

Users may only view, print, copy, download and text- and data-mine the content, for the purposes of academic research. The content may not be (re-)published verbatim in whole or in part or used for commercial purposes. Users must ensure that the author's moral rights as well as any third parties' rights to the content or parts of the content are not compromised.

Subjective Evaluation of Road Situation Alerts Using Visual, Speech and Auditory Modalities with Elderly Drivers

Luka Rukonic¹, Marie-Anne Pungu Mwange²

¹ Institute for Language and Communication, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

² AISIN, Avenue de l'Industrie 19, 1420 Braine-L'Alleud, Belgium
luka.rukonic@uclouvain.be
marie-anne.pungumwange@awtce.be

Abstract. Drivers often rely on navigation systems and traffic alerts to anticipate the road events ahead, such as obstacles, accidents, and roadworks. We designed simple road situation alerts using visual, speech, and auditory modalities to warn the drivers about the upcoming road events. The prototype was tested with a driving simulator and evaluated with elderly drivers. In this study, we evaluated drivers' subjective trust, cognitive workload, and situational awareness in three experimental conditions. We also collected electrocardiograms to measure the workload and stress as a response to the stimuli. Results show that visual warnings were difficult to notice and distractive. Speech and sound combination resulted in the lowest cognitive load, highest trust while maintaining the highest situational awareness. Both speech and visual warnings reduced distrust compared to the baseline. The weather did not affect any of the subjective measures. The physiological analysis showed that visual warnings induce lower stress compared to speech warning alerts. Speech alerts allow highest situational awareness.

Keywords: Human Factors · User Experience · Driving simulation · Road Situation alerts · Elderly drivers

1 Introduction

As automated driving is gaining momentum and the car industry is focusing on the personalization of driver-car interaction [7,12] or the design of takeover requests, many people (45.8%) still prefer to drive manually [14]. In addition, elderly people make a significant part of the drivers on the roads [16]. They need to deal with constant technological advancements, especially now that new cars must be equipped with active safety systems and advanced driving assistance systems (ADAS). Research shows that drivers generally have low awareness of the existence of ADAS in their cars, little knowledge or training about it, and low willingness to use it [1]. Also, drivers often use

navigation systems built in their cars and various mobile applications to navigate to their destination and to receive warnings about obstacles, roadworks, or traffic jams ahead. This type of systems is called road situation notification alert system and it warns the driver about upcoming hazards [2]. We explored the use of auditory and visual modalities to design road situation alerts that inform the driver about the roadworks ahead of the vehicle while driving on the highway. We researched the influence of timely warnings about stationary obstacles ahead would affect drivers' subjective impression of the driving experience and impact their anticipation of the obstacles. Finally, while most studies focus on warning alerts in the context of semi-autonomous vehicles [11,13] and transition from autonomous to manual driving, we focused on the manual driving context. This study has been conducted within the scope of a research project focused on the improvement of ADAS technologies focused on elderly drivers. Our goal is to improve the user experience (UX) and safety of manual and assisted driving.

2 Experimental Design and Procedure

This simulator study investigated the effect of auditory, speech, and visual warnings, on drivers' subjective cognitive workload, situational awareness (SA), and trust while driving manually in three different weather conditions. We also collected participants' electrocardiograms to measure whether the warning alerts induced noticeable physiological responses such as mental stress [3]. The test scenario comprised highway driving and avoiding a stationary truck in the right-most lane, roadworks, or boxes on the road. We instructed participants to drive in the right-most lane as much as possible, as advised by traffic regulations. This way, we ensured that they would always face the obstacle and would need to perform an avoidance maneuver.

This study was conducted as part of a formative UX design approach while working towards the design of warning alerts for the takeover requests in semi-autonomous vehicles. We designed three experimental conditions. First, the baseline (C0) was established in which no warning was given. Then, the two other experimental conditions followed. In condition 1 (C1) the warnings consisted of speech (S) and auditory (A) modalities. In condition 2 (C2), the warnings consisted of visual (V) and auditory (A) modalities. We also manipulated the weather conditions as it is known that adverse weather effects are more stressful for drivers [8]. There were three of them, namely, clear night, rainy night, and foggy night. Therefore, the within-subject independent variable was the condition (C0, C1, C2), while the between-subject independent variable was the weather (foggy, rainy, clear). Trust, cognitive load, and situational awareness were the dependent variables. Auditory warnings were implemented using a beep sound from <http://freesound.org> available under Creative Commons license, free for anyone to use, modify, and distribute. It had a duration of 2.797 seconds and would precede the visual or speech alert. Speech warnings consisted of alerts about approaching roadworks ("Roadworks in 500 meters" and "Roadworks on two right lanes in 200 meters"). The visual warning consisted of an icon depicting the four lanes of the highway; the indication of where the obstacle was positioned and a visual clue of the action to take (Figure 1). Additionally, in all three conditions, we played music from an online radio station to simulate the immersive driving environment. Warnings were triggered 500 meters and 200 meters before reaching the obstacle.

We recruited 9 participants, holding a valid driver's license, between 55 and 68 years old (mean: 61.6; SD=4.7, 4 females, 5 males). They were all in good physical shape and had no impediments. Two participants had a vision corrected with glasses. Before the start, the experimenter connected ECG electrodes, then instructed participants to drive for 15 minutes to become familiar with the simulator. We divided participants into three groups, where each group experienced a different weather condition. Each trial lasted around 3 minutes. Between conditions, we administered NASA-TLX [6], trust [9], and Situational Awareness Rating Technique (SART) [5,15]) post-trial questionnaires. After the trials were completed, we conducted a short interview with the participants. We asked them to describe their opinion and attitude about the road notification alert system and their driving experience with it.

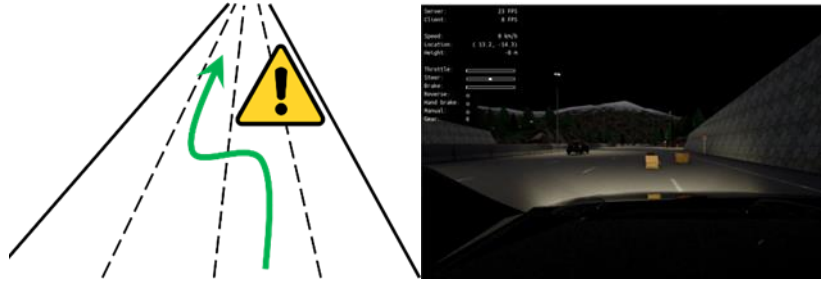


Fig. 1. Visual warning alert depicting the roadworks (left), boxes as an obstacle (right)

2.1 Simulator Platform

The high-fidelity driving simulator consisted of a 180-degree field of view using three screens. The side mirrors and the rear-view mirror were implemented and shown on the main screens. We used CARLA as a driving simulation software platform with which we implemented our scenarios [4]. The Logitech G27 steering wheel was mounted in front of the screens and the pedals were placed underneath the desk. A smaller 10-inch size touch screen was placed on the right side of the driver simulating the infotainment screen and showing visual warnings. To enhance the feeling of reality, wind and engine noise was played from the speakers placed behind the screens.

3 Results

Fig. 1 shows the unweighted workload scores per condition and dimension of the NASA-TLX. The combination of speech and auditory alert (C1) achieved the lowest overall mean TLX score (49.72, SD=15.26). The baseline induced the highest workload of all three conditions (65.65, SD=20.38), followed by C2 (57.3, SD=21.23). Mental demand, physical demand, and effort were the main sources of cognitive workload across the three conditions. Using a repeated-measures ANOVA we found a significant effect of condition on physical demand subscale, $F(2,14) = 9.229$, $p=.003$. However, because the steering wheel required a larger movement amplitude compared to a real car, we treat the scores for physical demand as potentially excessive. There were no significant effects on other subscales ($p>0.05$).

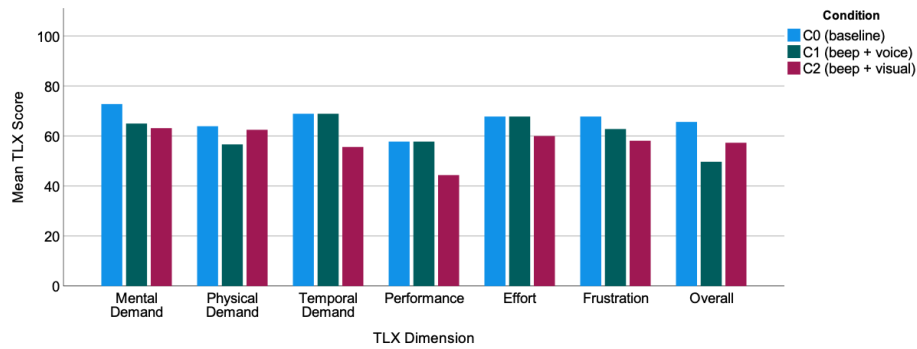


Fig. 1. TLX scores per condition and dimension of NASA-TLX.

Condition C1 also achieved the highest mean trust score amongst participants (4.39), followed by the condition C2 (4.24) and then the baseline (3.99). Looking at Fig. 2, we can see that trust levels remained almost the same across conditions, suggesting that the choice of modalities used to communicate the road notification alerts do not significantly affect the trust. The higher y-value indicates a higher level of trust. The median is highest for C1. The trust scale consists of two subscales, trust and distrust. We found a strong negative correlation between them ($r = -.606$, $p = .001$). Additionally, condition C2 scored lowest on trust ($M=3.1$, $SD=1.596$) and highest on distrust ($M=5.23$, $SD=1.129$). Baseline scored highest on trust ($M=3.84$, $SD=1.008$) and lowest on distrust ($M=4.096$, $SD=1.072$), while C1 scored in-between for trust ($M=3.4$, $SD=1.627$) and distrust ($M=4.93$, $SD=1.017$). A repeated-measures ANOVA showed a significant effect of condition on distrust, $F(2,14)=5.814$, $p=.015$. Specifically, distrust decreased for both C1 and C2 compared to baseline, $p<.05$. However, there was no significant effect of condition on trust.

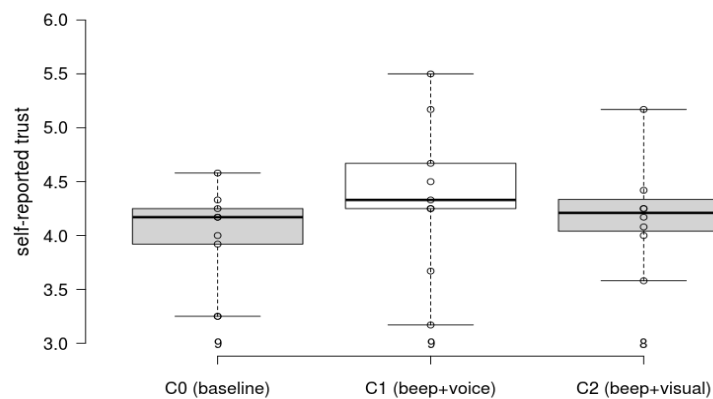


Fig. 2. Boxplots showing the trust scores distribution per condition.

The scores of the SART questionnaire were compared across conditions using repeated-measures ANOVA. There was no significant effect of conditions on overall

SART, $F(2,16) = 2.925$, $p > 0.05$. However, when analyzing each subscale, we found a statistically significant difference for information quality, $F(2,14) = 4.923$, $p = .024$. Post-hoc pairwise comparison using Bonferroni correction revealed that the information quality for C1 increased significantly compared to baseline, $p = .012$. Overall, five participants reported higher SA in C1 compared to C2 and the baseline. Two participants rated their overall SA highest in the baseline condition, and one rated their SA equal in both C1 and C2 (Fig. 3). Table 1 shows the composite SART scores per condition. The highest SART score was 23 (SD=4.3) for the condition C1, followed by a baseline (19.33; SD=3.84), and C2 (7.22; SD=8.2). Across conditions, C1 had the lowest score for attentional demand and the highest score for understanding of the situation. Supply of attentional resources, however, was rated highest in the baseline (C0), although no alerts were presented to the participants. We assume C1 scored well on SA as the participants did not have to look away from the road, remaining focused longer.

Table 1. Mean composite SART scores and SD per experimental condition.

	Overall SART	Demand	Supply	Understanding
C0 (baseline)	19.33 (3.84)	13.89 (3.92)	20.78 (3.56)	12.44 (3.4)
C1 (beep + voice)	23 (4.3)	10.89 (3.06)	18.11 (3.76)	15.78 (2.86)
C2 (beep + visual)	17.22 (8.2)	11.44 (5.61)	16.67 (6.73)	12 (5.39)

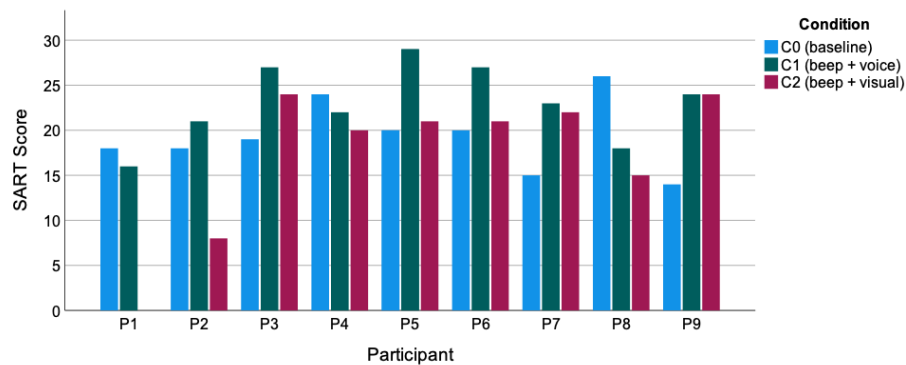


Fig. 3. SART scores per participant.

We split the collected subjective measures data based on whether the participant had a crash during the experiment or not. Whenever a crash happened before the obstacle appeared, we repeated the trial. Because the data was not normally distributed, a Mann-Whitney U test was conducted to check for differences between the two groups. We found that a group that did not crash had significantly lower trust ($U=39.5$, $p < .05$) and cognitive workload ($U=25$, $p < .01$), but there were no significant differences for SA ($U=75.5$, $p = .73$). Finally, a one-way ANOVA was performed to compare the effect of weather on the three subjective measures in foggy, rainy, and clear weather conditions. There were no statistical differences between weather conditions on overall SART, $F(2,23) = 1.786$, $p = .19$. There was also no effect of weather condition on trust, $F(2,23) = 3.205$, $p = .059$, and on overall TLX, $F(2,23) = 0.122$, $p = .88$.

3.1 Qualitative Analysis

Only three participants saw the visual warnings as soon as they were displayed. This suggests that showing the visual warnings in the driver's peripheral vision was not convenient while driving. Participants commented that they did not pay much attention to the infotainment screen while driving. Two participants misinterpreted the meaning of the visual warnings (Fig. 1 left), mistaking them for driving instructions from a navigation system. Participants reported that the voice messages were effective at conveying information about obstacles and were easy to process while driving.

Qualitative data collected from the interviews generally shows that drivers are in favor of such warnings alerts and they would like to use precise road situation notifications while driving. Their main concerns were about the accuracy of the received data. When asked whether they would prefer the system to give them information about obstacles or instructions on how to avoid the obstacles, five participants were in favor of information, and four in favor of receiving instructions. Additionally, participants reported that the beep sound draws their attention and enables them to anticipate that they should be more attentive to the road, but also prepares them to hear the voice warning.

3.2 Physiological Analysis

We used the Biopac MP36 to record the electrocardiogram (ECG) data. For this, we attached electrodes on both the participants' wrists and the third one on their right ankle. This setup can introduce noise in the collected data, given that they had to drive and move their hands. It was mostly for modesty and convenience reasons that we opted for the placement on less intrusive body locations. We were not able to collect good physiological data for all participants. Sometimes, electrodes would detach during the test. After sorting out all the unusable data, we performed the analysis on 8 participants. In general, we observed a decrease in the participants' mean heart rate of each subsequent trial that followed the order of administration of trials. We could attribute this effect to the experimental design and participants becoming more familiar with the simulator.

For each participant, we calculated their average heart rate (HR) value in a 15s window directly before the obstacle and the average value in a 15s window starting 10s after facing the obstacle. The delay was introduced because the change in heart rate is not instantaneous. The heart rate reactivity is the difference between the two HR values (HR after – HR before obstacle). A Pearson's correlation was computed to assess the relationship between the mean HR and overall TLX score, trust, and SA. There was a moderate negative correlation between HR and TLX ($r = -.32$), HR and trust ($r = -.25$), but no correlation between HR and SA ($r = .01$). However, none of the three were significant at $p < 0.05$ level.

Physiological responses being very different for each person, they are very difficult to compare. However, Jouven et al. [10] have proven that the average mild stress level is associated with the reactivity of 8.9 BPM. Therefore, we can see in Fig. 4 that no matter what modality was presented to participants, there was no significant increase in heart rate that could lead to mental stress. Additionally, the comparison of reactivity also shows that the heart rate when encountering an obstacle seems to decrease when participants were warned with the auditory and visual modality (C3). Overall, the mean

heart rate was highest for the baseline (83.05, SD=9.68), followed by C1 (75.27, SD=5.32) and C2 (73.6, SD=7.62).

This could imply that the choice of modalities is associated with driver's stress which could be used when designing driver-car interactions. However, future research is needed to validate these basic findings. We suggest investigating the use of speech and visual modalities to warn drivers depending on the distance to the road situation and obstacles.

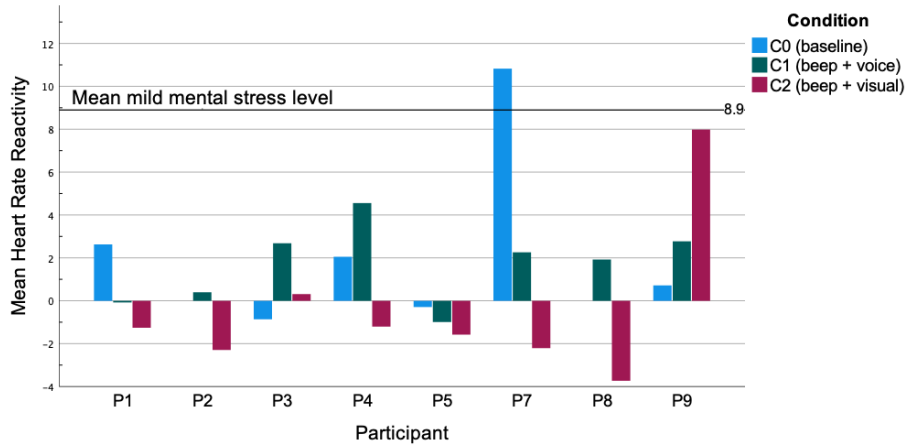


Fig. 4. Heart rate reactivity at obstacle per participant.

4 Conclusion

The combination of speech and auditory modalities used to convey traffic alerts was a preferred solution by the participants. Also, both visual and speech alerts decreased the amount of distrust in the road situation alert system. Speech alerts enabled participants to achieve higher situational awareness, as they could maintain better attention to the road. The results suggest that the weather conditions did not affect drivers' cognitive load, trust, and situational awareness. We have not observed a high increase of mental stress on drivers when using speech or visual modalities as road situation alerts. In fact, visual warnings induced less heart rate reactivity than speech alerts. Regarding the simulator platform, we recognized the importance of using a realistic and immersive driving simulator. A limitation of our study lies in the fact that the simulator was quite different from a real car, thus imposing an unnecessary effort on the participants. However, they reported that after a familiarization drive, they were able to grasp the driving feeling and felt more comfortable with it. Furthermore, a larger sample study should be conducted to make further generalizations for the use of warning modalities, but we believe our results provide useful indications towards design guidelines for driver-car interaction. This study contributed to a better understanding of the output modalities' effect on driver's experience.

Acknowledgments. The authors acknowledge the support by the project VIADUCT under the reference 7982 funded by Service Public de Wallonie (SPW), Belgium.

References

1. Bronson, K., Page, S. L., Robinson, K. M., Moon, A., Rismani, S., & Millar, J. (2019). Drivers' Awareness, Knowledge, and Use of Autonomous Driving Assistance Systems (ADAS) and Vehicle Automation. arXiv preprint arXiv:1911.10920.
2. Chung, W. Y., Chong, T. W., & Lee, B. G. (2019). Methods to detect and reduce driver stress: a review. *International journal of automotive technology*, 20(5), 1051-1063.
3. Costin, R., Rotariu, C., & Pasarica, A. (2012, October). Mental stress detection using heart rate variability and morphologic variability of EEG signals. In 2012 International Conference and Exposition on Electrical and Power Engineering (pp. 591-596). IEEE.
4. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. *Proceedings of the 1st Annual Conference on Robot Learning*, in PMLR 78:1-16
5. Gawron, Valerie J. (2019). Measures of Situational Awareness in Human Performance and Situation Awareness Measures. Ed. Gawron Valerie. CRC Press. Routledge Books Online.
6. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.
7. M. Hasenjäger, M. Heckmann and H. Wersing. (2020). A Survey of Personalization for Advanced Driver Assistance Systems. In *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 335-344, June. doi: 10.1109/TIV.2019.2955910.
8. Hill, J. D., & Boyle, L. N. (2007). Driver stress as influenced by driving maneuvers and roadway conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(3), 177-186.
9. Jian, J.-Y. (1998). Foundations for Empirically Determined Scale of Trust in Automated Systems.
10. Jouven, X., Schwartz, P. J., Escolano, S., Straczek, C., Tafflet, M., Desnos, M., ... & Ducimetière, P. (2009). Excessive heart rate increase during mild mental stress in preparation for exercise predicts sudden death in the general population. *European heart journal*, 30(14), 1703-1710.
11. Kasuga, N., Tanaka, A., Miyaoka, K., & Ishikawa, T. (2020). Design of an HMI System Promoting Smooth and Safe Transition to Manual from Level 3 Automated Driving. *International Journal of Intelligent Transportation Systems Research*, 18(1), 1-12. <https://doi.org/10.1007/s13177-018-0166-6>
12. Lilis, Y., Zidianakis, E., Partarakis, N., Antona, M., & Stephanidis, C. (2017). Personalizing HMI elements in ADAS using ontology meta-models and rule based reasoning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10277 LNCS, 383-401. https://doi.org/10.1007/978-3-319-58706-6_31
13. Politis, I., Langdon, P., Adebayo, D., Bradley, M., Clarkson, P. J., Skrypchuk, L., Mouzakitis, A., Eriksson, A., Brown, J. W. H., Revell, K., Stanton, N. (2018). An evaluation of inclusive dialogue-based interfaces for the takeover of control in autonomous cars. 23rd International Conference on Intelligent User Interfaces, Proceedings IUI'18, (March), 601-606. <https://doi.org/10.1145/3172944.3172990>
14. Schoettle, B., and Sivak, M. (2016). Motorists' Preferences for Different Levels of Vehicle Automation: 2016. <http://www.umich.edu/~umtriswt/PDF/SWT-2016-8.pdf>. Accessed July 22, 2015.
15. Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. AGARD. *Situational Awareness in Aerospace Operations*, 17, 23-53.
16. U.S. Census Bureau. (2018). An aging nation: Projected number of children and older adults. Retrieved from www.census.gov/library/visualizations/2018/comm/historic-first.html.