Novel Auditory Displays in Highly Automated Vehicles: Sonification Improves Driver Situation Awareness, Perceived Workload, and Overall Experience

Chihab Nadri[1], Sangjin Ko[1], Colin Diggs[1], Michael Winters[2], Sreehari V. K.[1], and Myounghoon Jeon[1]
1 – Virginia Polytechnic Institute and State University, 2 – Georgia Institute of Technology

Highly automated driving systems are expected to require the design of new user-vehicle interactions. Sonification can be used to provide contextualized alarms and cues that can increase situation awareness and user experience. In this study, we examined user perceptions of potential use cases for level 4 automated vehicles in online focus group interviews (*N*=12). Also, in a driving simulator study, we evaluated (1) visual-only display; (2) non-speech with visual display; and (3) speech with visual display with 20 young drivers. Results indicated participants' interest in the use cases and insight on desired functions in highly automated vehicles. Both audiovisual display conditions resulted in higher situation awareness for drivers than the visual-only condition. Some differences were found between the non-speech and speech conditions suggesting benefits of sonification for both driving and non-driving related auditory use cases. This study will provide guidance on sonification design for highly automated vehicles.

INTRODUCTION

Automation technology is changing the way drivers interact with their vehicles. As technology advances, users of automated vehicles will increasingly disengage with the driving task and perform non-driving tasks. Highly automated vehicles, defined by SAE International as level 4 vehicles (SAE Committee, 2014), are expected to reduce driver stress and increase productivity (Litman, 2020). Challenges exist in the adoption and acceptance of higher levels of autonomy in vehicles, revolving around safety concerns with the vehicles, and can be addressed through novel interactive displays for safe transitions in automation (Jeon, 2019), such as robot agents (Lee, Ko, Sanghavi, & Jeon, 2019) or augmented reality displays (von Sawitzky, Wintersberger, Riener, & Gabbard, 2019). Furthermore, researchers have suggested the use of adaptive auditory alerts (Šabić, Henning, & MacDonald, 2019), spatial sound (Petermeijer, Bazilinskyy, Bengler, & de Winter, 2017), or other auditory displays as ways to quickly provide feedback for drivers to takeover or respond to different vehicle states.

Sonification, which is transcribing data into non-speech sound (Nees & Walker, 2011) is a display method that has been suggested for automated vehicles (e.g., driving data sonification) (Landry, Jeon, FakhrHosseini, & Tascarella, 2016). This display method has been used to transcribe vehicle state, intentions, or user emotions to increase situation awareness (SA) (Gang et al., 2018; Landry et al., 2016). However, this method has not yet been evaluated for the operation of level 4 automated vehicles. Furthermore, it is also important to assess occupant interest in sonification approaches and identify occupant preferences.

Attitudes and user information requirements for automated driving (Hock, Kraus, Walch, Lang, & Baumann, 2017; Lee, Nadri, Sanghavi, & Jeon, 2020) indicate an interest in receiving information about the vehicle's state, intentions, or environmental information. The use of sonification could be extended to both driving and non-driving related tasks and activities as an additional display modality to improve occupant experience, trust, and acceptance of automated vehicle technology.

We present in this paper the results of focus group sessions and a preliminary driving simulation study aimed at generating new use cases that can utilize sonification in different driving conditions and environments. The study also sought to compare this modality with alternate display conditions and provide insight into the effects of sonification for use cases developed for highly automated driving.

METHOD

We conducted online focus group interviews (FGI) and a driving simulator study with young drivers to evaluate the usability and effect of auditory alerts for a select set of use cases in level 4 automated vehicles.

Online Focus Group Interviews

Participants

For the FGI study, 12 participants were recruited (female = 7, male = 5) and three sessions were held. All FGI participants were university students in the United States, with an average age of 23.33 years old (*SD* = 3.52).

Procedure

Participants received a video conference link via Zoom to join the online FGI session. All participants provided consent for the study by signing an informed consent form approved by the Institutional Review Board (IRB) of the university.

At the start of the online session, the purpose of the FGI and seven use cases selected by researchers were presented along with brief examples and user journeys (Table 1).

Table 1: Use cases presented during the focus group interviews

Use Cases
1. Automation level change notification
2. Takeover/handover request
3. Live journey sonification
4. Virtual conference/video calls

Use Cases
5. Battery status alert
6. In-vehicle meditation
7. Network security notification

The seven use cases were generated following a review of relevant research for highly automated vehicles (Fakhrhosseini, Landry, Tan, Bhattarai, & Jeon, 2014; Gang et al., 2018; Hock et al., 2017; Landry et al., 2016). Auditory cues were generated for each of the seven use cases, with both speech and non-speech sounds. A male, native English speaker voice was used for speech cues. Non-speech included both earcons (short abstract sounds) (Blattner, Sumikawa, & Greenberg, 1989) and, in the case of the live journey sonification event, auditory icons (sounds that are related to their referent object, event, or process) (Gaver, 1986) in the form of natural soundscapes for different environments traversed (birds chirping for a forest, seagulls and waves crashing for a beach (Mauney & Walker, 2004)). For other use cases, earcons were designed through pitch mapping tones for related use cases (e.g., battery level, automation level), using relevant melodies and tones (e.g., virtual calls).

This was followed by a group discussion regarding the situations shown. Participants were then instructed to listen to alerts made for each use case. After completing an evaluation of sounds for each use case, a group discussion was briefly held to gather subjective feedback and user preferences. Throughout the FGI, one of the researchers took notes of the discussion. At the end of the FGI, participants discussed their favorite use cases in terms of perceived usefulness and preference.

Driving Simulator Study

Participants

For the driving simulator study, 20 participants were recruited (female = 6, male = 13, other gender identity = 1). Participants were university students and residents from the area with a valid driving license, with an average age of 24.3 years old (*SD* = 6.32). Only one participant had prior experience with automated vehicles. All participants provided consent.

Experimental design

A within-subject design was implemented with three conditions (Visual alert only (V), Non-speech with visual (N), Speech with visual (S)). In each of the three driving scenarios participants completed, seven similar events were encountered. The events were developed from use cases for highly automated vehicles (Table 2). Each use case developed during the online FGI was represented by an event, except for meditation. This use case was not included due to its longer auditory component and potential variations, and user responses in the FGI which indicated less interest in the use case compared to others shown.

Table 2: Journey events encountered in each driving scenario

Event list
1. Battery is full notification
2. Battery is low notification
3. Takeover request – Unexpected pedestrian crossing
4. Incoming call
5. Vehicle request for an increase in automation level
6. Vehicle request for a decrease in automation level
7. Live journey sonification – Notification of local environment (forest, beach)

The order of some events was fixed, with events 1 and 5 always appearing before events 2 and 6 respectively, while events 3, 4, and 7 appeared in a random order. The order of events reflected a single driving session with an initially charged vehicle that eventually loses battery and subsequently downgrades in automation level and was chosen as an ecologically valid succession of events. Sound conditions were randomized to minimize learning effects.

Apparatus and stimuli

A motion-based driving simulator (Nertech) was used in the study. Visuals were displayed on three 48" displays, and the simulator was equipped with a surrounding sound dome. During the simulated drive, participants received information from visual or audiovisual cues. Visual cues consisted of a visual icon displayed on the vehicle dashboard (Figure 1).

Figure 1: User view and dashboard in the driving simulator during the live journey event at a beach (left) and a forest (right)

Auditory cues were speech or non-speech messages depending on the scenario condition. The set of non-speech sounds used in the driving simulator study originated from non-speech cues previously created and evaluated during the FGI. Selection was based on perceived recognizability and user feedback during the FGI. For the live journey sonification event, natural sounds were used in the non-speech with visual condition. A speech message indicating the vehicle's current location was used in the speech with visuals condition (message in Table 3). All sounds lasted at most seven seconds per event.

Table 3: Auditory message component for driving events and conditions

Event list	Non-Speech cue	Speech cue
1. Battery is full notification	Six tone melody with an ascending pitch contour	"Battery is full"
2. Battery is low notification	Six tone melody with a descending pitch contour	"Battery is low"
3. Takeover request	Two dominant frequencies (880, 1760 Hz) repeated four times (Jeon, 2019)	"Please take control of the vehicle"
4. Incoming call	Typical phone ring sound with a six tone melody	"Incoming call"
5. Increase in automation level	Series of three notes increasing in pitch	"Increase in automation level"
6. Decrease in automation level	Series of three notes decreasing in pitch	"Decrease in automation level"
7. Live journey sonification – Beach - Forest	- Sound of waves crashing and seagulls - Sound of small birds chirping	"Passing by a beach" "Passing by a forest"

Procedure

Before the study started, participants provided consent for the study by signing an informed consent form approved by the Institutional Review Board (IRB) of the university. Participants initially had a short driving session to test for simulation sickness (Gable & Walker, 2013). Participants were familiarized with the driving simulator and were informed that they would complete three 12 minute trials. The vehicle was assumed to be a level 4 automated vehicle, and participants were told they could perform non-driving tasks and only needed to takeover control if the vehicle requested it. After each event happened (a second after the end of the auditory signal), researchers interrupted the driving simulation, which was paused and blocked out from participants' view, before querying participants about their perception, understanding, and future projection of the present situation, following the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1988). Participants then answered queries regarding the event for a 0 or 1 score for each SA component. Participants received a 1 score for all three event SA components if they (1) perceived the in-vehicle notification; (2) understood why the notification is happening; and (3) could predict subsequent events or actions that happen next. Participants could receive partial scores based on answering one or two SA components correctly. After completing each run, participants answered an adapted version of the Trust in Automated Systems scale (Jian, Bisantz, & Drury, 2000) for automated vehicles (α = .90) with a seven point Likert-scale. Participants filled out the electronic version of the NASA-TLX workload index at the end of each simulated run.

RESULTS

Online Focus Group Interview Results

FGI participants expressed their interest in use cases and ranked usefulness and preference (as seen in Table 4).

Table 4: Usefulness and preference choice of use cases

Usefulness	Display condition	
	Use case	Count
1st	Automation level change notification	7
2nd	Takeover/handover request	4
3rd	Battery status alert	1

Preference	Display condition	
	Use case	Count
1st	Live journey sonification	4
	Battery status alert	4
3rd	Takeover/handover request	3

The automation level change notification was chosen as the most useful use case by seven participants, who mentioned "It seems like the most important aspect of an automated car and would be important to my understanding" (P1 in FG1) and "it is essential to keep drivers concentrated" (P3 in FG3).

In terms of personal preference, four participants chose live journey sonification and battery status alerts respectively. Live journey sonification was said to be relaxing (P2 in FG2 and P2 in FG3), and "I really like this one as it provides an immersive experience I couldn't experience outside the car" (P2 in FG3). Battery status alerts were chosen for their helpfulness (P2 in FG1) and importance (P1 in FG2).

Driving Simulator Study Results

The non-parametric Kruskal-Wallis was used for SA scores and the Trust in Automated Systems scale, as the data did not follow a normal distribution. If a significant difference was found, a Wilcoxon signed-ranks test was used for pairwise comparisons, using a Bonferroni correction (with an adjusted $\alpha = 0.05/3 = 0.0167$). A repeated-measures ANOVA was used for NASA-TLX scores, and we ran post-hoc paired-samples t-tests using a Bonferroni correction (with an adjusted $\alpha = 0.05/3 = 0.0167$) for pairwise comparisons.

Table 5 presents results of the analysis of SA scores for the different display conditions. Significant differences were found between the different display conditions for all events, except for the decrease in automation event. Pairwise comparisons showed that the audiovisual conditions induced higher SA for battery alerts, takeover requests, and live trip sonification. The non-speech with visual condition resulted in significantly higher SA than speech with visual for the increase in automation event.

Table 5: Situation Awareness scores for different display conditions (V = visual-only, N = non-speech with visual, S = speech with visual)

Event	Display condition			Sig.	Pairwise comparisons
	V	N	S		
Full Battery	0.550	0.883	0.966	$p < 0.01$	V < N = S
Low Battery	0.633	0.883	0.883	$p < 0.01$	V < N = S
Takeover Request	0.733	0.950	1.000	$p < 0.01$	V < N = S
Incoming Call	0.866	0.983	0.950	$p = .031$	V < N
Increase in Automation	0.550	0.916	0.700	$p < 0.01$	V = S < N
Decrease in Automation	0.816	0.833	0.933	$p = .137$	V = N = S
Live Trip Sonification	0.650	0.883	0.933	$p < 0.01$	V < N = S

There were no statistically significant differences among the three conditions for trust scores. However, average scores for the non-speech with visual display ($M = 4.71$, $SD = 0.87$) and the speech with visual display ($M = 4.74$, $SD = 0.89$) were numerically higher than for the visual-only display ($M = 4.22$, $SD = 1.02$).

For the NASA-TLX subscale scores, a main effect for display type was found for physical demand $F(2, 38) = 3.623$, $p = 0.036$. The non-speech with visual condition ($M = 20.75$, SD = 13.50) resulted in significantly lower scores than the visual-only display ($M = 30$, $SD = 22.71$), $p = 0.012$. There were no statistically significant differences between the speech with visual display ($M = 26.75$, $SD = 17.57$) and the non-speech with visual display. A main effect for display type was found for effort $F(2, 38) = 5.525$, p < 0.01. The non-speech with visual display ($M = 31.75$, $SD = 17.27$) resulted in significantly lower scores than the visual-only display ($M = 51.25$, $SD = 25.80$), $p = 0.002$. There were no statistically significant differences between the speech with visual display ($M = 39$, $SD = 23.20$) and the non-speech with visual display. A main effect for display type was found for frustration $F(2, 38) = 4.005$, $p = 0.026$. The speech with visual display ($M = 30.25$, $SD = 19.57$) resulted in significantly lower scores than the visual-only display ($M = 46.5$, $SD = 29.11$), $p = 0.009$. There were no statistically significant differences between the non-speech with visual display ($M = 35.25$, $SD = 21.85$) and the speech with visual display. There were no statistically significant differences among the three conditions for mental demand, temporal demand, and performance.

DISCUSSION

We explored drivers' subjective evaluations on several potential use cases in highly automated vehicles. In the FGI, we were able to obtain user preferences and priorities in level 4 automated driving. Changes in automation level were perceived as the most important aspect of the drive, with participants placing more emphasis on planned rather than sudden changes in automation. Level 4 automated vehicles are expected to be self-driving in most conditions (SAE Committee, 2014; Litman, 2020) and inform drivers ahead of time to takeover. Thus, this observation and distinction between sudden and planned transitions in automation suggests that clear and precise information will need to be displayed ahead of time for users. This also falls in line with previous research identifying this as a key information need (Hancock et al., 2020; Lee et al., 2020) which could affect user trust, and in turn, acceptance of automated vehicle technology (Choi & Ji, 2015; Haspiel et al., 2018).

Trust scores in the driving simulator study did not significantly differ among display conditions, although audiovisual conditions (non-speech with visual, speech with visual) led to numerically higher rating scores than visual alert only. Additionally, SA scores further suggest that both the non-speech with visual and the speech with visual display conditions led to better user awareness of the driving environment, which is in line with previous studies about multimodal displays (Jeon, 2019; Liu, 2001; Petermeijer et al., 2017). Differences in score between the increase and decrease in automation events can be explained by a learning effect due to the order of the two events, as the decrease in automation event occurred after the increase in level for simulated runs.

In terms of perceived workload, participants experienced lower physical demand and effort scores for the non-speech with visual condition. This effect was statistically lower than the visual alert only condition and numerically lower than the speech with visual condition. This suggests that non-speech sounds elicited faster and more intuitive responses, because responding to the speech alert took more time due to its length and the need to process information received. This would fall in line with previous research on the use of auditory displays (Nees & Walker, 2011). This result may also indicate high learnability associated with our set of non-speech sounds, and the benefits of designing auditory displays that can sonify driving states and intent back to drivers. The use of both natural sounds and earcons in the non-speech set may have helped balance the learnability of earcons and auditory icons (Walker et al., 2013). The speech with the visual condition resulted in numerically lower frustration levels when compared to non-speech with visual and might reflect the ease of comprehension and learnability associated with speech (Dingler, Lindsay, & Walker, 2008).

In terms of the use of live journey sonification, participants expressed interest in the use case during both the FGI and the driving simulator study. Participants noted the novelty of the use case and were able to identify several applications for the use case (e.g., tourism, relaxation). This suggests that, in addition to its use for increasing situation awareness (Gang et al., 2018; Schoop, Smith, & Hartmann, 2018) or sonifying driver states and data (Landry et al., 2016), sonification could also be used as a tool in infotainment displays. While interest in the meditative use case was low, and there were experimental limitations with including this use case within the driving simulator study, the effect of music and sonification on driver performance has been shown in previous studies(Fakhrhosseini et al., 2014), so this use case may be important for driver comfort and should be pursued in future work. Limitations in the study include low participant numbers and learning effects associated with the order of events in each driving scenario. While the current study was limited in scope and did not include any emotional event in the present simulator study, music has already been shown an effective tool in reducing drivers' anger in automotive conditions (Fakhrhosseini et al., 2014). Future investigation of the use of sonification for meditative applications could provide more insight on user preferences and auditory modalities suitable for this use case.

CONCLUSION

In the present study, we explored the use of sonification for vehicle-occupant interactions for both safety and infotainment display situations in highly automated vehicles. Based on feedback from the FGI, participants expressed interest in the use of sonification and informed the choice of stimuli for a driving simulator study. It was also noted that there may be different design considerations and user interaction requirements between sudden and unexpected transition in automation. The driving simulator study indicated that both sounds improved driver situation awareness and that sonification may even provide advantages in terms of workload when compared to speech-based auditory displays. Future studies can address limitations present with the current study and are necessary to determine implications of the use of sonification for vehicle-occupant interactions.

REFERENCES

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human–Computer Interaction, 4*(1), 11-44.

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction, 31*(10), 692-702.

Dingler, T., Lindsay, J., & Walker, B. N. (2008). *Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech*. Proceedings of the International Community for Auditory Display.

Endsley, M. R. (1988). *Situation awareness global assessment technique (SAGAT).* Proceedings of the IEEE 1988 national aerospace and electronics conference.

Fakhrhosseini, S. M., Landry, S., Tan, Y. Y., Bhattarai, S., & Jeon, M. (2014). If you're angry, turn the music on: Music can mitigate anger effects on driving performance. *AutomotiveUI 2014 - 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, in Cooperation with ACM SIGCHI - Proceedings*(September). doi:10.1145/2667317.2667410

Gable, T. M., & Walker, B. N. (2013). *Georgia Tech Simulator Sickness Screening Protocol*. Retrieved from

Gang, N., Sibi, S., Michon, R., Mok, B., Chafe, C., & Ju, W. (2018). *Don't Be Alarmed: Sonifying Autonomous Vehicle Perception to Increase Situation Awareness*. Paper presented at the Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction, 2*(2), 167-177.

Hancock, P. A., Kajaks, T., Caird, J. K., Chignell, M. H., Mizobuchi, S., Burns, P. C., . . . Noy, I. Y. (2020). Challenges to human drivers in increasingly automated vehicles. *Human Factors, 62*(2), 310-328.

Haspiel, J., Du, N., Meyerson, J., Robert, L. P., Tilbury, D., Yang, X. J., & Pradhan, A. K. (2018). Explanations and Expectations: Trust Building in Automated Vehicles. *ACM/IEEE International Conference on Human-Robot Interaction*(Dc), 119-120. doi:10.1145/3173386.3177057

Hock, P., Kraus, J., Walch, M., Lang, N., & Baumann, M. (2017). Elaborating Feedback Strategies for Maintaining Automation in Highly Automated Driving. 105-112. doi:10.1145/3003715.3005414

Jeon, M. (2019). *Multimodal Displays for Take-over in Level 3 Automated Vehicles while Playing a Game.* In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71.

Landry, S., Jeon, M., FakhrHosseini, M., & Tascarella, D. (2016). Listen to Your Drive : An In-Vehicle Sonification Prototyping Tool for Driver State and Performance Data. *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '16)*, 21-26. doi:10.1109/CIMCA.2005.1631387

Lee, S. C., Ko, S., Sanghavi, H., & Jeon, M. (2019). Autonomous driving with an agent: Speech style and embodiment. *Adjunct Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*(September), 209-214. doi:10.1145/3349263.3351515

Lee, S. C., Nadri, C., Sanghavi, H., & Jeon, M. (2020). *Exploring User Needs and Design Requirements in Fully Automated Vehicles.* Paper presented at the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.

Litman, T. (2020). Autonomous vehicle implementation predictions: Implications for transport planning.

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveler information systems. *Ergonomics, 44*(4), 425-442. doi:10.1080/00140130010011369

Mauney, B. S., & Walker, B. N. (2004). *Creating functional and livable soundscapes for peripheral monitoring of dynamic data*. Proceedings of the International Community for Auditory Display.

Nees, M. A., & Walker, B. N. (2011). Auditory Displays for In-Vehicle Technologies. *Reviews of Human Factors and Ergonomics, 7*(1), 58-99. doi:10.1177/1557234X11410396

Petermeijer, S., Bazilinskyy, P., Bengler, K., & de Winter, J. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204-215. doi:10.1016/j.apergo.2017.02.023

Šabić, E., Henning, D., & MacDonald, J. (2019). Adaptive Auditory Alerts for Smart In-Vehicle Interfaces. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 1545-1549. doi:10.1177/1071181319631404

SAE Committee. (2014). Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. *SAE Standard J 3016*, 1-16.

Schoop, E., Smith, J., & Hartmann, B. (2018). *HindSight: Enhancing Spatial Awareness by Sonifying Detected Objects in Real-Time 360-Degree Video.* Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.

von Sawitzky, T., Wintersberger, P., Riener, A., & Gabbard, J. L. (2019). *Increasing trust in fully automated driving: route indication on an augmented reality head-up display.* Proceedings of the 8th ACM International Symposium on Pervasive Displays.

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors, 55*(1), 157-182. doi:10.1177/0018720812450587