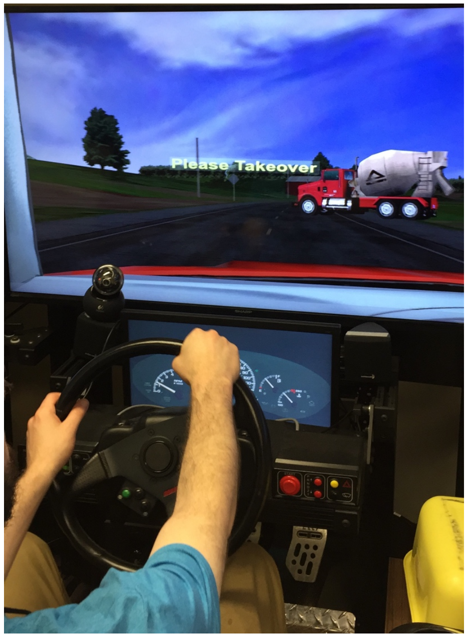


Figure 3: Experimental Scene with the simulator.

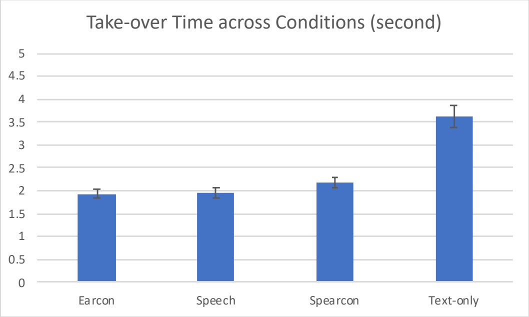


Figure 4: Mean of take-over time for each display type.

2.4 Design and Procedure

This experiment used a within-subjects design with four display conditions (earcon+text, speech+text, spearcon+text, and text-only). 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 was no oncoming traffic and there was no intersection. The speed limit was set at 50 miles per hour. Situations on the road that triggered the participant to take over included deer, a parked car, and a service vehicle, all blocking most or all of the driving lanes (Figure 3).

After the consent form procedure, participants were given an overview of the experiment and learned how the driving simulator worked. When the participants felt ready, the experiment began with one of the randomly selected scenarios. After a short driving, a ding noise signaled participants that the vehicle would take over driving and then, participants had to take their hands off the steering wheel and their foot off the gas pedal. During the time the vehicle self-drove, the participants played the game, 2048 (Figure 2) on a laptop placed next to them on the center console (Figure 1). After a while, one of the four take-over displays was presented for the participants to take over control of the vehicle due to a situation on the road. The time was measured from the moment the display was presented to the moment the participants took over control of the vehicle (grabbing the steering wheel). A few seconds after avoiding (or crashing into) the hazard on the road, the car made the ding noise again, which told the participants it would take over control of driving once again. The participants then returned to playing 2048 as instructed by the researcher. This process continued for three obstacles in the scenario. Once the scenario was completed, participants completed the subjective questionnaire. This process was repeated for the four conditions. After all four scenarios were completed, the participants completed a demographic questionnaire.

3 RESULTS

Literature [9] identified objective measures of the sound-relevant research as efficiency (reaction time), accuracy (number of errors), and learning rate (as a function of time). In the present study, we measured all of these objective measures. The number of crashes reflects the accuracy of the display or errors caused from the display. Take-over time reflects efficiency and take-over times across the three laps reflect the learning rate of each display. We also used a survey to explore user experience, which influenced driving performance and provided practical guidelines for take-over display design.

3.1 Number of Crash

In total, eleven participants made 17 crashes. As expected seven participants made crashes in the text-only condition. No participants made a crash in the earcon+text condition. For both speech+text condition and spearcon+text condition, two participants each made crashes. These cases were not sufficient to make inferential statistics, but it provided a sense of where participants made errors and where they did not miss the take-over display. Take-over time was analyzed with only successful take-over cases without crashes.

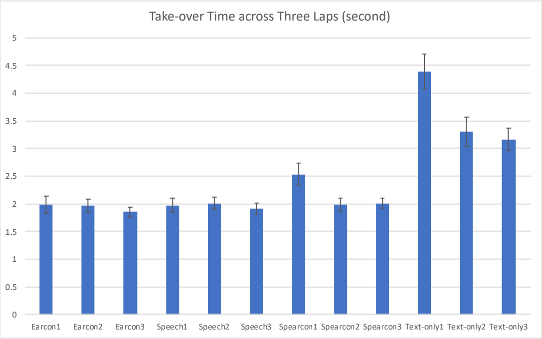


Figure 5: Take-over time across three laps.

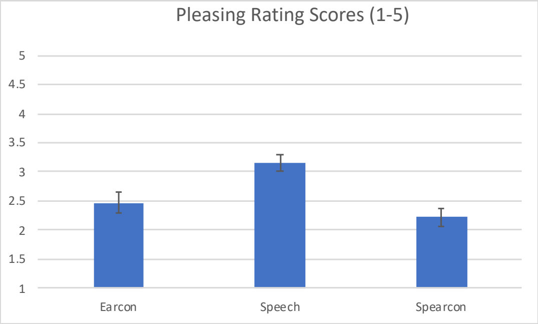


Figure 6: Pleasing rating scores across conditions.

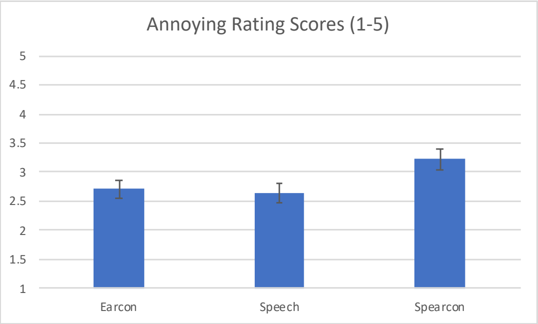


Figure 7: Annoying rating scores across conditions.

3.2 Take-over Time

In the driving task, reaction time is not just relevant to efficiency, but also directly relevant to road safety. Figure 4 shows overall mean take-over time for each display type.

Results were analyzed with a 4 (Display type) x 3 (Lap) repeated measures analysis of variance (ANOVA), which revealed a statistically significant difference among display types in mean take-over time, *F*(1.25, 39.84) = 63.61, *p* < .001, η_p^2 = .67 (using Greenhouse-Geisser due to the Sphericity violation). Lap also showed a statistically significant difference, *F*(1.32, 42.23) = 9.26, *p* = .002, η_p^2 = .22. There was a statistically significant interaction between Display type and Lap, *F*(2.29, 73.29) = 6.17, *p* = .002, η_p^2 = .16. Figure 5 shows that text-only and spearcon+text have rapidly decreasing patterns between Lap 1 and Lap 2 but earcon+text and speech+text show a gentle slope across Laps. This interaction was further validated in the learning rate analysis below. 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 a more conservative alpha level (critical alpha level: 0.05/6 pairs = .00833). Participants got back to drive significantly faster in all of the multimodal display conditions than the text-only condition. Take-over time in the text-only condition (*M* = 3.61, *SD* = 1.53) was significantly slower than that in the earcon+text condition (*M* = 1.96, *SD* = 0.74), *t*(40; there are three missing data points for text-only) = 9.14, *p* < .001, the speech+text condition (*M* = 1.99, *SD* = 0.74), *t*(40) = 9.63, *p* < .001, and the spearcon+text condition (*M* = 2.17, *SD* = 0.78), *t*(40) = 7.58, *p* < .001. In addition, the spearcon+text condition was significantly slower than the earcon+text condition, *t*(43) = -3.11, *p* = .003 and the speech+text condition, *t*(43) = 2.86, *p* = .007. In short, all auditory displays decreased the take-over time compared to visual-only display and earcon and speech even showed faster take-over time than spearcon.

To check the learning rate of each display type, we conducted repeated measures ANOVA (Figure 5). Text-only showed a significant difference among the Laps, *F*(1.39, 48.54) = 12.71, *p* < .001, η_p^2 = .27 (using Greenhouse-Geisser due to the Sphericity violation). Take-over time of Lap 1 in the text-only condition (*M* = 4.36, *SD* = 2.13) was significantly slower than that in Lap 2 (*M* = 3.30, *SD* = 1.84), *t*(35; there are 8 missing data points) = 3.85, *p* < .001 and in Lap 3 (*M* = 3.19, *SD* = 1.41), *t*(35) = 3.75, *p* = .001. However, Lap 2 and Lap 3 did not show the statistical difference.

Spearcon+text also showed the same pattern. There was a significant difference among the Laps, *F*(1.33, 54.34) = 11.37, *p* < .001, η_p^2 = .22 (using Greenhouse-Geisser due to the Sphericity violation). Take-over time of Lap 1 in the spearcon+text condition (*M* = 2.53, *SD* = 1.28) was significantly slower than that in Lap 2 (*M* = 1.99, *SD* = 0.77), *t*(43) = 3.61, *p* = .001 and in Lap 3 (*M* = 2.02, *SD* = 0.76), *t*(41; there are two missing data points) = 3.75, *p* = .001. However, Lap 2 and Lap 3 did not show the statistical difference. There were no significant differences between Laps for the earcon+text condition and speech+text condition.

3.3 Sound User Experience Questionnaire

- Pleasing: Results were analyzed with a 3 (Sound types) repeated measures ANOVA, which revealed a statistically significant difference among Sound types in mean “pleasing” rating score, *F*(2, 80) = 8.43, *p* < .001 (Figure 6). Paired-samples t-tests showed that there were statistically

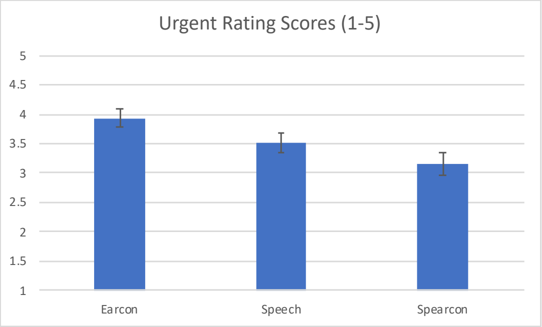


Figure 8: Urgent rating scores across conditions.

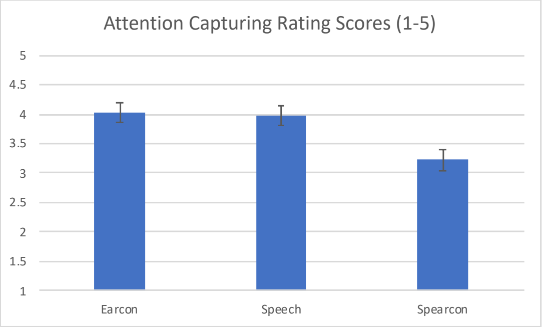


Figure 9: Attention capturing rating scores across conditions.

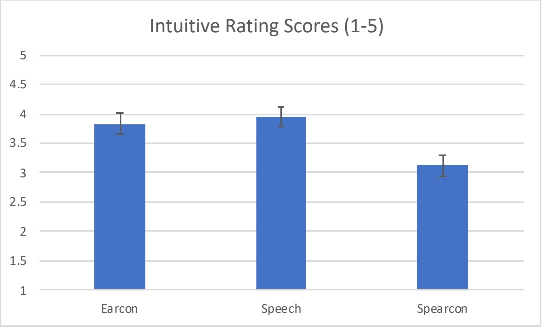


Figure 10: Intuitive rating scores across conditions.

significant differences between speech (*M* = 3.15, *SD* = 0.94) and spearcon (*M* = 2.22, *SD* = 1.04), *t*(40) = -3.82, *p* < .001 and speech and earcon (*M* = 2.46, *SD* = 1.23), *t*(40) = -3.29, *p* = .002.

- Annoying: Repeated measures ANOVA showed a statistically significant difference among Sound types in mean "annoying" rating score, *F*(2, 80) = 3.16, *p* < .05 (Figure 7). Paired-samples *t*-tests showed that there was a statistically significant difference between speech (*M* = 2.63, *SD* = 1.10) and spearcon (*M* = 3.22, *SD* = 1.18), *t*(40) = -2.50, *p* = .00832. Earcon (*M* = 2.70, *SD* = 1.04) and spearcon (*M* = 3.22, *SD* = 1.18), *t*(40) = -1.85, *p* = .036 tended to be different, but not significantly different.
- Urgent: ANOVA showed a statistically significant difference among Sound types in mean "urgent" rating score, *F*(2, 80) = 5.17, *p* < .001 (Figure 8). Paired-samples *t*-tests showed that there was a statistically significant difference between earcon (*M* = 3.93, *SD* = 1.02) and spearcon (*M* = 3.15, *SD* = 1.28), *t*(40) = 2.92, *p* = .00285.
- Attention capturing: Repeated measures ANOVA showed a statistically significant difference among Sound types in mean "attention capturing" score, *F*(2, 80) = 6.68, *p* < .05 (Figure 9). Paired-samples t-tests showed that there were statistically significant differences between earcon (*M* = 4.02, *SD* = 1.09) and spearcon (*M* = 3.22, *SD* = 1.14), *t*(40) = -3.73, *p* < .001 and between speech (*M* = 3.98, *SD* = 1.12) and spearcon, *t*(40) = -2.78, *p* = .0042.
- Intuitive: Repeated measures ANOVA showed a statistically significant difference among Sound types in mean "intuitiveness" rating score, *F*(2, 80) = 8.94, *p* < .001 (Figure 10). Paired-samples t-tests showed that there were statistically significant differences between earcon (*M* = 3.83, *SD* = 1.15) and spearcon (*M* = 3.12, *SD* = 1.17), *t*(40) = -3.41, *p* < .001 and between speech (*M* = 3.95, *SD* = 1.17) and spearcon, *t*(40) = -3.67, *p* < .001.
- Startling: Spearcon showed the highest "startling" rating score, but there was no statistically significant result.
- Commanding: Earcon showed the highest "commanding" rating score, but there was no statistically significant result.
- Would Turn off this audio in my own vehicle: Repeated measures ANOVA showed a statistically significant difference among Sound types in mean "would turn off" score, *F*(2, 80) = 4.41, *p* < .05 (Figure 11). Paired-samples t-tests showed that there was a statistically significant difference between speech (*M* = 2.46, *SD* = 1.06) and speearcon (*M* = 3.20, *SD* = 1.27), *t*(40) = -3.19, *p* < .0027.

4 DISCUSSION

We evaluated take-over time and subjective measures for four types of take-over displays for semi-automated vehicles in the presence of an engaging secondary task (online game). The results showed that adding the auditory cues for take-over displays can significantly improve take-over time compared to the visual-only (text) display. However, not all auditory displays are similarly applicable. Among the auditory displays, the earcon and the speech conditions showed the best performance (lowest take-over time) from the first lap when combined with visual text, indicating that no learning is required for these displays. In contrast, the spearcon+text and text-only conditions showed significantly worse takeover times. Their performance was enhanced per laps,

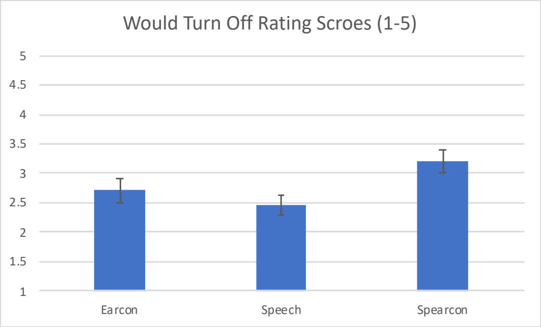


Figure 11: Would turn off rating scores across conditions.

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.

but this means that these displays require users' learning. Based on the data from the subjective questionnaire, we can infer the application directions of multimodal take-over displays. Among the cues, the speech display was rated the most pleasing and least turned off. As expected, earcon showed highest urgency and commanding effect compared to other displays. Both speech and earon was rated as more attention-capturing and intuitive than spearcon. Because spearcon is compressed speech, it was expected to show higher urgency, but it did not show such a trend. Spearcon showed the highest annoying, startling, and would turn off rating scores. Its annoying and startling ratings seemed to lead to faster reaction time than text-only, but it was still slower than earcon or speech. Research shows that the auditory warning should include urgency, but it could/should avoid startling [10]. More importantly, spearcon showed the highest rating score for "would turn off this display". This result is in line with literature showing that aesthetic and annoyance issues are more important in auditory displays than in visual displays [11]. In sum, earcon and speech seem to work best when combined with text for take-over displays, guaranteeing performance and user experience. More research can be done with specific parameters of earcons (e.g., pulse duration and interburst interval) and speech (e.g., gender and tone) with additional measures.

REFERENCES

[1] Vivien Melcher, Stefan Rauh, Frederik Diederichs, Herald Widlroither, and Wilhelm Bauer. 2015. Take-over requests for automated driving. Procedia Manufacturing, 3, 2867-2873.

[2] Ioannis Politis, Stephen Brewster, and Frank Pollick. 2015. 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 (pp. 3-10). ACM. http://doi.acm.org/10.1145/2799250.2799262

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

[4] Bruce N. Walker, Jeffrey Lindsay, Amanda Nance, Yoko Nakano, Dianne K. Palladino, Tilman Dingler, and Myounghoon Jeon. 2013. Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. Human Factors 55, 1: 157-182.

[5] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and icons: Their structure and common design principles. Human–Computer Interaction 4, 1: 11-44.

[6] Salim Roucos and Alexander Wilgus. 1985. High quality time-scale modification for speech. In Proceedings of the Acoustics, Speech, and Signal Processing, vol. 10, pp. 493-496. doi:10.1109/ ICASSP.1985.1168505

[7] John L. Campbell, James, L. Brown, Justin S. Graving, Christian M. Richard, Monica G. Lichty, Thomas Sanquist,...and Justin F. Morgan. 2016. Human factors design guidance for driver-vehicle interfaces (Report No. DOT HS 812 360). Washington, DC: National Highway Traffic Safety Administration.

[8] Draft, ISO Working. Road vehicles-Ergonomic aspects of transport information and control systems-Specifications for in-vehicle auditory presentation," ISO Standard 15006:2011(E).

[9] Myounghoon Jeon. 2015. Auditory user interface design: Practical evaluation methods and design process case studies. The International Journal of Design in Society, 8, 2: 1-16.

[10] Judy Edworthy, Scott Reid, Siné McDougall, Jonathan Edworthy, Stephanie Hall, Danielle Bennett, James Khan, and Ellen Pye. 2017. The recognizability and localizability of auditory alarms: Setting global medical device standards. Human factors, 59, 7: 1108-1127.

[11] Stephen Brewster. 2008. Chapter13: Nonspeech auditory output. In A. Sears and J. Jacko (Eds.), The human computer interaction handbook. New York: Lawrence Erlbaum Associates (pp. 247-264).