Advances in Transportation, Vol. 60, 2022, 25–32
https://doi.org/10.54941/ahfe1002430

The Effects of Multi-Modal Takeover Request on Distracted Drivers' Takeover Performance and Perception

Ye Yang, Mengying Deng, Yaohua Zhou, Mingjiang Shi, and Ruixuan Gao

Banma Network Technology Co., Ltd., Shanghai, China

ABSTRACT

Visual and cognitive distractions caused by non-driving related tasks (NDRT) engagement and over-trust were exposed and considered a serious driving safety hazard for the automated driving system (ADS). However, the typical HMI design of takeover request (TOR) in the market (displaying a symbol in the digital cluster with an auditory warning) can be easily unnoticed, especially in auditory conflicts scenarios, and should be optimized urgently. This study introduced four multi-modality combinations of TOR. It examined the differences in drivers' reaction time, reaction quality, and subjective perception in situations where the drivers were experiencing visual and cognitive distractions. Eighteen drivers participated in the driving-simulator experiment. Results showed that the symbol and speech combination was difficult to notice, yet it was the easiest to understand. Vibration and speech combination could effectively alert the distracted drivers within the shortest time, but there was negative experience feedback from the drivers. Ambient light and speech combination could guide drivers but was inconvenient in time-critical situations. Vibration, light, and speech combination has shown a better performance and user experience in general. Our results provide multi-modal TOR design implications that improve takeover performances and user experiences.

Keywords: Automated driving, Takeover request, Multi-modality, Non-driving related Task, Takeover performance, User experience

INTRODUCTION

In conditional automation, the automated drive system administrates the object and event detection and response task (OEDR). However, when the vehicle encounters a situation that is not within the operational design domain (ODD), the driver must take over the OEDR tasks when they receive a TOR. Reports show that drivers may undertake NDRT as ADS conditionally liberates drivers from OEDR tasks (Du et al., 2020) and decrease their attention on environmental observation and system operation, such as perception and reaction of the TOR (Sun and Li, 2019). Therefore, it is necessary to monitor the distracted state of drivers in automated driving and evaluate the most effective type of TOR alert. Multi-modal warning messages have been demonstrated to have benefits, that at least one modality is reachable to

© 2022. Published by AHFE Open Access. All rights reserved.

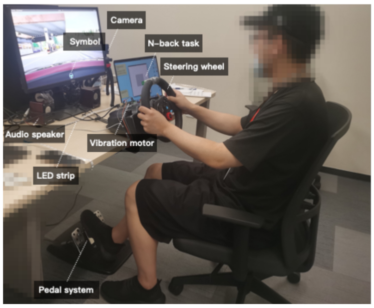

Figure 1: Simulator set-up.

the driver and have been recommended by NHTSA in DVI Design Guidance (NHTSA, 2021). In this study, we incorporated the concept of multimodality and compared 4 TOR combinations with a visual and cognitive distracted driver, and examined the differences in drivers' takeover performance and subjective perception. Our goal is to help improve takeover performance and user experience. With auditory as a fixed modality in the combination, the collocation modalities we compared are focal vision, ambient vision, haptic, and ambient visual with haptic. Focal visual and auditory modalities are used in the traditional HMI design of TOR; haptic modality is adapted by some HMI design, primarily used in safety-related alerts, such as lane departure warning; ambient visual shows potential to draw attention – unlike focal visuals, which only covers 3-5 degrees of the visual field (Vision and NRC, 1985) visual stimuli within the entire visual field can be perceived with ambient vision.

METHOD

Experiment Set-Up

A static driving simulator was set up in a dedicated lab consisting of three parts: driving simulator, modular hardware, and representation of visual and cognitive distraction NDRT (N-back memory task (Jaeggi et al., 2008)). The virtual driving scenarios were achieved by using CARLA, a driving simulation software platform, and the automated-driving algorithm was from AutoPilot by CARLA. For modular hardware, a Logitech G29 steering wheel and pedal system were mounted. Arduino UNO with LED strip, vibration motor, and the audio speaker was also applied to provide modular interaction (Fig. 1). The N-back memory task requires the participants to play a memory game (programmed by Pygame) during the experiment. When participants start the game, a number ranging from 1 to 9 will appear randomly in a Sudoku. Then, they were required to press a button if the current number was the same as the number that is 1 step back in the process.

Experimental Design

The four types of stimuli were selected to correspond with the sensory transmission channels (focal vision, ambient vision, haptic and auditory): symbol,

Table 1. Descriptions of TOR groups.

Group	TOR Combinations	Descriptions
C0	Symbol + Speech*	Red icon with warning text remains still on display until TOR is completed.
E1	Light + Speech*	LED light strip flashes in orange 3 times.
E2	Vibration + Speech*	The vibration motor vibrates 3 times.
E3	Vibration + Light + Speech*	LED light strip flashes 3 times, and the vibration motor vibrates 3 times at the same frequency.

*Speech contains the message "Autopilot assistance is about to exit, please take over driving." (In Chinese) and is broadcasted once in every TOR combination.

light, vibration, and speech. The auditory stimulus was set as a constant, and all TOR combinations included auditory as one of the modalities. In the experiment, we designed four experimental conditions as independent variables. The controlled group (C0) was established with symbol and speech, as this combination is common in the current HMI design. Experimental group 1(E1) consisted of light and speech, group 2 (E2) consisted of vibration and speech, while the last group (E3) consisted of light, vibration, and speech. The virtual driving scenario is controlled throughout the experiment. In all conditions, Town04 Map from CARLA was adapted, which consists of an infinite loop with a highway and a small town. In order to ensure the accuracy of the experiment, the traffic density was also a controlled variable in which we included 0 oncoming vehicles in the driving scenario.

Eighteen company employees between 18 and 50 years old (3 females, 15 males) were recruited to participate in the simulator experiment. All participants had a valid driver's license, were physically healthy, and all had a normal or correct-to-normal vision. The average driving experience of the participants was 8.53 years.

Dependent Variables

The takeover performance was evaluated by reaction time and quality. Reaction time refers to the time that the drivers took to start performing the takeover action after each TOR; the action could be pressing on the brake or turning the steering wheel. Reaction quality is measured by drivers' maximum steering speed (Degree/s), brake-pressing speed (0-100%/s), and brake-pressing depth (0-100%). The subjective experience about the TOR combinations was measured by the "In-Car Warning Perception Scale (ICWPS)", which was designed based on a 10-point semantic differential scale and divided into 8 dimensions (from 0 to 10 points): "Complicated - Simple", "Uninteresting - Interesting", "Not urgent - Urgent", "Conservative - Creative", "Obstructive - Supportive for driving", "Confusing - Understandable", "Difficult - Easy to perceive", and "Inefficient - Efficient". The ICWPS has distributed a total of 72 copies, with the Cronbach's α coefficient being 0.744, the KMO value being 0.736, and Bartlett's sphere test being $\chi 2=428.688$, $p < 0.01$, which is indicated to be suitable for analysis.

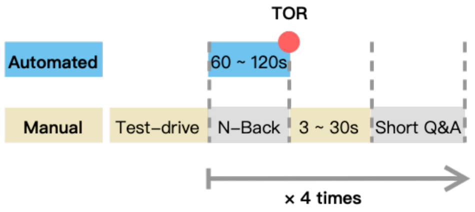

Figure 2: Sequence of events in the experiment procedure.

Procedure

Before starting the experiments, the participants were asked to sign the consent contract and began with a familiarization drive to accustom themselves to the simulator. When the ADS mode was turned on, the drivers were asked to complete the N-back 1 game on the laptop. The vehicle would drive for 60 -120 seconds in the automated state. The driving simulator would randomly issue one of four TOR combinations, which was designed to be unexpected for the drivers. After the execution of TOR, the drivers were expected to act to end the automated mode and transition the vehicle into manual operation. Once drivers successfully took over the driving, they would drive a few seconds and brake to stop. The drivers would then be given the ICWPS to fill out; this step also indicates that the first run of the experiment is completed. The procedure above was required to perform 3 more times with different TOR combinations (Fig. 2).

RESULT

Performance

The descriptive statistical result showed that the combination of vibration and speech (E2) achieved the lowest average reaction time of 3.05 seconds ($SD = 1.26$), followed by the combination of light, vibration, and speech (E3), which carried an average reaction time of 3.36 seconds ($SD = 1.47$). The combination of symbol and speech (C0) induced the highest average reaction time of 4.48 seconds ($SD = 2.29$). The one-way ANOVA analysis indicated a significant effect of TOR on drivers' reaction time, $F\,(3, 68) = 3.38$, $p = 0.02$. However, the result of a Bonferroni-adjusted post-hoc analysis revealed no significant difference between each combination of TOR ($p > 0.05$). For reaction quality, the one-way ANOVA result showed that all the tested factors (max. brake speed, max brake depth, and steering speed) were not affected by the TOR ($p > 0.05$); the post-hoc also revealed that there is no difference between each combination of TOR ($p > 0.05$) (Fig. 3).

Subjective Perception

The ANOVA results indicated that there is an effect of TOR on some measuring factors of subjective perception, in which "attractiveness" ($F = 6.2$, $p < 0.001$), "urgency" ($F = 7.2$, $p < 0.001$), and "creativity" ($F = 12.2$, $p < 0.001$) were affected the most, followed by "perceptiveness" ($F = 4.2$, $p = 0.009$). Among the 8 measuring factors, "simplicity", "supportability",

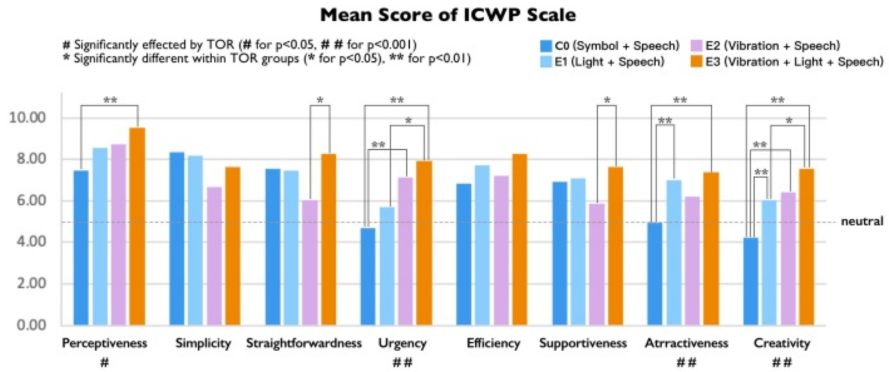

Figure 3: Mean Score of ICWPS.

"straightforwardness", and "efficiency" have shown to be not affected by the TOR ($p > 0.05$). The post-hoc analysis, as shown in the Figure below, reveals that discrepancy occurs between C0 group (symbol + speech) and E3 group (light + vibration + speech) the most. The discrepancy between groups mostly happens in "attractiveness", "urgency", "creativity", and "perceptiveness". In "creativity", C0 group is shown to have a significant difference from all three experimental groups; the difference between C0 and E3 was the largest ($p < 0.01$), followed by E2 group (vibration + speech, $p = 0.01$).

DISCUSSION

The four examined TOR combinations had notable effects on the reaction time of the drivers, with low cognitive load and visual distraction. However, TOR had no direct effect on the reaction quality. The speed or depth drivers pressed on the brake or rotated the steering wheel could depend on personal factors, such as driving habits, physical strength, etc.

Symbol + Speech (C0) The controlled group induced the highest average reaction time; the average score in the "perceptiveness" section of the ICWPS was also the lowest, which shows that the participants felt that this TOR combination was the most difficult to perceive in comparison with the other three combinations. In the interview after the experiment, 4 participants reported that they did not see the symbol that appeared on display. Some mentioned that they "felt a change" in the display, but they "did not notice the specific information." This is because focal vision is required to process and recognize symbols and objects (Vision and NRC, 1985), but when drivers are focused on a secondary task, their focal vision is no longer staying on the road or the front display. However, C0 group has the highest mean for "simplicity", and the instructional warning text makes it easy to understand (second-highest average score for "straightforwardness"). This showcases that although visual alerts cannot conveniently draw distracted drivers' attention, it still executes instruction and explanation to inform the drivers of what they must do.

Light + Speech (E1) In average reaction speed, the results are similar between E1 and C0 (symbol and speech). The ICWPS also reflected a neutral score (mean = 5.71) of this combination in "urgency". After the experiment, the participants reported that the ambient light was similar to a guiding message and was easier to perceive than symbols with text because their sight was not on the road. In terms of visual perception, light can be perceived in a wider range, "No matter where I looked, I could still feel it," one participant stated. When the driver is distracted by a secondary task, lights can play a role in "awakening" the driver, providing guidance, and relieving the unnecessary tension caused by sudden warnings. Therefore, light and speech TOR combination may be more suitable for warnings that are not time-critical or act as auxiliary functions.

Vibration + Speech (E2) Vibration with speech was the TOR combination that obtained the lowest reaction time. The descriptive statistic of the ICWPS results shows that, in the subjective rating of participants, vibration and speech was considered to be more urgent and was highly perceivable (mean score in "urgency" = 7.14; mean score in "perceptiveness" = 8.73). On the other hand, the intensity of the vibration used in the experiment frightened some drivers, especially when they were immersed in the secondary task with cognitive load. Vibration and speech TOR also induced the lowest average score in "supportiveness" (mean = 5.87) and "straightforwardness" (mean = 6.03) among the four TOR combinations, which shows that it may negatively impact driving. Vibration would be convenient for alerting the drivers in the shortest time. Still, this TOR type should be utilized with caution, or it could be incorporated with another form of sensory stimulus that could alleviate drivers' stress and effectively alert them at the same time.

Light + Vibration + Speech (E3) The ICWPS reflected that E3 group, which consists of light, vibration and speech, demonstrated a comprehensively satisfactory result among the four combinations tested. The ambient light used in this TOR combination has been demonstrated to have a stress-relieving effect – light is able to improve the communications between the regions of the brain that are responsible for handling emotions and stress (Minguillon et al., 2017). In the interview after the experiments, participants reflected that the ambient light used in the experiment made them feel calmer compared to the TOR combination of vibration and speech. The ICWPS scores also have shown that vibration, light, and speech obtained the highest average score in "supportiveness", "urgency", and "perceptiveness".

Design Implication

The results of our study provide design implications for multi-modal TOR alerts. Firstly, the design of the TOR should especially consider the advantages of reaching users with different modalities in different NDRT scenarios. For visual distraction scenarios, focal visuals should be the most fundamental element to help drivers understand the current situation and the corresponding action they need to take. Therefore, we need to introduce other modalities (e.g., ambient light) to help drivers' vision return straight-ahead. Secondly, considering basic safety requirements and the enhancement of experience, multi-modal TOR may have negative implications, such as the individual differences in light perception and startle perception of vibration intensity mentioned above, and need to be avoided after more studies are being done. Eventually, the observed discrepancies from multiple aspects among E1, E2, and E3 show that the utility value of E3 has no accumulation pattern with E1 and E2. Therefore, all combinations of TOR design should be tested separately for user experience and driving behavior.

CONCLUSION

Generally, the light, vibration, and speech (E3) TOR combination were most preferred by the participants under visual and cognitive distraction. Vibration and speech was the TOR combination that induced the shortest reaction time. Symbol and speech combination had shown to be the least convenient TOR when drivers are under distraction, as it had been reflected to be relatively easier unnoticed.

Further study should be done with different cognitive loads to observe the effect. Moreover, the takeover performance needs to be tested in different dimensions, such as drivers' trust, eye movements, or heart rates, to indicate the cognitive and emotional states of the drivers. Lastly, a larger sample study should be conducted with real cars and real automated driving scenarios to make additional generalizations for the application of TOR warning design. Nevertheless, our findings provide useful insights and indications towards the concept of multi-modality TOR design.

REFERENCES

Borojeni S S, Chuang L, Heuten W, et al. (2016). "Assisting drivers with ambient takeover requests in highly automated driving". Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications, 237–244.

Dogan, Ebru, et al. (2021). "Manual takeover after highly automated driving: Influence of budget time and Lane Change Assist on takeover performance." European Conference on Cognitive Ergonomics 2021, 33, 1–6.

Du, Na, et al. (2020). "Evaluating effects of cognitive load, takeover request lead time, and traffic density on drivers' takeover performance in conditionally automated driving." 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. (Online Virtual Conference), 66–73.

Jaeggi, Susanne M., et al. (2008). "Improving fluid intelligence with training on working memory." In Proceedings of the National Academy of Sciences, 105.19, 6829–6833.

Minguillon, J. et al. (2017) 'Blue lighting accelerates post-stress relaxation: Results of a preliminary study', PLoS ONE, 12(10), p. e0186399. doi: 10.1371/journal.pone.0186399.

National Highway Traffic Safety Administration . (2021). Distracted Driving 2019, TRAFFIC SAFETY FACTS

SAE International Surface Vehicle Recommended Practice, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE Standard J3016_202104, Rev. Apr. 2021.

Sanghavi, Harsh, Yiqi Zhang, and Myounghoon Jeon. (2020). “Effects of anger and display urgency on takeover performance in semi-automated vehicles.” 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, (Online Virtual Conference), 48–56.

Sun, F. and Li, K., (2019). Dian dong qi che gong cheng shou ce =. Beijing: Ji xie gong ye chu ban she.

Vision, N.R.C. (US) C. on (1985). “TWO MODES OF VISUAL PROCESSING.” Emergent Techniques for Assessment of Visual Performance. National Academies Press (US).

Wickens, Christopher D.,(2008). “Multiple resources and mental workload.” Human factors, 50,3, 449–455.