Ergonomics in Design, Vol. 47, 2022, 151–158
https://doi.org/10.54941/ahfe1001932

Employing Synesthesia-Based Warnings to Enhance Road Safety During an Automated Driving Scenario

Liang Zhuo[1] and Danhua Zhao[1,2]

[1]School of Design, Hunan University, Lushan Road (S), Yuelu District, Changsha, Hunan Province, China
[2]State Key Laboratory of Advanced Design and Manufacture for Vehicle Body, Lushan Road (S), Yuelu District, Changsha, Hunan Province, China

ABSTRACT

With the rapid employment of advanced driver assistance systems (ADAS), drivers will perform more varied and demanding tasks, which may result in reduced perception and response to emergencies. Existing studies have proved the effectiveness of multimodal warning signals and found that they can reduce drivers' response time. However, most of these studies only focus on the superposition effect between different channels. There is a lack of research on the correlation between different channels and their impact on drivers' workload level. Therefore, a multimodal warning signal model based on the theory of Audio-visual synesthesia is proposed in this paper, and the effects of this model on road safety in the scenario of automatic driving are explored through a controlled experiment of synesthesia and non-synesthesia signals. The results show that compared with the non-synesthetic signal, the multimodal warnings based on synesthesia can transmit the relevant information more quickly, and also reduce the workload level of the driver in the assisted driving scenario. The findings will aid the design of an early warning system for future autonomous vehicles.

Keywords: Autonomous vehicles, ADAS, Audio-visual synaesthesia, Multimodal interaction, Workload

INTRODUCTION

With the expansion of the autonomous vehicle (AVs) market and the continuous development of intelligent technologies, advanced driver assistance systems (ADAS) are increasingly being applied to people's daily driving processes.The development of AVs is currently at the L2-L3 level (Shadrin et al. 2019). Under this level, ADAS and the driver are jointly involved in the control tasks of the car, and the way ADAS interacts with the user has a huge impact on driving behavior and driving safety. Warning is one of the main problems related to ADAS. When a road emergency occurs, ADAS needs to quickly issue a warning to convey vehicle driving and environmental information to help users judge the driving situation. Successful design of ADAS warnings can effectively reduce road accidents (Kuehn et al. 2009). Therefore, how to improve the validity of ADAS and build a more efficient way of human-machine interaction is the problem that researchers are facing.

MULTIMODAL INTERACTION

Multimodal interaction refers to a way of human-computer interaction through multiple senses such as vision, hearing, touch, smell, and movement. Multimodal interaction is considered to be a more natural way of human-computer interaction because it conforms to the human cognitive model. Compared with single-channel interaction (Turk, 2014), it can realize human-machine dialogue more accurately and effectively, and improve the user's interactive experience.

With the development of in-vehicle intelligent technology, emerging single-channel perception technologies, such as speech recognition, artificial intelligence, gesture recognition, AR, VR, and other technologies, are becoming more mature, and multi-modal interaction forms also have greater potential. The multimodal interaction of automobiles is becoming a topic that more and more researchers and automobile manufacturers are conducting in-depth research and development. Almén et al. studied the effect of sound and tactile input on the driver's attention in the driving subtask (Almén, 2002), and the results showed that the combination of the two-channel inputs is better than the single-channel input ,and can shift the driver's attention back to the main driving mission faster. Robert et al. proposed an in-vehicle interaction model that allows control of the car (eg, window switches) through a combination of voice, gaze, and gestures (Robert, 2016).

SYNESTHESIA THEORY

Synesthesia is a feature mapping process between different senses in psychology (Sun et al. 2021), which refers to the resonance of two or more perceptual sensations caused by a single channel of sensory stimuli. As a system, human sensory organs are interconnected in a certain sense, and the senses such as vision, auditory, touch, taste and smell can be interconnected and functioned.

At present, many researchers have studied the phenomenon of synesthesia. Marks studied the interaction between vision and hearing, and confirmed that higher pitches and louder sounds are associated with lighter colors (Marks, 1974). Evans et al. demonstrated a certain cross-talk between auditory frequencies and visual location, size, and spatial frequencies (Evans et al. 2010). The use of synesthesia effect to intervene in the multi-modal autonomous vehicle interaction mode and improve the synergy of each channel can further improve the reliability and effectiveness of ADAS and enhance users' driving experience.

HYPOTHESIS

At present, autonomous vehicles are still mainly at the L2-L3 level. Under this level of automated driving, the driver still needs to take over control of the vehicle when special circumstances occur to ensure the safety of driving. The occurrence of the user's takeover behavior largely depends on the user's judgment of the road conditions and environment. Therefore, when the situation occurs, the speed, accuracy and effectiveness of the assisted driving system to transmit relevant road condition information will greatly affect the user's final takeover situation.

Current research has shown that multimodal early warning methods can effectively improve the user's response speed in emergency situations. The research on multimodal early warning mainly focuses on the combination or superposition of different sensory channels. There is a lack of research on the correlation between different channels and their impact on drivers' workload level. As a phenomenon of interaction between different sensory channels, synesthesia can help produce better coordination effects between the various channels of multimodal interaction.

Relevant studies have shown that vision and hearing are the most important cognitive senses, and 97% of the information people receive daily comes from vision and hearing. Therefore, we propose that a multimodal warning system based on audio-visual synesthesia can effectively improve the driver's information receiving efficiency in emergency situations and reduce the driver's task load.

EXPERIMENTS

PARTICIPANTS

We recruited 30 volunteers to participate in the experiment, including 14 women and 16 men. All participants had a valid driver's license and at least one year of driving experience, with normal vision, normal hearing, and no language impairment. In order to prevent the accumulation of simulated driving experience from affecting the results of subsequent experiments, participants were randomly divided into two groups: Group A (non-synesthesia) and Group B (synesthesia) to make experimental tests under different modes. The experiment lasted about 20 minutes for each participant.

MATERIALS

The experiment uses a highly realistic driving simulator containing a seat, a steering wheel, brakes, display devices, and a camera. We recorded the entire driving process including relevant road conditions in advance through the simulation driving software citycardriving, and used Premiere to edit and combine relevant sounds and images. In the experiment, the participants sat in the driver's seat and watched the driving picture on the screen. When encountering an emergency and the assisted driving system issued a warning, they could intervene driving process by holding the steering wheel or stepping on the brake pedal according to their own judgment of the situation. During the experiment, we would ask participants to complete a simple secondary task (sending a message or answering a phone call), and participants could use their own mobile phones to complete the task to avoid differences in user familiarity with mobile phones.

DESIGN

The experiments were carried out by comparing group A (non-synesthesia) with group B (synesthesia). Participants in both groups were tested in an

Figure 1: Different forms of synaesthesia combinations.

autonomous driving simulation scenario at L2 to L3 levels. In the driving scenario, two types of emergencies are set up – common low load scenario LLS (steering reminder, speed limit reminder) and emergency high load scenario HLS (pedestrian insertion, vehicle in front of emergency brake). In the two scenarios, there are differences in the prompt sounds issued by ADAS: In the case of LLS, low-frequency beep prompt sounds are used, while in the case of HLS, high-frequency beep prompt sounds are used.

In the two groups of experiments, the warning information sent by ADAS was transmitted synchronously through the visual channel and auditory channel, which was composed of prompt sound, atmosphere light band and HUD display. In Group A, the visual channel of warning message did not produce synesthesia with the auditory channel, the standard white light source of 6000K was used for both the light band and the HUD warning message, and the area of HUD warning in the screen accounted for 3% (a). In Group B, there was a synesthesia effect between the audiovisual channels of the warning message. According to the research of Sun X et al, low-frequency sounds have a high correlation with blue, and high-frequency sounds have a high correlation with red. Meanwhile, there is an association effect between low-frequency sound and visuals with a uniform area ratio, and an association effect between high-frequency sound and visuals with a large difference in area ratio and strong contrast (Sun et al. 2018). Therefore, in case of LLS, we adopted the synaesthesia combination (b) "low-frequency prompt tone – blue light band – 3% area ratio HUD warnings". In case of HLS, we adopted the synaesthesia combination (c) "high-frequency prompt tone – red light band – 6% area ratio HUD warnings".

Participants in both groups were asked to complete a secondary task (such as watching a video on a mobile phone) as prompted by the experimenter, to reflect the efficiency of receiving the warning message in the distracted state (the audio-visual channel was occupied). The whole experiment will be recorded, and the time when the participants apply the brakes will also be recorded by the equipment.

PROCEDURE

Upon arrival, participants filled in their basic personal information and read the experiment instructions. Participants first undergo a 5-minute L3 level automatic driving simulation experience to familiarize themselves with the experimental environment and driving operation, so as to put in the real driving experience. During the experiment, participants pass through LLS and HLS successively for an 8-minute simulated driving experience. In the case of LLS, the AVs will experience two scenarios:"turn right at traffic lights" and "keep spacing between vehicles". While in the case of HLS, the AVs will experience two scenarios: "pedestrian insertion" and "vehicle in front of emergency brake". In all scenarios, participants are given the option to hit the brakes and take over, or to continue watching the video without doing anything. After the experiment, participants are required to complete a Nasa-tlx scale and do an interview.

DATA COLLECTION

Quantitative: The quantitative data measured by the experiment mainly consists of two parts. The first part is the driver's braking reaction time T. we define T as the time from when ADAS issues a warning to when the driver activates the brake pedal. The second part of the data is the Nasa-tlx workload scale, which includes six indicators, including mental demands, physical demands, temporal demands, performance, effort, and frustration. Each indicator responds to a scale of 11, with 0 corresponding to "very low" and 10 corresponding to "very high".

Qualitative: After each experiment, we conducted interviews with each participant. Some questions were asked about the effectiveness of the multimodal warnings provided by the ADAS, and the effectiveness of the synesthesia combination of auditory and visual channels. These include: 1. Can the warning provided by ADAS remind you of the occurrence of emergencies and return to the driving task effectively? 2. Does the combination of multimodalities (visual and auditory) affect your driving? 3. Which combination of visual and auditory information is more effective?

RESULTS

QUANTITATIVE RESULTS

Pearson bivariate analysis was conducted on driver braking response time (T) and experimental task (synesthesia or non-synaesthesia) in LLS and HLS. T and experimental task have a significant correlation ($p < 0.01$). It can be seen that the response time T of the synesthesia group was smaller than that of the non-synaesthetic group especially in HLS (see Figure 2). It shows that synesthetic improves the speed of multimodal warning delivery in both LLS and HLS.

Pearson correlation coefficient analysis was carried out between experimental tasks and 6 indicators of Nasa-tlx. The correlation coefficient between the overall workload and experimental tasks $p < 0.01$, indicating a significant correlation. The correlation coefficients of mental demand, temporal demand, effort and experimental task $p < 0.01$, indicating significant correlation.The correlation coefficient between frustration , performance and experimental task $p < 0.05$, indicating a significant correlation. There was no significant correlation between physical demand and experimental task.

The mean and standard deviation of the 6 indicators of the experimental task (synesthesia and non-synaesthesia) and Nasa-tlx were analyzed (see

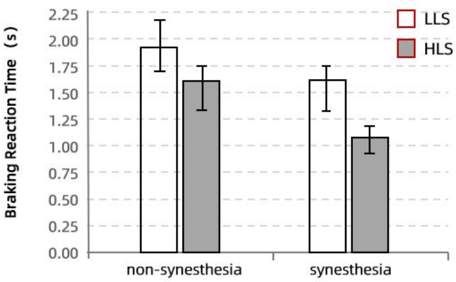

Figure 2: Response times to warning signals in LLS and HLS.

Table 1. Mean and standard error of subjective ratings for Nasa-tlx.

NASA-TLX SCALE														
Warning	Mental Demands		Physical Demands		Temporal Demands		Performance		Effort		Frustration		Total Score	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Non-synesthesia	3.78	0.62	2.56	0.34	4.78	0.80	6.22	0.60	3.56	0.41	2.44	0.65	23.33	1.90
Synesthesia	1.78	0.40	1.78	0.36	3.56	0.74	7.22	0.80	2.11	0.42	1.89	0.42	18.33	2.03

Table 1). The comparison results show that the multi-mode warning combined with synesthesia can effectively reduce the workload of automatic driving.

QUALITATIVE RESULTS

Summarize the interview results. In question 1, almost all participants believed that the warning information provided by the ADAS could effectively remind the occurrence of emergencies. In question 2, most participants believed that multi-modal warning messages would affect their driving behavior, and most participants believed that sound signals were more suggestive than visual signals (light band, HUD). In question 3, Group A of the synesthesia high frequency/low frequency combined form of sound or visual information has no obvious tendency, Participants in Group A have no significant tendency to the two combined forms of non-associative audiovisual information,The number of people in Group B who choose the combination form of "high-frequency sounds & red light & large area of HUD warnings", is far more than that choose "low-frequency sounds & blue light & small area of HUD warnings". This also reflects that synesthesia effect can affect multimodal warning signals to some extent.

CONCLUSION

This paper investigates whether multimodal warning based on synesthesia can reduce driver reaction time and reduce their workload in an autonomous driving scenario. we designed a form of synaesthesia "sound frequency - light color - HUD ratio", and divided the participants into Group A (non-synaesthesia) and Group B (synaesthesia) respectively for a comparative experiment in both LLS and HLS.

For the first question, it was found that, the braking reaction time T of participants in Group B was significantly smaller than that of participants in Group A in HLS, which could prove that synesthesia-based multimodal warning could improve the reaction speed of drivers in emergency situations. Compared with that in HLS, the reaction time T between the two groups in LLS are not so different, which is probably because people have a certain degree of trust in AVs, so they are not in a nervous state to make immediate response when non-emergency situations occur. At the same time, the number of participants who raised their head or started braking in LLS was significantly lower than that in HLS, which also proved the above phenomenon to some extent.

For the second question, the results of Nasa-tlx scale showed that the overall workload of group B was significantly smaller than that of group A, which in particular could effectively reduce the mental demand, temporal demand, effort level and frustration level of drivers. This revealed that the synesthesia-based multimodal warning could effectively reduce driver's workload compared with non-synesthetic warning. Through sorting out the interview results, it wa found that participants have a higher recognition of the synesthesia combination of "high-frequency sound, red light and large HUD area ratio", which is probably because the overall workload of the driver is higher in the emergency scene.

The limitation of this study is that the experiment is carried out on a simulator. On the one hand, it is different from the real autonomous driving environment. On the other hand, the driver does not face actual risk, and the safe indoor simulation environment may affect their state and judgment. Therefore, further simulation is needed in the future to prove the validity of the experimental results.

REFERENCES

Almén, L. (2002, September). Comparing audio and tactile inputs as driver attention control. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 46, No. 21, pp. 1777–1781). Sage CA: Los Angeles, CA: SAGE Publications.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. Journal of vision, 10(1), 6–6.

Kuehn, M., Hummel, T., & Bende, J. (2009, June). Benefit estimation of advanced driver assistance systems for cars derived from real-life accidents. In 21st International Technical Conference on the Enhanced Safety of Vehicles ESV (Vol. 15, p. 18).

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. The American journal of psychology, 173–188.

Robert Jr, L. P. (2016, February). Monitoring and trust in virtual teams. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 245–259).

Shadrin, S. S., & Ivanova, A. A. (2019). Analytical review of standard Sae J3016 ≪taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles≫ with latest updates. Avtomobil'. Doroga. Infrastruktura., (3 (21)), 10.

Sun, X., & Zhang, Y. (2021). Improvement of Autonomous Vehicles Trust through Synesthetic-Based Multimodal Interaction. IEEE Access, 9, 28213–28223.

Sun, X., Li, X., Ji, L., Han, F., Wang, H., Liu, Y., ... & Li, Z. (2018). An extended research of crossmodal correspondence between color and sound in psychology and cognitive ergonomics. PeerJ, 6, e4443.

Turk, M. (2014). Multimodal interaction: A review. Pattern recognition letters, 36, 189–195.