INVESTIGATING THE EFFECT OF EARCON AND SPEECH VARIABLES ON HYBRID AUDITORY ALERTS AT RAIL CROSSINGS

Chihab Nadri[1], Seul Chan Lee[2], Siddhant Kekal[1], Yinjia Li[1], Xuan Li[1], Pasi Lautala[3], David Nelson[3], and Myounghoon Jeon[1]

Mind Music Machine Lab[1]
Virginia Polytechnic Institute
and State University,
Department of Industrial
and Systems Engineering,
1185 Perry Street Blacksburg,
VA 24061 USA
cnadri, skekal, yli0502,
leonxuan,
myounghoonjeon@vt.edu

Department of Industrial and
Systems Engineering /
Engineering Research Institute[2]
Gyeongsang National
University, Gyeongsangnam-do,
Jinju, Republic of Korea, 501
seulchan@gnu.ac.kr

Department of Civil and
Environmental Engineering[3],
Michigan Technological
University, Houghton, MI
49931, United States
ptlautal, dannelso@mtu.edu

ABSTRACT

Despite rail industry advances in reducing accidents at Highway Rail Grade Crossings (HRGCs), train-vehicle collisions continue to happen. The use of auditory displays has been suggested as a countermeasure to improve driver behavior at HRGCs, with prior research recommending the use of hybrid sound alerts consisting of earcons and speech messages. In this study, we sought to further investigate the effect of auditory variables in hybrid sound alerts. Nine participants were recruited and instructed to evaluate 18 variations of a hybrid In-Vehicle Auditory Alert (IVAA) along 11 subjective ratings. Results showed that earcon speed and pitch contour design can change user perception of the hybrid IVAA. Results further indicated the influence of speech gender and other semantic variables on user assessment of HRGC IVAAs. Findings of the current study can also inform and instruct the design of appropriate hybrid IVAAs for HRGCs.

1. INTRODUCTION

Train-vehicle collisions are a major issue in the US and across the world. While less frequent than vehicle-vehicle collisions, collisions at HRGCs are much deadlier due to the 4000-to-1 train-to-vehicle mass ratio [1]. Although the rail industry has introduced different warning devices at rail crossings, including but not limited to road markings, active warning devices (lights and gates), and passive warning devices (stop or yield signs) [2], different agencies still seek to improve safety at these locations. For instance, the European SAFER-LC project has evaluated HRGCs from the users' perspective and sought to take into account human error and other key criteria to improve safety at HRGCs [3]. In many of these initiatives, the inclusion of an auditory alert has been used to improve situation awareness and help alert drivers to an upcoming risk [4-6].

The use of auditory displays for preventing vehicle accidents has been well documented in past research, from frontal and directional collision warnings [7, 8], to intersection warnings [9-12], and automated vehicle takeover alerts [13-16]. In these driving situations, auditory display design plays an important role in determining the effectiveness of the warning display.

Research has found that sound manipulation techniques can also improve driving performance. Looming (increasing intensity of alarm over time) [17], radio sound level manipulation [18], and proper timing for alerts [19] can significantly reduce brake reaction time for drivers. Balancing loudness with semantics can also help driver performance [8], with a trade-off between perceived Urgency and Annoyance [20] requiring Urgency coding through Pulse Rates and other parameters [21]. The design of optimized auditory displays should consider these parameters to succeed at reducing accident rates at HRGCs.

Previously, we investigated the use of IVAAs for HRGCs through a combination of both subjective and simulator-driven experiments. In Landry et al. [22] auditory display types were evaluated through subjective ratings, and a hybrid alert was suggested as an appropriate IVAA for HRGC situations. As the study did not comprehensively explore the effect of auditory variables for earcon, speech, and hybrid alerts, another subjective assessment was conducted, and investigated the effect of auditory display variables for each display type [23]. From these studies we were able to extract the effect of a variety of acoustic variables. From our earcon results, we were able to determine the importance of earcon semitone range and pulse rate on IVAA ratings. We found that hybrid auditory alerts were associated with better hazard level identification, attention capturing ability, and created less desire to turn off than speech alerts. These results for auditory display types and variables are largely supported by previous research in related fields [24-27]. Lastly, we observed an interaction effect between participant gender and IVAA speech gender. Indeed, female participants appeared to have preferred male speech, rating male speech better in terms of commanding nature and hazard level, an effect which has not been observed in previous research [28].

However, limitations existed in our previous study. Speech and hybrid alert evaluation was constrained due to the use of TTS voice clips, and hybrid alerts were only varied based on speech content variables (speech length, speech rate, spatiality of audio), without investigating the effect of varying the hybrid alert earcon component. In this follow-up study, we seek to investigate some of these effects while using speech clips from native English speakers.

2. IVAA DESIGN

Following the findings of the previous subjective HRGC IVAA study [23], a further investigation on Hybrid alerts was conducted. Specifically, Hybrid IVAAs tested during this study were primarily manipulated based on earcon characteristics. The earcon component lasted around 0.9 s and was generated using the software Max/MSP to closely resemble an airplane intercom ding composed of two dings with a frequency of 523.25 Hz (two C5 notes). The speech component of the hybrid alert lasted around 4.6 s and only varied based on the speech gender of the alert. Speech clips were generated through recording native English speakers, unlike the previous experiment which used Text-To-Speech (TTS). All IVAAs used the same speech message “Slow down. Rail crossing ahead. Look left and right at crossing”, which was validated in a previous study [22].

As seen in Table 1, variables that were manipulated were earcon pitch contour (EP1: flat contour, EP2: ascending contour with the second note representing D5 with 589.25 Hz, EP3: descending contour with the second note representing an A#4 or 467.72 Hz; corresponding to changes of 3 semitones), earcon speed (ES1: regular speed, ES2: faster earcon by a factor of 25%, ES3: slower earcon by a factor of 25%), and speech gender.

Table 1: Experimental variables and variations for all IVAAs used in the study.

Earcon Pitch Contour EP	Earcon Speed ES	Speech Gender	
		Male M	Female F
Flat EP1	Regular ES1	EP1ES1M	EP1ES1F
	Fast ES2	EP1ES2M	EP1ES2F
	Slow ES3	EP1ES3M	EP1ES3F
Ascending EP2	Regular ES1	EP2ES1M	EP2ES1F
	Fast ES2	EP2ES2M	EP2ES2F
	Slow ES3	EP2ES3M	EP2ES3F
Descending EP3	Regular ES1	EP3ES1M	EP3ES1F
	Fast ES2	EP3ES2M	EP3ES2F
	Slow ES3	EP3ES3M	EP3ES3F

In total, 18 Hybrid IVAAs were created and evaluated.

3. EXPERIMENT

3.1. Participants

Nine college-aged participants (*M* = 25.1, *SD* = 2.7; five males and four females) completed the study. Every participant provided informed consent to the study and procedure prior to the start of the study, and all were compensated for their participation at a rate of ten dollars. Every participant had a valid driver’s license and reported normal hearing ability.

3.2. Experimental Design

A within-subjects, or repeated-measures, design was adopted for the study, with each participant listening to all 18 sound alerts. The order of auditory displays shown was randomized using a Latin-Square design.

3.3. Apparatus

Loudspeakers were used to generate each sound alert. The driving simulator used in the current study was a mid-fidelity National Advanced Driving Simulator (NADS) (as seen in Figure 2), which has previously been used in HRGC studies [22].

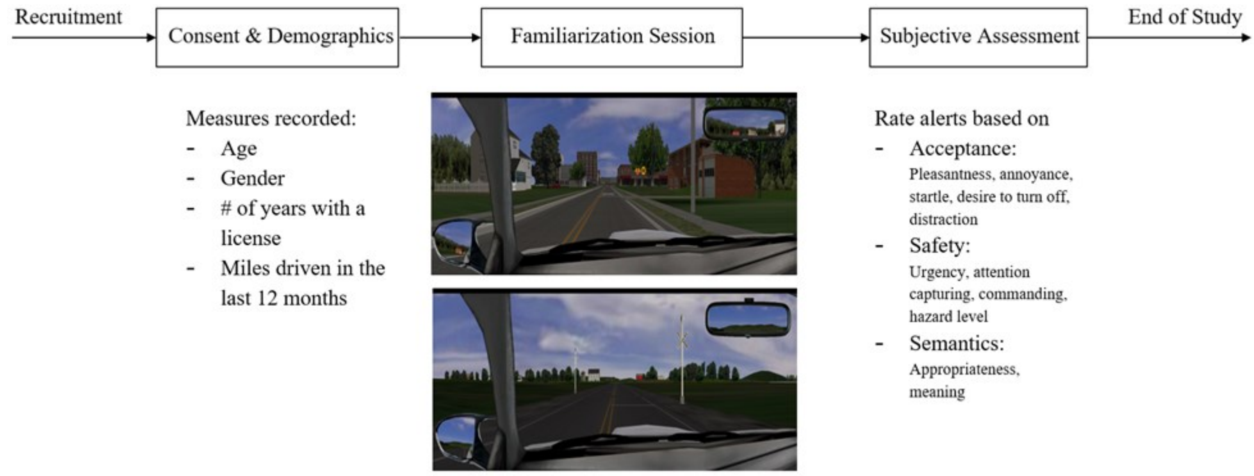

Figure 1: Study procedure, with the driving simulator section used to display urban (top) and rural (bottom) HRGCs.

Figure 2: Driving simulator used for the familiarization.

3.4. Procedure

After first providing informed consent and filling out a demographic questionnaire, participants were instructed to complete a short simulator scenario aimed at familiarizing participants with HRGCs before listening to the IVAAs.

During this session, participants drove through an environment that contained an urban and a rural HRGC, as was done in the prior study (shown in Figure 1). We did not play any IVAAs during this session.

Upon reaching the HRGC, no alert sound was triggered, but study investigators monitored and informed participants of proper safety behavior at the crossings. The session lasted five to seven minutes. Next, investigators explained the meaning of each of the 11 subjective ratings before playing the IVAAs. Dimensions measured participants' opinion on the following IVAA factors: pleasantness, urgency, annoyance, appropriateness, meaning, startle, attention capturing, commanding, desire to turn off, and distraction.

The first ten dimensions were rated on a five-point Likert scale, with a one indicating strong disagreement and five indicating strong agreement with the statement. For the last subjective rating, the perceived hazard level, participants were asked to rate the alert on a four-point scale (Notice, Caution, Warning, Danger), which were identified as an urgency order in Human Factors research [29]. When participants were ready, investigators played each alert at least three times. Participants could request the alert to be repeated and rated each alert along all 11 subjective ratings. The same rating procedure was repeated for all 18 alerts. Participants completed the study upon rating all IVAAs.

4. RESULTS

A repeated-measures ANOVA model was used to analyze subjective rating results. Demographic factors (age, gender, number of years with a US license, miles driven in the last 12 months) were considered in addition to the independent variables that were investigated in the study. For all pairwise comparisons, we used paired samples t-tests with a Bonferroni adjustment to control for Type-I error ($\alpha = 0.05/6 =.0083$). Subjective ratings were grouped in clusters identified in the previous experimental study [23] as follows: acceptance (pleasantness, annoyance, startle, desire to turn off, distraction), safety (urgency, attention capturing, commanding, hazard level), and semantics (appropriateness, meaning).

4.1. Acceptance cluster

For the pleasantness dimension, there was a statistically significant main effect of earcon speed $F(2, 139) = 4.84$, $p = .0093$. As seen in Table 2, pairwise comparisons showed that faster earcons ES2 were rated lower than slower earcons ES3 $t(139) = -3.00$, $p = .0032$. There was no statistically significant difference with regular speed earcons ES1 within the adjusted alpha level.

For the annoyance dimension, no statistically significant effect was found.

For the startle dimension, there was a statistically significant main effect of participant gender $F(1, 4) = 9.29$, $p = .0381$, earcon pitch contour $F(2, 139) = 4.88$, $p = .0090$, and earcon speed $F(2, 139) = 7.83$, $p = .0006$. As seen in Table 2, female participants rated IVAAs less startling than male participants. A flat pitch contour EP1 was rated lower than a descending pitch contour EP3 $t(139) = -2.81$, $p = .0057$. As for earcon speed, slow alerts ES3 were found to be significantly lower in startle ratings than both fast ES2 $t(139) = -3.78$, $p = .0002$ and regular ES1 earcon speeds $t(139) = -2.91$, $p = .0042$.

For the desire to turn off dimension, there was a statistically significant effect of speech gender $F(1, 139) = 5.68$, $p = .0185$ and earcon pitch contour $F(2, 139) = 3.48$, $p = .0336$). As seen in Table 2, female speech in the hybrid IVAA was rated lower than male speech. Pairwise comparisons did not find a statistically significant difference between pitch contours within the adjusted alpha level, but a flat pitch EP1 was rated numerically lower on this scale.

For the distraction dimension, no statistically significant effect was found.

Table 2: Statistically significant differences in subjective ratings for acceptance cluster ratings. Statistically significant pairwise comparisons are also presented.

Rating	Variable	Mean	SD	Comparison
Pleasantness (Earcon Speed)	ES1	3.24	0.87	ES2 < ES3
	ES2	2.93	1.01	
	ES3	3.35	0.8	
Startle (Earcon Pitch Contour, Earcon Speed, Participant Gender)	EP1	2.33	0.91	EP1 < EP3
	EP2	2.78	1.18	
	EP3	2.81	1.12	
	ES1	2.76	1	ES3 < ES1 ES3 < ES2
	ES2	2.91	1.15	
	ES3	2.26	1.01	
	Male	3.01	1.16	F < M
	Female	2.18	0.79	
Desire to turn off (Earcon Pitch Contour, Speech Gender)	EP1	2.61	1.07	______
	EP2	2.93	1.2	
	EP3	2.89	1.16	
	Male	2.94	1.22	F < M
	Female	2.68	1.06	

4.2. Safety cluster

For the urgency dimension, there was a statistically significant main effect of earcon speed $F(2, 139) = 8.19$, $p = .0004$ and speech gender $F(1, 139) = 4.90$, $p = .0285$. As seen in Table 3, slow earcons ES3 were significantly lower than fast earcons ES2 $t(139) = -3.89$, $p = .0002$ and regular speed earcons ES1 $t(139) = -2.89$, $p = .0044$. Female speech was rated as less urgent than male speech.

For the attention capturing dimension, there was a statistically significant effect of earcon speed $F(2, 139) = 7.25$, $p = .001$. As seen in Table 3, slow earcons ES3 were significantly lower on this scale than fast earcons ES2 $t(139) = -3.70$, $p = .0003$. No statistically significant difference was found with regular speed earcons ES1 within the adjusted alpha level.

For the commanding dimension, there was a statistically significant effect of earcon pitch contour $F(2, 139) = 3.10$, $p = .0482$ and earcon speed $F(2, 139) = 6.12$, $p = .0028$. Although pairwise comparisons showed no statistically significant difference between earcon pitch contours at the adjusted alpha level, an ascending pitch EP2 was rated numerically lower than a flat EP1 or descending pitch contour EP3, as seen in Table 3. As for earcon speed, slow earcons ES3 were significantly less commanding than fast earcons ES2 $t(139) = -3.22$, $p = .0016$ and regular speed earcons ES1 $t(139) = -2.79$, $p = .0060$.

For the hazard level dimension, a statistically significant effect was found for earcon speed $F(2, 139) = 7.12$, $p = .0011$. As seen in Table 3, pairwise comparisons show that IVAAs with slow earcons ES3 were rated significantly lower on this scale than fast earcons ES2 $t(139) = -3.51$, $p = .0006$ and regular speed earcons ES1 $t(139) = -2.95$, $p = .0037$.

Table 3: Statistically significant differences in subjective ratings for safety cluster ratings

Rating	Variable	Mean	SD	Comparison
Urgency (Earcon Speed, Speech Gender)	ES1	3.28	1.02	ES3 < ES1 ES3 < ES2
	ES2	3.48	1.22	
	ES3	2.69	0.99	
	Male	3.32	1.07	F < M
	Female	2.98	1.16	
Attention Capturing (Earcon Speed)	ES1	3.94	0.56	ES3 < ES1 ES3 < ES2
	ES2	4.06	0.45	
	ES3	3.67	0.61	
Commanding (Earcon Pitch Contour, Earcon Speed)	EP1	3.17	1.13	———
	EP2	2.74	1.1	
	EP3	2.91	1.03	
	ES1	3.07	1.04	ES3 < ES1 ES3 < ES2
	ES2	3.15	1.05	
	ES3	2.59	1.12	
Hazard Level (Earcon Speed)	ES1	1.94	0.63	ES3 < ES1 ES3 < ES2
	ES2	2	0.58	
	ES3	1.65	1.09	

4.3. Semantics cluster

For the appropriateness dimension, there was a statistically significant main effect of earcon speed $F(2, 139) = 6.19$, $p = .0027$). As seen in Table 4, slow earcons ES3 were rated significantly lower on this scale than fast ES2 $t(139) = -3.17$, $p = .0018$ and regular speed earcons ES1 $t(139) = -2.90$, $p = .0044$.

Table 4: Statistically significant differences in subjective ratings for semantics cluster ratings

Rating	Variable	Mean	SD	Comparison
Appropriateness (Earcon Speed)	ES1	3.81	0.95	ES3 < ES1 ES3 < ES2
	ES2	3.85	0.92	
	ES3	3.43	1.09	

For the meaning dimension, no statistically significant effect was found.

5. DISCUSSION

To design appropriate and effective auditory displays for HRGCs, we previously conducted an experimental study evaluating participants' subjective responses to different auditory alert types. In this follow-up study, we evaluated hybrid sound alerts specifically and investigated the effect of modulating earcon and speech variables for the same hybrid sound alert. The study results indicate consistent effects of earcon variables in a hybrid sound alert. IVAAs with slow earcons were associated with lower ratings with regards to urgency, attention capturing ability, commanding tone, appropriateness, and hazard level perception. These results largely support previous research on auditory display design that indicates the influence of earcon speed and pulse rate on urgency [25, 30, 31]. Additionally, it was found that IVAAs with slow earcons were rated lower in terms of appropriateness by participants. This effect could be explained by a difference between the urgency of the earcon component and the HRGC situation, supporting the importance of perceived affordance in auditory sound design [32]. IVAAs with fast earcons did not differ much in perception than those with regular speeds, being rated numerically lower for pleasantness. While this does not align with previous effects of earcon pulse rate and speed [23], this can be explained by the multimodal nature of the auditory alert. As the fast earcon component is short and perceived as appropriate by users, the hybrid alert's speech component helps mitigate the high perceived urgency effects identified in previous research [23-25]. The findings indicate that the use of medium to high earcon speed should be used for IVAAs at HRGCs.

With regards to earcon pitch contour, a flat pitch range between the two auditory earcon "dings" was associated with a lower perceived startle response, and numerically better ratings for the desire to turn off and commanding dimensions. These results seem to indicate that a deviation in the earcon content may inhibit user response to IVAAs. As with earcon speed, this might be due to a mismatch of affordances. The observed effect can also be attributed to the choice of semitone range used for the earcon, as a larger semitone range was associated with higher urgency and appropriateness [23, 33]. As a result, a flat pitch contour should be used for hybrid IVAAs at HRGCs.

Speech gender influenced user response to the hybrid IVAAs. Male speech induced higher ratings in terms of the startle, desire to turn off, and urgency dimensions. Urgency differences between male and female speech conform to previous studies [23, 25, 34]. These results indicate the potential tradeoffs between the use of male and female speech for IVAAs.

In this follow-up study, no interaction effect between speech gender and participant gender was observed. The previously observed interaction effect between female participants and speech gender was not supported in this follow-up study [23]. This might be partially attributed to the use of TTS speech previously, which has been found to underperform natural speech [35].

The findings suggest that candidate Hybrid IVAAs should possess regular or fast earcon speed and a flat Pitch Contour. The refined hybrid auditory alerts were associated with positive subjective ratings, with scores at levels better than neutral for almost all subjective ratings. Additionally, the urgency and hazard level associated with the hybrid auditory alerts was aligned well with the hazard level associated with HRGC situations, which are supposed to be *caution* signals that should play at all rail crossings regardless of train arrival.

6. LIMITATIONS

This short follow-up study is limited by its small pool of participants due to the COVID 19 pandemic. We were able to hypothesize the effect of speech gender found in the previous study was due to the TTS delivery method [23], other factors such as tone or pitch could have played a role [36]. Additionally, as the study consisted of a series of subjective assessments, the effects identified would need to be investigated in simulator or naturalistic driving conditions.

7. CONCLUSION AND FUTURE WORKS

In this study, we sought to further refine the design of hybrid auditory alerts for HRGC situations. The results of this follow-up subjective assessment validated the previous results on the effect of earcon variables on auditory alerts, as earcon speed and pitch contour influenced subjective ratings of IVAAs. Additionally, semantic variables such as speech gender and the perception of speech alerts were found to help modulate and mitigate the influence of earcon variables. These results further support the recommendation to use hybrid IVAAs at rail crossings. Further work needs to be done to validate these results in a driving simulator environment, which we plan to conduct with refined hybrid auditory alerts that were identified in this study.

8. REFERENCES

[1] X. Yan, L. D. Han, S. Richards, and H. Millegan, "Train–vehicle crash risk comparison between before and after stop signs installed at highway–rail grade crossings," *Traffic injury prevention,* vol. 11, no. 5, pp. 535-542, 2010.

[2] M. G. Lenné, C. M. Rudin-Brown, J. Navarro, J. Edquist, M. Trotter, and N. Tomasevic, "Driver behaviour at rail level crossings: Responses to flashing lights, traffic signals and stop signs in simulated rural driving," *Applied ergonomics,* vol. 42, no. 4, pp. 548-554, 2011.

[3] G. M. Havârneanu, A. Dreßler, J. Grippenkoven, A. Silla, E. Prieto, and M.-H. Bonneau, "SAFER-LC project: Safer Level Crossings by integrating and optimizing road-rail infrastructure management and design," *Proceedings of 7th Transport Research Arena TRA,* pp. 1-6, 2018.

[4] E. Dijkema, "How smart data driven solutions can help to make level crossings safer," presented at the ILCAD Conference, Amersfort, Netherlands, 6 Jun, 2019, 2019.

[5] A. Nikiforiadis, G. Aifadopoulou, J. M. S. Grau, and N. Boufidis, "Determining the optimal locations for bike-sharing stations: methodological approach and application in the city of Thessaloniki, Greece," *Transportation Research Procedia,* vol. 52, pp. 557-564, 2021.

[6] X. Wang, J. Li, C. Zhang, and T. Z. Qiu, "Active warning system for highway-rail grade crossings using connected vehicle technologies," *Journal of advanced transportation,* vol. 2019, 2019.

[7] S. M. Belz, G. S. Robinson, and J. G. Casali, "Auditory icons as impending collision system warning signals in commercial motor vehicles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1998, vol. 42, no. 15: SAGE Publications Sage CA: Los Angeles, CA, pp. 1127-1131.

[8] C. L. Baldwin and J. F. May, "Loudness interacts with semantics in auditory warnings to impact rear-end collisions," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 14, no. 1, pp. 36-42, 2011, doi: 10.1016/j.trf.2010.09.004.

[9] S. Bakhtiari *et al.*, "Effect of Visual and Auditory Alerts on Older Drivers' Glances toward Latent Hazards while Turning Left at Intersections," *Transportation Research Record,* vol. 2673, no. 9, pp. 117-126, 2019, doi: 10.1177/0361198119844244.

[10] F. Bella and M. Silvestri, "Effects of directional auditory and visual warnings at intersections on reaction times and speed reduction times," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 51, pp. 88-102, 2017, doi: 10.1016/j.trf.2017.09.006.

[11] J. Werneke and M. Vollrath, "How to present collision warnings at intersections? - A comparison of different approaches," *Accident Analysis and Prevention,* vol. 52, pp. 91-99, 2013, doi: 10.1016/j.aap.2012.12.001.

[12] X. Yan, Y. Zhang, and L. Ma, "The influence of in-vehicle speech warning timing on drivers' collision avoidance performance at signalized intersections," *Transportation Research Part C: Emerging Technologies,* vol. 51, pp. 231-242, 2015, doi: 10.1016/j.trc.2014.12.003.

[13] C. Gold, M. Körber, D. Lechner, and K. Bengler, "Taking over Control from Highly Automated Vehicles in Complex Traffic Situations," *Human Factors,* vol. 58, no. 4, pp. 642-652, 2016, doi: 10.1177/0018720816634226.

[14] M. Jeon, "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. 2019.

[15] S. Ko, Y. Zhang, and M. Jeon, "Modeling the effects of auditory display takeover requests on drivers' behavior in autonomous vehicles," presented at the Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, Utrecht, Netherlands, 2019.

[16] H. Sanghavi, "Exploring the Influence of anger on takeover performance in semi-automated vehicles," Doctoral dissertation, Virginia Tech, 2020.

[17] R. Gray, "Looming auditory collision warnings for driving," *Human Factors,* vol. 53, no. 1, pp. 63-74, 2011, doi: 10.1177/0018720810397833.

[18] J. Fagerlönn, S. Lindberg, and A. Sirkka, "Graded auditory warnings during in-vehicle use," In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 85-91. 2012., doi: 10.1145/2390256.2390269.

[19] S. Winkler, J. Werneke, and M. Vollrath, "Timing of early warning stages in a multi stage collision warning system: Drivers' evaluation depending on situational influences," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 36, pp. 57-68, 2016, doi: 10.1016/j.trf.2015.11.001.

[20] C. L. Baldwin *et al.*, "Multimodal cueing: The relative benefits of the auditory, visual, and tactile channels in complex environments," *Proceedings of the Human Factors and Ergonomics Society,* pp. 1431-1435, 2012, doi: 10.1177/1071181312561404.

[21] C. L. Baldwin, J. L. Eisert, A. Garcia, B. Lewis, S. M. Pratt, and C. Gonzalez, "Multimodal urgency coding: Auditory, visual, and tactile parameters and their impact on perceived urgency," *Work,* vol. 41, no. SUPPL.1, pp. 3586-3591, 2012, doi: 10.3233/WOR-2012-0669-3586.

[22] S. Landry, M. Jeon, P. Lautala, and D. Nelson, "Design and assessment of in-vehicle auditory alerts for highway-rail grade crossings," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 62, pp. 228-245, 2019, doi: 10.1016/j.trf.2018.12.024.

[23] C. Nadri *et al.*, "Effects of auditory display types and acoustic variables on subjective driver assessment in a rail-crossing context.," *Transportation Research Record,* 2021.

[24] B. N. Walker *et al.*, "Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus," *Human Factors,* vol. 55, no. 1, pp. 157-182, 2013, doi: 10.1177/0018720812450587.

[25] M. A. Nees and B. N. Walker, "Auditory Displays for In-Vehicle Technologies," *Reviews of Human Factors and Ergonomics,* vol. 7, no. 1, pp. 58-99, 2011, doi: 10.1177/1557234X11410396.

[26] E. Šabić, J. Chen, and J. A. MacDonald, "Toward a Better Understanding of In-Vehicle Auditory Warnings and Background Noise," *Human Factors,* 2019, doi: 10.1177/0018720819879311.

[27] K. Kutchek and M. Jeon, "Takeover and Handover Requests using Non-Speech Auditory Displays in Semi-Automated Vehicles," In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1-6. 2019.

[28] S. Moran, E. Skovron, M. Nare, and K.-P. L. Vu, "Accent and Gender Bias in Perceptions of Interactive Voice Systems BT - Engineering Psychology and Cognitive Ergonomics," In International Conference on Engineering Psychology and Cognitive Ergonomics, pp. 457-470. Springer, Cham, 2018.

[29] E. Hellier, J. Edworthy, B. Weedon, K. Walters, and A. Adams, "The perceived urgency of speech warnings: Semantics versus acoustics," *Human Factors,* vol. 44, no. 1, pp. 1-17, 2002.

[30] Dingler, Tilman, Jeffrey Lindsay, and Bruce N. Walker. "Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech." International Community for Auditory Display, 2008.

[31] International Organization for Standardization, "ISO 15006: Road vehicles -- Ergonomic aspects of transport information and control systems -- Specifications for in-vehicle auditory presentation," 2011.

[32] M. Jeon, "Exploring Design Constructs In Sound Design With A Focus On Perceived Affordance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2019, vol. 63, no. 1: SAGE Publications Sage CA: Los Angeles, CA, pp. 1199-1203.

[33] J. Campbell et al., "Human factors design guidance for driver-vehicle interfaces," National Highway Traffic Safety Administration, Washington, DC, DOT HS, vol. 812, p. 360, 2016.

[34] P. Bazilinskyy and J. C. F. de Winter, "Analyzing crowdsourced ratings of speech-based take-over requests for automated driving," *Applied Ergonomics,* vol. 64, pp. 56-64, 2017, doi: 10.1016/j.apergo.2017.05.001.

[35] C. Stevens, N. Lees, J. Vonwiller, and D. Burnham, "On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference," *Computer speech & language,* vol. 19, no. 2, pp. 129-146, 2005.

[36] S. Lattner, M. E. Meyer, and A. D. Friederici, "Voice perception: sex, pitch, and the right hemisphere," *Human brain mapping,* vol. 24, no. 1, pp. 11-20, 2005