scientific reports

Check for updates

OPEN

Effects of visual interface presentation location and auditory information on takeover performance in multimodal takeover interfaces for high-speed trains

Yunan Jiang & Jinyi Zhi✉

Due to the rapid development of high-speed train automatic driving, the combination forms of multimodal interface are various. In order to determine the differential effects of visual interface location and auditory interface on takeover efficiency and validity in high-speed train automatic driving scenario, 48 participants were selected to conduct multimodal interface simulation takeover test of high-speed train automatic driving, and the takeover reaction time, takeover completion time, takeover errors, SUS scale and generalized anxiety disorder scale were analyzed by variance analysis and Mann-Whitney U test. The results show that non-speech type auditory interface can attract drivers 'attention faster, and has significant advantage in takeover response, but it should avoid the anxiety caused by startle effect when used. Finally, reasonable TOR interface design suggestions for high-speed train autopilot are formed, which can provide theoretical basis for interface design of high-speed train autopilot, and provide new research methods for evaluation and optimization of train man-machine cooperation.

Keywords High-speed train, Automated driving, Multimodal, Takeover task

With increasing urbanization, high-speed trains are becoming the primary means of transportation between cities, providing residents with fast, convenient travel[1]. Accordingly, high-speed-train development and the related technologies have received extensive attention. In particular, autonomous driving technology has attracted considerable attention because it is directly related to public safety[2]. At present, research on autonomous driving technology and its application mainly occurs in the automobile field, especially with regard to new-energy vehicles, where the research is at a more mature stage. Thus, there remains insufficient research on automatic driving for high-speed trains. These trains have a long braking distance of 3.7 km when the speed reaches 270 km/h[3], thus requiring a safer and more efficient autopilot sudden takeover system. Currently, automatic driving of high-speed trains mainly adopts the driver-on-duty automatic driving mode, which is similar to the automatic driving mode of automobiles. In the case of an unexpected crisis or unclear driving environment, it will quickly send out a takeover request (TOR) through the warning interface, asking the driver to complete the takeover of the vehicle as soon as possible[4]. Since human driving reactivity is limited, it is necessary to activate the driving state to the driver in the shortest possible time and help the driver understand the takeover situation.

The current high-speed-train autopilot system operates as a conditional autopilot stage in which the driver needs to take over the task[5]. This stage is unavoidable and an important transitional stage in the development of automated high-speed-train driving. The manual takeover of automatic high-speed-train driving is relatively simple, prompted by the visual interface of the driver console display. This type of prompting method cannot ensure that the driver will notice and complete the takeover in a timely manner or accurately understand the current driving environment and takeover tasks. In research on automated driving takeover prompts, multimodal interfaces have been widely verified as an effective warning method for TOR. Yun et al.[6] found

Department of Industrial Design, Southwest Jiaotong University, Chengdu 610031, China. ✉email: luvfeather21@outlook.com

that purely visual warning interfaces had the worst takeover efficiency, while a combination of visual–auditory–tactile warnings showed the best performance. Monsaingeon et al.[7] noted that multimodal interface takeovers stimulate appropriate levels of driving attention and awareness. Lee et al.[8] found that a dual-mode interpretive takeover interface was more effective than a three-channel single interpretive interface. Such studies have revealed that the warning approach of multimodal interfaces is more conducive to improving TOR efficiency and driver takeover performance. However, train driving has its own unique aspects of operation—that is, moving according to the track, not being able to derail, no overtaking of another train on the same track, and traveling much faster than automobiles. Thus, train drivers handle information differently than automobile drivers[9]. As a result, research on multimodal warning interfaces for manual autopilot takeover in automobiles is not directly applicable to high-speed-train autopilot takeover. Studies on multimodal takeover in high-speed trains have noted that multimodal warning interfaces have faster response efficiency than unimodal warning interfaces. While auditory–visual multimodal interfaces have the highest takeover efficiency, haptic interfaces do not improve takeover efficiency[10]. Although studies have demonstrated that multimodal warning interface cues can help drivers complete TOR, more research is needed on the presentation content of different channels in addition to more detailed differentiation.

In addition to introducing different modalities to help drivers perform more efficient TOR takeovers, studies have investigated more detailed multimodal interface elements. There are two main forms of TOR cueing interfaces. One is same-side cueing, in which the interface is based on stimulus–response compatibility (SRC)[11]. For example, when the vehicle encounters an obstacle or other vehicles on the left side, the driver receives a warning cue on the left side of the dashboard, indicating that the driver needs to pay attention to the traffic situation on the left. The second type is contralateral cueing—that is, signaling the opposite of the desired response (reverse SRC) [12,13,14.15]. This category operates in the form of a warning cue presented on the right side of the dashboard when an obstacle or other vehicle is encountered on the left side, telling the driver to steer or drive to the right to avoid obstacles on the left. The underlying logic of these two alert positions is different. The same-side alert tells the driver to pay attention to the driving environment and situation on that side and make a judgment about safe driving behavior. The opposite-side prompt directly communicates driving operation and behavior so that the driver can more quickly apply the correct safety operation. Studies have investigated the advantages and disadvantages of the two types. Cohen-Lazry et al.[12] found that drivers responded faster to ipsilateral TOR, while Chen et al.[13] found that contralateral signaling had a shorter operational response time. However, in the driving environment of a high-speed train, there are no situations involving being parallel to other trains or being overtaken by the same track, and the driver's cab console has a far richer set of components for operation than is the case for automobile driving. The different components of a train driver's console partly surround the driver in a ring from left to right. Therefore, for TOR, the driver needs to quickly understand the takeover task and the train's traveling status, accurately find the components in the corresponding area, and complete the corresponding takeover task. Thus, the location of cues on the visual channel is also crucial for TOR in high-speed-train automation.

The auditory channel is also an important part of the multimodal takeover warning interface. Auditory TOR tends to capture the driver's attention faster and is preferred over TORs with visual information[14]. Research on the auditory elements of multimodal TOR warnings has mainly categorized sounds into two types: speech and nonspeech alarms[15]. Hong and Yang[16] found that speech-type sound elements, which convey information that could increase the driver's mental load, might be more distracting than simple nonspeech alarm sounds and therefore have implications for takeover response efficiency. Speech-based sounds can convey specific takeover tasks or operational TOR processes and elaborate on the current driving scenario. The driver might therefore have to work to make sense of the speech content, resulting in a certain cognitive load that leads to slower response times but provides a clearer understanding of the post-response takeover task. Choi and Ji[17] found that building trust is one of the most influential factors for drivers' adoption of autonomous driving technology. Speech-interpreted TOR, meanwhile, can help drivers understand the driving environment and takeover tasks, improve or calibrate trust in highly automated systems, and reduce perceived risk owing to the interpretation[18]. Moreover, nonspeech alarm tones can quickly draw the driver's attention from other tasks, cause the driver to quickly notice the current driving environment, and help the driver think about what driving maneuvers need to be performed. Faster reaction times mean less risk for TOR, especially for faster vehicles such as high-speed trains.

In summary, the use of multimodal TOR in automated driving of high-speed trains is still imperfect. There is a need to investigate the effects of different combinations of visual and auditory elements on the response performance and operational performance of TOR in automated driving of high-speed trains. This study's research questions are as follows:

1: What is the effect of the location of the visual channel in multimodal TOR on the driver's takeover response and takeover operation performance in high-speed-train autopilot mode?

2: What are the effects of speech and nonspeech alarm tones in the auditory channel of multimodal TOR on driver takeover response and takeover operation performance in high-speed-train autopilot mode?

Based on this study's experimental results, optimization suggestions are provided for the audiovisual design of multimodal TOR for high-speed trains using autopilot.

Method

Participants

Forty-eight participants (24 males and 24 females) were recruited for the experiment and divided into four groups, with the same number of males as females in each group. All participants were 18–30 years old (M = 22.3; SD = 1.9). All participants were informed about the purpose of the experiment. Complete SCL90 psychological test, vision and hearing tests before the experiment. All participants were in good physical condition, had good

mental state, had no cognitive impairment, normal vision after correction, and no color blindness, weak color and other problems.

Experimental equipment

(1) Simulation table: Simulator: The simulator consists of a high-speed train driving simulator on a fan console, four Jih-Sun QC1009 touchscreens (13 In) and a large screen consisting of 16 small screens. The buttons on the fan console handle most of the functions of the high-speed train console, and the four touchscreens on the fan console are used to simulate the human-machine interface of the high-speed train. In the experiment, subjects were required to make multimodal TOR-based responses using the console. The large screen simulated the front window of a high-speed train cab and displayed the traveling conditions (Fig. 1).

(2) Multimodal TOR materials: Different visual and auditory autopilot takeover cue request materials were designed using PS 2019 cc software and Microsoft Voice. The design was based on takeover tasks, which were categorized into three types: (1) checking whether the driving conditions (battery voltage, total air cylinder pressure) were normal through the right HMI, (2) data transmission through the left CIR screen, and (3) braking operation through the left driver controller. For the visual material, prompts were provided through different icons and texts (Fig. 2). For the auditory material, prompts were provided through alarm tones and voices containing instructions to take over the task. The warning time of visual prompt information and auditory information is designed according to the real driving situation. When the driver finds the prompt and takes over the operation process, the warning prompt will always respond. When the driver needs to click the tablet after completing the takeover task, the multi-modal warning interface will stop responding at this time, so as to simulate the driver completing the takeover operation.

(3) Other equipment: three iPads 2018 for displaying the multimodal TOR and two Sony FX30 video cameras to record the experiment.

Variable design

The independent variables were the audiovisual elements of the multimodal TOR species: the location of the visual element presentation and the cue content of the auditory element. The location of the visual element was divided into the fixed location at the center of the driver's desk and the corresponding orientation of the cue according to the orientation of the task to be taken over. The cue content of the auditory element was divided into two categories: voice cues containing the operation of the takeover task and nonvoice alarm sounds. The two sets of independent variables were combined with each other for four sets of multimodal TOR combinations: visually fixed speech TOR, visually moving speech TOR, visually fixed nonspeech TOR, and visually moving nonspeech TOR.

The dependent variables included takeover reaction time, takeover task completion time, number of takeover task manipulation errors, the System Usability Scale (SUS), and the Generalized Anxiety Disorder scale. Takeover reaction time can reflect the driver's response speed under multimodal TOR. Takeover task completion time and the number of takeover task operation errors can reflect the driver's takeover efficiency. SUS reflected the driver's subjective assessment of the usability of multimodal TOR. The Generalized Anxiety Disorder scale assessed drivers' takeover stress and anxiety after different multimodal TORs. The specific indicators are listed below:

(1) Takeover response time. Time calculation started with the multimodal TOR starting to respond to the cue and ended with the driver starting to perform the takeover task operation.

(2) Takeover task completion time. Time was calculated from the start of the takeover task by the driver; the end time was calculated from the completion of the takeover task by the driver.

Fig. 1. High-speed train driving simulator and presentation of different visual takeover requests.

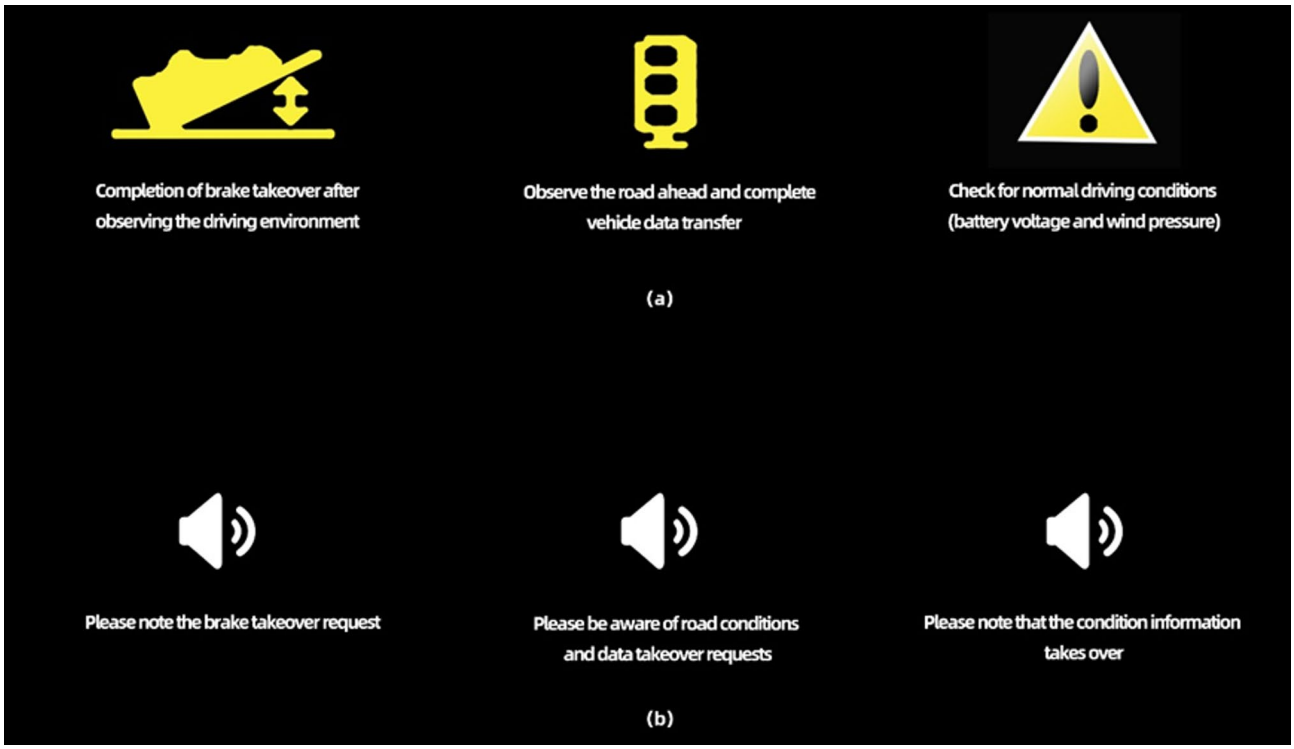

Fig. 2. Visual takeover cue icon with textual cue for takeover task (**a**), auditory takeover cue content for voice type (**b**), and normal alarm tone for non-voice type auditory takeover cue.

(3) Number of takeover operation errors. The number of errors that occurred during the takeover task was recorded; the more the errors, the more inefficient the takeover.

(4) SUS. This scale consists of 10 questions, rated on a five-point Likert scale (completely disagree, disagree, average, agree, and strongly agree).

(5) Generalized Anxiety Disorder Scale. A seven-item version was selected to derive an anxiety score (from 0 to 21), with each item answered on a 4-point scale (0 = not at all, 1 = weak, 2 = moderate, 3 = strong). The original version of the questionnaire screened for the frequency of anxiety symptoms in the past two weeks, i.e., the trait score. To assess the transient nature of stress, the questions were adapted to create a response schedule for the day (e.g., today, do you feel nervous, anxious, or tense?).

Experimental procedure

First, the participants were given an explanation of the purpose and tasks of the experiment. They learned to operate the multimodal TOR and takeover task and were tested to ensure they had sufficient familiarity and could complete the train autopilot takeover operation. A pre-experiment was conducted to ensure that the participants understood the process of the experiment.

Second, the participants were divided into four groups, corresponding to four multimodal TORs with different audiovisual combinations. There were 12 participants in each group with an equal number of males and females. Personal information was recorded.

Then, a video of high-speed-train driving was played through the simulation table to simulate the automatic driving state of high-speed trains. Participants were seated in front of the simulated driving platform, and different numbers of iPads were placed according to the group: one iPad was placed at the center of the driving platform for the visual fixation group. For the visual mobility group, one iPad was placed in the left, middle, and right positions of the driving platform; these were used to provide multimodal TOR prompts corresponding to the different takeover task orientations. Meanwhile, different auditory elements were played according to the group. To ensure that the auditory channels of the multimodal TOR were in the same location, both speech and nonspeech TOR were played from the iPad at the center of the driver's desk.

Next, the participants needed to perform the takeover task operation the first time the multimodal TOR was found during the autopilot process. Each participant needed to perform the takeover task operation three times during the whole process. Each takeover task operation was different.

Finally, after the participants completed the train autopilot takeover simulation experiment, the experimenter had the participants fill out the questionnaire and conducted short post-experiment interviews to ask the participants about their feelings during the experiment. Acknowledgments were made and the experiment was concluded.

Results

The experimental data on takeover reaction time, takeover task completion time, and questionnaire scores were subjected to the S-W test and ANOVA chi-square test. The data conformed to normal distribution and ANOVA chi-square ($p > 0.05$). Therefore, two-way ANOVA was used for takeover reaction time and takeover task completion, while a nonparametric Mann–Whitney U test was performed on the experimentally obtained number of takeover errors, SUS scale, and anxiety stress scale.

Takeover reaction time

Two-factor ANOVA was conducted to investigate the effect of multimodal visual position and the auditory elements of TOR on takeover reaction time. Visual position did not show significance ($p > 0.05$), while the auditory elements did show significance ($p < 0.05$), as shown in Table 1. This indicates that visual position did not produce a differential effect on takeover reaction time, while auditory elements did produce a differential effect on takeover reaction time. Moreover, there was no significance ($p > 0.05$) in the relationship between visual location and auditory elements, indicating that there was no second-order effect between the two. One-way ANOVA on the auditory element differences revealed that takeover reaction times were significantly lower ($p < 0.05$) in the nonspeech mode than in the speech mode. Data with significant differences corresponded to effect sizes greater than 0.8, giving high validity. Figure 3 shows a comparison of takeover reaction times for multimodal TORs with different auditory element types.

Takeover completion time

Two-way ANOVA was conducted to investigate the effect of visual position and auditory elements on takeover completion time. Visual position showed significance ($p < 0.05$), and auditory elements also showed significance ($p < 0.05$). There was no significance between visual position and auditory elements ($p > 0.05$), as shown in Table 1. This indicates that there was a main effect of the two types of variables on takeover completion time, but there was no second-order effect between the variables. Therefore, one-way ANOVA was performed on the visual position and auditory elements. The nonspeech modality took significantly longer than the speech modality in terms of takeover completion time ($p < 0.05$). The modality in which vision followed the takeover task took less time than the modality in which vision was fixed at the center ($p < 0.05$). Data with significant differences corresponded to effect sizes greater than 0.8, giving high validity. Figure 4 shows a comparison of takeover completion times for multimodal TORs with different visual positions and auditory element types.

Number of takeover operation errors

A nonparametric Mann–Whitney U test was performed on the number of takeover operation errors. It revealed that visual location did not have a significant effect on the number of takeover operation errors ($p > 0.05$), whereas the auditory element did have a significant effect on the number of operation errors ($p < 0.05$), as shown in Table 2. A later comparison of the medians of the data revealed that the TOR in the nonspeech mode was significantly higher ($p < 0.05$) than that in the speech mode in terms of the number of takeover operation errors ($p < 0.05$). The corresponding effect value $|d|=0.384$, a medium effect. Figure 5 shows a comparison of the median number of takeover operation errors for different types of multimodal TORs.

SUS

Participants were asked to fill out a SUS questionnaire corresponding to the group and task after completion of the experiment to measure the perceived usability of different combinations of multimodal TORs of different types in the high-speed-train driving takeover task. After transforming the participants' SUS scores, it was found that there was no significant difference regarding visual location in the SUS scores based on the Mann–Whitney U test ($p > 0.05$); there was, however, a significant difference between auditory elements in the SUS scores ($p < 0.05$), as shown in Table 2. Comparing the median SUS scores of each group, it was found that the SUS scores for the speech-based TORs were significantly lower than those for the nonspeech-based TORs. The corresponding effect value $|d|=0.752$ is a large effect. Figure 5 shows a comparison of the median SUS scores for different types of multimodal TORs.

	Multimodal TOR groups	Mean and standard deviation (s)	F	Cohen's d	*p*
Takeover reaction time	Visual position fixation	1.81 ± 0.45	0.124	0.09	0.726
	Visual position shift	1.86 ± 0.62			
	Speech-based auditory cues	2.27 ± 0.33	99.689	2.84	0.000**
	Nonspeech auditory cues	1.40 ± 0.28			
Takeover completion time	Visual position fixation	8.31 ± 1.19	52.380	2.09	0.000**
	Visual position shift	5.89 ± 1.13			
	Speech-based auditory cues	6.22 ± 1.57	17.963	1.21	0.000**
	Nonspeech auditory cues	7.98 ± 1.30			

Table 1. Mean, standard deviation, and one-way ANOVA for takeover reaction time and takeover task completion time in different multimodal TOR groups. **represents $p < 0.001$.

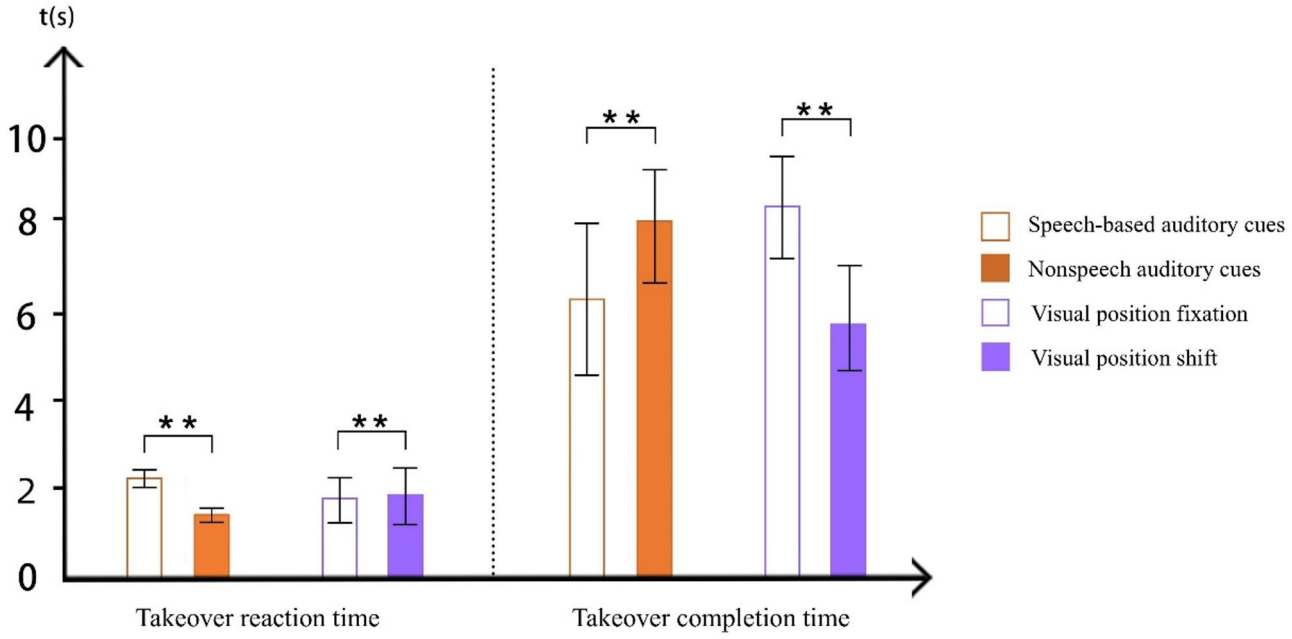

Fig. 3. Average takeover reaction time and average takeover completion time for different TOR combinations (**represents $p < 0.01$).

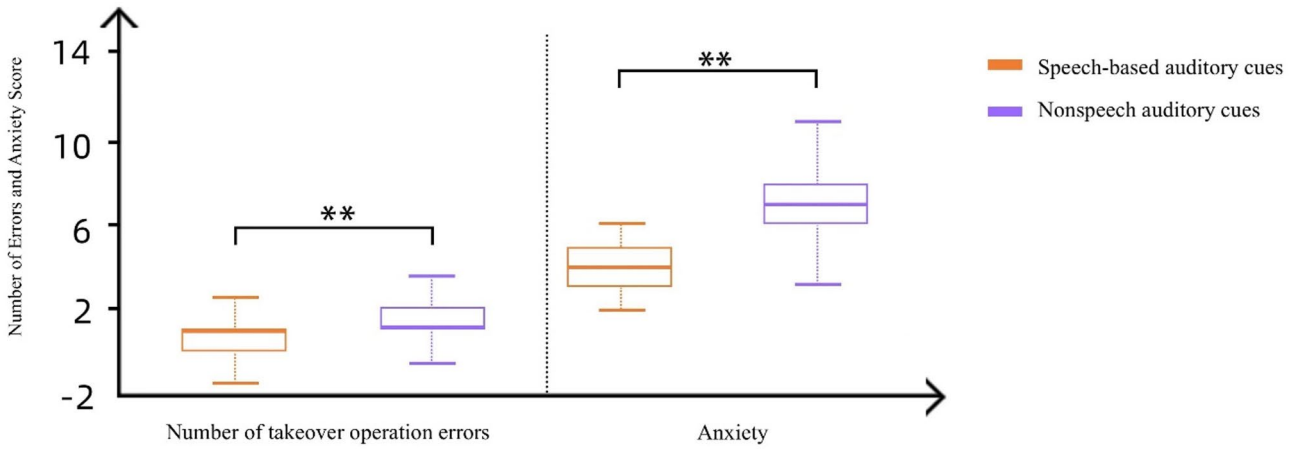

Fig. 4. Speech vs. non-speech auditory takeover cues in the number of takeover errors vs. median Generalized Anxiety Disorder Scale score box plots (**represents $p < 0.01$).

Generalized anxiety disorder scale

The participants' perceived anxiety or stress during the experiment was obtained based on the scales they filled out before and after the experiment. The total scores before the experiment were subtracted from the total scores after the experiment. The Mann–Whitney U test showed that there was no significant difference ($p > 0.05$) in visual location regarding the participants' anxiety or stress; there was, however, a significant difference ($p < 0.05$) in the auditory element regarding the participants' anxiety or stress, as shown in Table 2. Comparing the median anxiety or stress scores of the participants in each group, it was found that the anxiety or stress scores for the speech TOR were significantly lower than those for the nonspeech TOR. The corresponding effect value

	Multimodal TOR groups	Median	z	Cliff's d	p
Number of receivership errors	Visual position fixation	1.000	−1.688	−0.257	0.091
	Visual position shift	1.000			
	Speech-based auditory cues	1.000	−2.521	−0.384	0.012*
	Nonspeech auditory cues	1.000			
SUS score	Visual position fixation	68.750	−1.043	−0.172	0.297
	Visual position shift	70.000			
	Speech-based auditory cues	72.500	−4.560	−0.752	0.000**
	Nonspeech auditory cues	67.500			
Generalized Anxiety Disorder scale	Visual position fixation	5.500	−0.229	−0.038	0.819
	Visual position shift	5.500			
	Speech-based auditory cues	4.000	−4.836	−0.806	0.000**
	Nonspeech auditory cues	7.000			

Table 2. Median and descriptive statistics for the number of errors on the takeover task, SUS scores, and generalized anxiety disorder scale scores under different multimodal TOR groups. *represents $p < 0.05$, **represents $p < 0.01$.

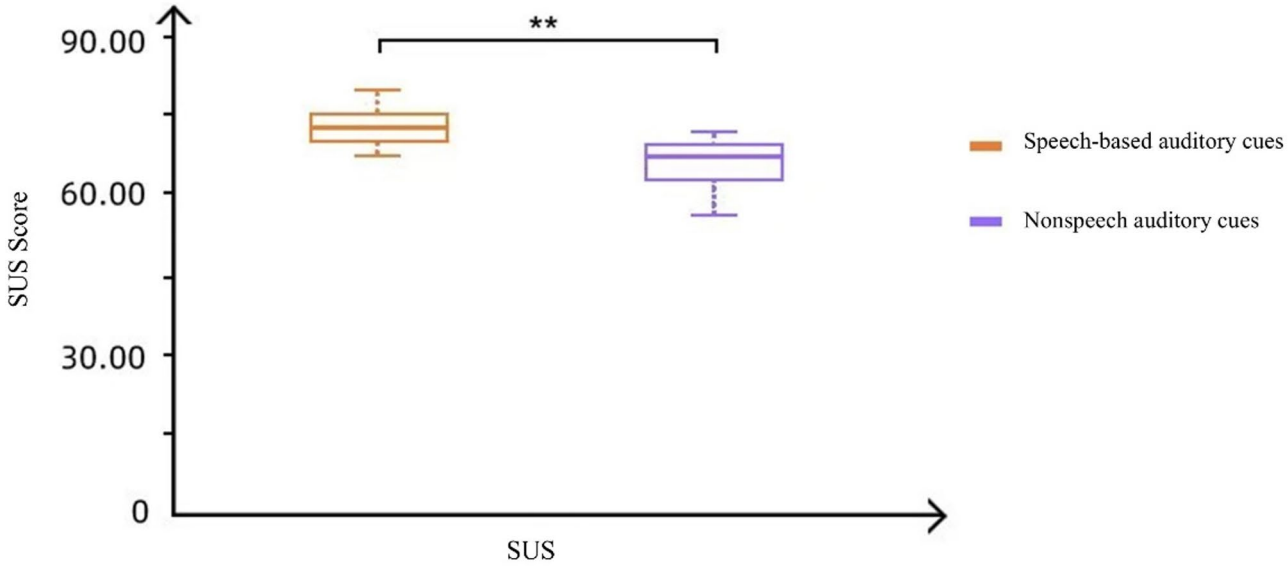

Fig. 5. Speech vs. non-speech auditory takeover cues in the median SUS usability score box plot (**represents $p < 0.01$).

|d|=0.806 is a large effect. Figure 5 shows a comparison of the median anxiety or stress scores for different types of multimodal TORs.

Discussion

This section presents important findings obtained from the analysis of experimental data on a high-speed train autopilot takeover task. The effects of different combinations of multimodal TORs on the performance of high-speed train autopilot takeover are explored. It also collates and retraces the questions asked by some users after the experiment and their feedback on the experimental experience, which is used to provide some explanations and illustrations of the experimental results.

Regarding sound elements in multimodal TOR, voice-type alert tones were not as fast as nonvoice-type alert tones in terms of takeover reaction time in the high-speed-train autopilot takeover task. That is, the non-speech type auditory interface is more efficient in takeover. For takeover reaction time, non-speech type of takeover prompts compared to speech type of prompts, due to their alerting audio, are able to engage the driver faster, wake the driver up quickly from non-driving tasks, react faster and start taking over the task as compared to the flat speech alarm audio. In a study by Edworthy et al.[19] it was stated that sudden warning tones can be quickly picked up by the human auditory system and are easier to localise, and that such sounds are not as easily overshadowed by other sounds. This is one of the potential reasons why drivers are able to respond quickly during takeover cues. This is consistent with the results of studies conducted by Hong and Yang[16] and Forster et al.[20], who found that compared with nonvoice-type alarms, voice-type alarms—which have content information

in addition to sound, which serves as a reminder to the driver while also conveying new information—distract the driver's attention from alarms, resulting in a slower takeover response. Moreover, the information conveyed by voice-type alarms leads to higher cognitive conformity, and because it is a common language that drivers are familiar with, drivers cannot help but want to listen to the meaning of the information conveyed in its entirety[21]. This also contributes to an increase in takeover reaction time, which was confirmed in the participants' post-experiment interviews in this study. The participants mentioned that upon hearing voice-type alerts, they would react to the need to start the takeover but would also wait until they had fully heard the instructions for the takeover task before becoming more comfortable. Furthermore, this type of familiar information can be less stimulating to the driver[22], which may lead to a lesser perception of the sense of urgency that the takeover cue hopes to bring. Other studies, however, have found that alarm tones, owing to their abruptness and warning nature, can attract the attention of a driver who is engaged in a nondriving task more quickly, but they might also have a startle effect on the driver, which can increase takeover reaction time[23]. Therefore, the design of multimodal TOR sound elements needs to be considered in various aspects. It requires the rapid capture of the driver's attention through nonspeech-based alarm sounds while also avoiding sounds that can startle the driver. The first step of cueing can optionally be done through a non-speech annunciation tone from small to large volume, which attracts the driver's attention. Then, through voice-type cues, guidance is provided for specific takeover tasks, and the use of voice-type messages as guidance rather than alarms may be a more usable method of auditory takeover cueing.

Drivers hearing voice-type alarm tones performed better than those hearing nonvoice-type alarms in terms of time to completion and the number of takeover errors. This finding is explained by Koo et al.[18], who showed that the interpretation of voice alarm tones for takeover commands helped drivers understand the current scenario and the required task operations, which could increase trust in autonomous driving and thus reduce risk. Meanwhile, Richard et al.[24] noted that most learning is guided by the comprehension process. Thus, voice-type alarm tones, while prompting the takeover, increased the driver's understanding of the takeover task and clarified the operation, thus helping the driver accomplish the takeover task. In addition, speech-type alarm tones can reduce visual workload during driver–automation interaction[25]. Naujoks et al.[26] similarly found that drivers tended to pay less attention to the visual human–computer interaction interface when provided with an additional speech output. When multiple channels of information are delivered, it can help avoid cognitive overload caused by too much information. In this regard, to effectively help accomplish takeover task prompts in high-speed-train autopilot situations, takeover task guidance can be provided by supplementing voice-like alarm tones. At the same time, if it can be combined with a human-machine collaboration system that provides step-by-step voice guidance based on the driver's operating steps, it may become a takeover cue with a lower error rate.

Visual presentation position There is no significant difference between the two visual interface presentation modes in terms of takeover reaction time for multimodal TOR in high-speed trains. There is no significant advantage or disadvantage in prompting the driver to take over, either by fixing it in the centre of the driver's console or by following the takeover task operation position. This may be due to the fact that auditory cues are dominant for cautionary messages such as alerts or takeover cues[27]. For the current stage of automated driving in high-speed trains, drivers need to observe visual information such as road and vehicle conditions as well as signals at all times, even though it is automated driving. Therefore, information in the auditory channel is more easily captured when there is sufficient visual cognitive load[28]. This also resulted in no significant difference in takeover responses between the two visual cue interfaces. This finding also provides design ideas for multimodal takeover interfaces, where when there is too much information in one modality, it may be possible to intervene by introducing another modality with potentially better results.

Visual takeover cues presented in different positions can facilitate takeover task completion at a faster rate than visual takeover cues that are always presented in the same position. Cohen-Lazry et al.[12] similarly found that ipsilateral visual cues facilitated faster takeover speeds. Meanwhile, Chen et al.[13] found that contralateral visual cues resulted in faster takeover speeds. It has also been found that differences in the location where visual takeover cues are presented do not present a significant difference in takeover speed[14,29]. These studies were conducted on self-driving takeovers in cars. When driving a train, however, the driver is seated in the middle of a fan table and needs to pay attention to far more screens and buttons than is the case when driving a car. There is no overtaking or being overtaken when driving a train, so there is no need to look at the left rear versus the right rear. In this regard, there is no situation where the visual cues are presented on the opposite side. Looking at this study's experimental results, since the train's driver platform has a fan-like shape, when the visual takeover cue was given in the center position, the driver's field of view switched when completing the takeover task, especially for the left- and right-side operations. This could also explain why the visually presented position-following task movement group had faster takeover completion. It was also mentioned in the post-experiment interviews that although the visual takeover cue could appear on any of the three flat panels, the auditory takeover cue compensated for the fact that the driver might not have noticed the visual cue on the two flat panels in the first place. Therefore, after hearing the audio takeover cue, he or she would quickly look for the location where the visual cue appeared and be able to take over the operation in that location. In this regard, based on the operating environment and task characteristics of the train driver's platform, the visual presentation position of the multimodal TOR can be designed to be closer to the direction of the takeover operation.

Meanwhile, regarding the SUS scores, the voice-based alarm tones and the presentation of visual takeover cues based on the location of the takeover operation had higher scores. With the nonvoice alarm tones, the participants' anxiety perceptions were also significantly higher than in the case of voice alarm tones. The fact that nonvoice alarm tones tended to produce takeover anxiety in drivers may have been the reason for their low usability scores. Nonvoice-based alarm tones can attract attention more quickly but can have a startle effect, produce anxiety, and fail to improve takeover performance. Thus, using fading nonspeech alarm tones

as a prelude to auditory elements, followed by speech-like alarm tones for takeover task instructions and visual takeover cues presented to follow the location of the operating task, could be better for multimodal high-speed-train autopilot takeover TORs.

Conclusion

This study compared participants' takeover reaction time, takeover performance, usability score, and takeover task anxiety for multimodal TORs with different visual cue presentation locations and different auditory cue elements in combination with each other in a high-speed-train driving autopilot takeover experiment. It was found that for takeover reaction time, nonspeech auditory takeover cues had a significant advantage over speech auditory takeover cues owing to their ability to attract drivers' attention faster. In terms of takeover task completion performance, drivers with voice-based auditory takeover cues performed better and were able to understand the takeover task operation through voice messages, which helped them understand the current takeover situation and perform the takeover operation more quickly. In terms of the presentation position of visual cues, the completion time of visual cues distributed according to the location of the takeover operation was less than that of the one presented in a fixed position. The presentation of following the location of the operation task reduced the driver's line-of-sight switching during the takeover operation and enabled the driver to complete the takeover task more quickly. Regarding SUS scores, the voice-based takeover prompts and the visual presentation of following the location of the takeover operation both had high scores. While nonvoice-based takeover cues can speed up the driver's takeover response, the anxiety caused by the startle effect needs to be avoided. Therefore, the abruptness of nonvoice-based prompts should be weakened. This study's results can provide a theoretical basis for the design of multimodal TOR for high-speed-train autopilot takeover and also provide a research method for train autopilot takeover.

This study has some limitations at this time due to the objectives of the study and the number of subjects. Since some driving experience is required, the subjects selected were 48 driving students who had gone through at least 3 months of relevant driving courses in traffic schools. The number of subjects just meets the experimental conditions, which is not very sufficient. At the same time, the results of the experiment show that non-voice prompts and voice prompts have their own advantages, but due to the current experimental environment and the subject situation, this study has not been able to present a more accurate quantitative study. In this regard, in the subsequent study, we will ensure the adequacy of the sample of subjects by finding more subjects to participate in the driver training programme and thus participate in the experiment. Meanwhile, we will further divide the broadcast time of non-verbal cues and verbal cues, and combine the two to investigate whether the combined auditory cues can bring about less anxiety perception and better takeover performance. Further enriching the continuity and usefulness of the study.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 10 March 2025; Accepted: 4 June 2025
Published online: 05 June 2025

References

1. Meiqi Wang, Y. et al. A hybrid triboelectric-piezoelectric-electromagnetic generator with the high output performance for vibration energy harvesting of high-speed railway vehicles, nano energy, **132**, 110417, ISSN 2211–2855, (2024). https://doi.org/10.1016/j.nanoen.2024.110417
2. Singh, P. et al. Deployment of autonomous trains in rail transportation: current trends and existing challenges, in IEEE access, **9**, pp. 91427–91461, (2021). https://doi.org/10.1109/ACCESS.2021.3091550
3. Heping, L. & Huting, L. A study on the design of fundamental brake system for high speed train. *China Railw Sci.* **24**, 8e13 (2003). accessed on 25 November 2022). http://www.cqvip.com/qk/97928x/200302/7662916.html
4. Naujoks, F., Purucker, C., Wiedemann, K. & Marberger, C. Noncritical state transitions during conditionally automated driving on German freeways: effects of Non-Driving related tasks on takeover time and takeover quality. *Hum. Factors.* **61** (4), 596–613 (2019). Epub 2019 Jan 28. PMID: 30689440.
5. Lisheng Jin, X. et al. Impact of non-driving related task types, request modalities, and automation on driver takeover: A meta-analysis, safety science, **181**, 106704, ISSN 0925–7535, (2025). https://doi.org/10.1016/j.ssci.2024.106704
6. Yun, H. & Yang, J. Multimodal warning design for take-over request in conditionally automated driving. *Eur. Transp. Res. Rev.* **12**, 34. https://doi.org/10.1186/s12544-020-00427-5 (2020).
7. Monsaingeon, N., Caroux, L., Langlois, S. & Lemercier, C. Multimodal interface and reliability displays: effect on attention, mode awareness, and trust in partially automated vehicles. *Front. Psychol.* **14**, 1107847. https://doi.org/10.3389/fpsyg.2023.1107847 (2023).
8. Lee, S. et al. Investigating effects of multimodal explanations using multiple In-vehicle displays for takeover request in conditionally automated driving, transportation research part F: traffic psychology and behaviour, **96**, Pages 1–22, ISSN 1369–8478, (2023). https://doi.org/10.1016/j.trf.2023.05.014
9. Naweed, A. & Balakrishnan, G. Understanding the visual skills and strategies of train drivers in the urban rail environment. *Work* **47**, 339–352 (2014).
10. Jing, C. et al. Influence of Multi-Modal warning interface on takeover efficiency of autonomous High-Speed train. *Int. J. Environ. Res. Public Health.* **20** (1), 322. https://doi.org/10.3390/ijerph20010322 (2023).
11. Proctor, R. & Vu, K. Stimulus-response compatibility principles: Data, theory, and application. CRC press. Retrieved from. (2006). https://books.google.com/books?hl=en&lr=&id=NISHh4ZJV4AC&oi=fnd&pg=PP1&dq=Stimulus+response+compatibility+principles:+Data,+theory,+and+application&ots=mnEBDhHIV1&sig=q2KkUY7YmaRLrnLaLHc_FYenc5I
12. Cohen-Lazry, G., Katzman, N., Borowsky, A. & Oron-Gilad, T. Directional tactile alerts for take-over requests in highly-automated driving. *Transp. Res. Part. F: Traffic Psychol. Behav.* **65**, 217–226. https://doi.org/10.1016/j.trf.2019.07.025 (2019).
13. Chen, J., Šabić, E., Mishler, S., Parker, C. & Yamaguchi, M. Effectiveness of lateral auditory collision warnings: should warnings be toward danger or toward safety? *Hum. Factors*. https://doi.org/10.1177/0018720820941618 (2020).

14. Petermeijer, S., Doubek, F. & De Winter, J. Driver response times to auditory, visual, and tactile take-over requests: A simulator study with 101 participants. *IEEE Int. Conf. Syst. Man. Cybernetics (2017b)*. https://doi.org/10.1109/SMC.2017.8122827 (2017).
15. Zhang, Q. & Jessie Yang, X. Robert what and when to explain?? A survey of the impact of explanation on attitudes toward adopting automated vehicles. *IEEE Access.* **9**, 159533–159540. https://doi.org/10.1109/ACCESS.2021.3130489 (2021).
16. Hong, S. & Yang, J. H. Effect of multimodal takeover request issued through A-pillar LED light, earcon, speech message, and haptic seat in conditionally automated driving. *Transp. Res. Part. F: Traffic Psychol. Behav.* **89**, 488–500. https://doi.org/10.1016/j.trf.2022.07.012 (2022).
17. Choi, J. K. & Ji, Y. G. Investigating the importance of trust on adopting an autonomous vehicle. *Int. J. Hum Comput Interact.* **31** (10), 692–702. https://doi.org/10.1080/10447318.2015.1070549 (2015).
18. Koo, J. et al. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *Int. J. Interact. Des. Manuf. (IJIDeM).* **9** (4), 269–275. https://doi.org/10.1007/s12008-014-0227-2 (2015).
19. Edworthy, J. Alarms and human behaviour: implications for medical alarms. BJA. *Br. J. Anaesth.* **97** (1), 12–17 (2006).
20. Forster, Y., Naujoks, F., Neukum, A. & Huestegge, L. Driver compliance to take-over requests with different auditory outputs in conditional automation. *Accid. Anal. Prev.* **109**, 18–28. https://doi.org/10.1016/j.aap.2017.09.019 (2017).
21. Jessen, A. et al. Native and Non-native speakers' brain responses to filled indirect object gaps. *J. Psycholinguist. Res.* **46**, 1319–1338. https://doi.org/10.1007/s10936-017-9496-9 (2017).
22. Velmans, M. Is human information processing conscious? *Behav. Brain Sci.* **14** (4), 651–669 (1991).
23. Rydström, A., Mullaart, M. S., Novakazi, F., Johansson, M. & Eriksson, A. Drivers' performance in Non-critical Take-Overs from an automated driving System-An On-Road study. *Hum. Factors*. **65** (8), 1841–1857 (2023). Epub 2022 Feb 25. PMID: 35212565.
24. Alterman, R., Zito-Wolf, R. & Carpenter, T. Interaction, comprehension, and instruction usage. *J. Learn. Sci.* **1**(3–4), 361–398. https://doi.org/10.1207/s15327809jls0103&4_4 (1991).
25. Bazilinskyy, P. & de Winter, J. Auditory interfaces in automated driving: an international survey. *PeerJ Comp. Sci.* **1**, e13. https://doi.org/10.7717/peerj-cs.13 (2015).
26. Naujoks, F., Forster, Y., Wiedemann, K. & Neukum, A. Improving usefulness of automated driving by Lowering primary task interference through HMI design. *J. Adv. Transp.* **2017**, 6105087. https://doi.org/10.1155/2017/6105087 (2017).
27. Opoku-Baah, C. et al. Visual influences on auditory behavioral, neural, and perceptual processes: A review. *JARO* **22**, 365–386. https://doi.org/10.1007/s10162-021-00789-0 (2021).
28. He, Y. et al. and. Effects of audiovisual interactions on working memory task Performance—Interference or facilitation Brain Sciences **12**, 7: 886. (2022). https://doi.org/10.3390/brainsci12070886
29. Petermeijer, S., Bazilinskyy, P., Bengler, K. & de Winter, J. Take-over again: investigating multimodal and directional TORs to get the driver back into the loop. *Appl. Ergon.* **62**, 204–215. https://doi.org/10.1016/j.apergo.2017.02.023 (2017a).

Acknowledgements

This study was approved by the Ethics Committee of Southwest Jiaotong University and conducted according to the principles of the Declaration of Helsinki. All the participants provided written informed consent before participating. We thank LetPub (www.letpub.com.cn) for its linguistic assistance during the preparation of this manuscript.

Author contributions

Jiang Yunan conducted experimental research and data analysis as well as the writing of papersJinyi Zhi carried out the determination of the topic of the paper, the analysis of the experimental data and the writing of the paper.

Declarations

Competing interests

The author(s) declared no potential conflicts of interest with respect to the research, author- ship, and publication of this article.

Additional information

Correspondence and requests for materials should be addressed to J.Z.