IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 29, 2021, accepted February 8, 2021, date of publication February 12, 2021, date of current version February 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3059071

Improvement of Autonomous Vehicles Trust Through Synesthetic-Based Multimodal Interaction

XIAOFENG SUN[1] AND YIMIN ZHANG[2]

[1]School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China
[2]Equipment Reliability Institute, Shenyang University of Chemical Technology, Shenyang 110142, China

Corresponding author: Yimin Zhang (zhangyimin@syuct.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant U1708254.

ABSTRACT Trust is the key factor for people to accept autonomous vehicles(AVs). Existing studies have reported that multimodal interaction would enhance people's trust in AVs. However, these researches mainly focus on the superposition effect between sensory channels, and lack on the research of correlation between different sensory channels and its influence on AVs trust. Therefore, we innovatively introduce synesthesia theory for the research of improving AVs trust. We present an AVs multimodal interaction model based on audio-visual synesthesia theory, and finally prove that the model has a definite effect on improving AVs trust by experiments. Firstly, 82 participants are recruited and assigned into two groups: Group A (non-synesthesia group) and Group B (synesthesia group). They conduct an experimental driving experienced normal traffic conditions (NTC) (turning, traffic lights, over and limit speed prompts) and emergency traffic condition (ETC) (sudden braking of the car in front, temporary lane change, pedestrian thrusting) while completing a secondary task. Then, we conduct a survey (questionnaire and interviews) to evaluate the attitude about trust, technical competence, situation management and perceived ease of use after participants finished experimental driving. The results demonstrate that synesthetic-based multimodal interaction (SBMI) can more effectively remind people of relevant information especially under ETC. SBMI model is more effective than single information stimulus or non-synesthetic audio-visual information stimulus not only in terms of information transmission efficiency and effect, but also in terms of output response/ action. The results also show that SBMI contributes to the improvement of AVs trust. These findings provide evidence on the importance of SBMI to the improvement of AVs trust. The findings of this study will be helpful to the future design of AVs interaction system.

INDEX TERMS Autonomous vehicles, trust, synesthesia, multimodal interaction.

I. INTRODUCTION

The development of AVs has been considered as one of the most promising directions by many experts in the field of automobile manufacturing all over the word. AVs will undoubtedly become the next revolution in the development of the automobile industry and will have a fundamental impact on human life. On March 9, 2020, the Ministry of Industry and Information Technology, PRC published the Classification of Auto Driving Automation on its official website, which is used to guide and regulate the development of AVs.

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

At present, the development of AVs is at the stage of level 2 to level 3, which is a ''Shared control'' model of human-computer co-driving. Accordingly, more and more researchers have been focused on the man-machine trust during driving time. User and public trust are considered as critical factors for the success of AVs. It is evident that we still lack effective means of establishing a basic understanding about trust in autonomous vehicles and many doubts and misunderstanding prevail in the general public. Thus, trust in automation is a critical factor that determines acceptability, attitude, behavioral intention, and actual use of AVs [1]. Therefore, how to enhance individual's trust in AVs plays a crucial role in the acceptance of AVs.

A. TRUST IN AVs

According to Ghazizadeh *et al.* [2], trust is a way that humans mitigate uncertainty, so trusting an automated system is essential in uncertain conditions. In the AVs' context, we build upon the trust definition provided by Lee and See [3], who state trust to be ''the attitude that an agent will help achieve an individual s goals in a situation characterized by uncertainty and vulnerability''. Trust therefore shapes an individual's attitudes and ultimately determines their behavior, such as their intention to use the system.

Initially, trust in automation evolves alongside the three dimensions of predictability, dependability, and faith [4]. Lee and Moray propose a general theory for trust in automation which describes trust in terms of three dimensions [5]: performance (what the system does), process (how the system is built), and purpose (why the system does something). Likewise, Ekman *et al.* state that trust is built on the possibility to observe the system's behavior (performance), understand the intended use of the system (purpose), as well as understand how it makes decisions (process) [6]. Additionally, Parasuraman and Miller note that one factor for trust is that the driver is able to compare the actions of the car and actions they would perform in different scenarios [7], so you have a psychological prediction about the reliability of the actions taken by the car system. In a word, individuals' trust in AVs should be based on their understanding of the function, usability and operation of the system.

Many scholars have conducted researches on improving trust in AVs, for example users' trust levels can be increased by presenting information about the system state and upcoming actions using augmented reality visualizations [8], speech commands [9], different degrees of anthropomorphism [10], etc. Additionally, Ekman *et al.* [11] present 11 key factors, such as Uncertainty Information, or Why and How Information to influence users' trust levels. However, these research only focus on the AVs' technology, design and so on while ignore the human factors. If vehicles respond slowly or inaccurately, individuals would assume the system is faulty and unreliable. Individuals can gain a deeper understanding of the system and create a sense of ''everything is under control'' by understanding the system in real time, thus increasing trust in the system, which is in line with the three levels of situation awareness (perception, understanding and projection) [12].Therefore, to improve the AVs trust, we should increase the technical competence situation management and perceived ease of use as well as the human-computer interaction mode.

In summary, AVs' trust focuses on people's feelings of the system and the feedback given by the system. Beggiato and Krems find that when participants had correct expectations regarding the behavior of the car's automated functionalities, they trusted the engaged more in vehicle [13]. Therefore, it is a key aspect of building trust to present the information received by the AVs system in a timely and effective way within the range of sensory channel information received by the driver or passenger according to the safety level.

B. SYNESTHESIA

Human behavior is not determined by objective factors, but rather by the user's subjective perceptions, based on their individual attitudes, expectations and experience. Thus, even a well-designed system that evidently performs effectively and without inflicting a negative or injurious outcome may not necessarily warrant a user s trust or acceptance. The cognitive theory holds that human behavior is determined by his perception and process of social context. As such, we look for guidance to psychology. Synesthesia is feature mapping process that between the different feelings in psychology. As a system, people's sense organs are interrelated in a certain sense. When a sensory organ is stimulated, it will cause other sensory responses in addition to its own response, so as to resonate among different senses and realize the expansion of sensation and cognitive enhancement. Synesthesia can occur between all five senses: auditory, visual, taste, touch and smell, and can occur in any combination.

Initially, synesthesia is studied as a case in a special population. However, with the progress of technology and the continuous efforts of scholars, the research on synesthesia is no longer limited to individual cases but a common phenomenon in some parts. At present, the focus of synesthesia research is audiovisual synesthesia. The classic conclusion is that the tone and hue of sound are mainly determined by the auditory wavelength energy and visual visible energy, both of which are wave energy [14]. Based on this theory, experiments often compare chromatic scales of sound with color rings, under which sounds of a particular frequency are found to correlate with colors of a particular spectrum. In addition, there are corresponding relations between the lightness of color and the loudness of sound, the saturation of color and the timbre of sound, the hue of color and the tone of sound. Experiments have shown that cross modal synesthesia between sound frequency and color brightness can guide visual attention. For example, high-frequency sounds will direct visual attention to light-colored objects, while low-frequency sounds will focus visual attention on dark-colored objects. Eye-tracking experiments have shown that this phenomenon is automatic, goal-less and subconscious and takes precedence over the effects of semantic associations, even when there is a clear indication to the contrary [15].

Several researchers start to investigate the impact of cross modal correspondences on human information processing using the speeded discrimination task. For instance, Bernstein finds that participants responded more slowly to visual stimuli when the pitch of sound in the task is inconsistent with them in his research [16]. Marks conduct a series of discrimination experiments between vision and hearing. He point out that there is a strong correspondence between certain properties of vision and hearing (e.g., pitch-brightness and loudness-lightness) and concluded that subjects responded

more quickly and accurately with matching stimuli than mismatching stimuli from the two modalities [17]. Evans demonstrates that there were strong cross modal correspondences between auditory pitch and visual location, size, spatial frequency [18]. Synesthesia, as a kind of phenomenon accompanied with human mental activity, plays an important role in the process of perceiving the properties of objects, receiving and communicating information.

At present, the AVs between L2 and L3 requires the participation of human, which will inevitably result in the information exchange, communication and feedback between human and the AVs system. In this process, the synesthesia effect can be used to design and change the environment experience in AVs, so as to realize the collaborative interaction between multi-senses and the AVs system, improve the efficiency of the interaction between people and the system, and improve people's experience and trust in the system.

C. MULTIMODAL INTERACTION (MMI)

With the development of artificial intelligence, virtual reality and augmented reality technologies, new ways and modes of HCI in AVs appeared in autonomous driving technology. Thus, MMI has become one of the most popular topics in modern HCI studies. MMI is a way of HCI which through speech, touch, smell, vision, gesture, somatosensory and other senses and it is widely agreed that MMI is more efficient.

The multi-resource theoretical model provides the theoretical basis for MMI. The theory is a method to explain the load problem from the perspective of resource capacity. It believes that operators have a group of psychological resources with limited capacity, similar nature and similar functions, which are the basis of various business activities [19]. It also points out that the load of information acquisition based on multimodal is lower than that of single modal. Allport *et al.* put forward the multimodal hypothesis and proved through experiments that the channels of perception occupy different mental resources [20]. The hypothesis provides the theoretical basis for MMI. Mayer proposed a model of human information processing system with two audio-visual channels [21]. As Christian and Friedland claim [22], there is no single best modality for the diversity of situations that are encountered while driving. Meanwhile, Sharon *et al.* developed a survey about multimodal interaction and discovered that users prefer multimodal over unimodal interaction when confronted with complex situations [23]. In a multimodal system, two or more of the modalities are combined, such as speech and gesture [24], speech and touch [25]. Multimodal technologies offer great potential to reduce short-comings of single modalities for interaction. At present, the existing MMI literatures focus on visual and auditory channels.

In the AVs' context, many scholars discussed how to improve the trust of AVs from single channel rather multimodal, such as the creation of the environmental light and interactive interface graphic design. For example, Alexander *et al.* make use of ambient visualizations to help drivers obtain awareness of the driving speed without the need to monitor the speedometer [26]. Robert *et al.* presents a prototype in-car system that allows car features (like turning lights and windows) to be controlled by combinations of speech, gaze, and micro-gestures [24]. Bastian *et al.* develop an interactive system-SpeeT [25], which combining touch gestures with speech in automotive environments to exploit the specific advantages of each modality. By comparison, MMI can provide a more efficient way. Driver and passenger can give instructions to vehicles through speech, gesture and other ways. Meanwhile, the vehicle can accurately judge the user's intention through a variety of information, so as to make human-car interaction more natural and easy.

D. SBMI

Information in the real world is composed of a variety of modality inputs such as visual, audio, and haptic information. Sensory information processing is inherently multimodal. In normal situations, an organism perceives the environment using all its senses simultaneously. Humans are known to integrate multimodal information. For example, humans can recognize speech effectively and accurately by using not only audio signals but also visual information (lip motion, eye motion, and gesture).

Synesthesia phenomenon is reported in psychology and neuroscience field as a multimodal perception of humans. It represents a special phenomenon that a stimulation of one sensory pathway automatically leads to experiences in other sensory pathway.

In a multimodal system, two or more of the modalities are combined [27]. Especially the combination of speech with other modalities has been focus of recent research [28]. A study about multimodal interaction discovered that users prefer multimodal over unimodal interaction when confronted with complex situations [29].

The multimodal interaction in AVs mainly focuses on vision and hearing. Human visual attention is oriented mainly contains goal-directed stimulus (endogenous) and external stimuli drive (exogenous). Therefore, we put forward a synesthetic-based multimodal interaction model by combing the synesthetic auditory information and visual information and applied it to the MMI of AVs. In the model, auditory information which are synesthetic with visual target information are used as exogenous stimulation-driven information, so as to guide visual attention and achieve the guidance of ''visual priority selection''.

As shown in Figure 1, firstly, non-synesthetic audio-visual stimuli are processed by visual perception and auditory perception respectively. Secondly, the information content will be transmitted to the visual and auditory processing areas of the brain, and finally the corresponding response signals and actions will be output after processing by the brain. As can be seen from the figure, non-synesthetic audio-visual information is transmitted in a single process, and each kind of information is perceived and processed in a single way, and there is no cross-modal interaction between the information.

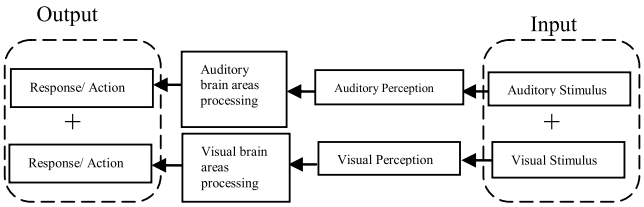


FIGURE 1. Non-synesthetic audio-visual information processing.

In the SBMI model, synesthesia fulfills its role as a high-level post-perception phenomenon. We use auditory and visual stimulus with synesthetic effects to stimulate people simultaneously in the model as shown in Figure 2. Audio-visual stimulus signals with synesthetic effect will produce cross-modal connection and combination with unconscious cognitive processes when they are perceived by people, so as to process the integration of auditory and visual information in high level heteromorphic cortical regions. Due to synesthesia can fully stimulate hyper connectivity between cortical areas potentially related to their synesthetic associations. Such hyper connectivity might not only give rise to a particular form of synesthesia but also result in enhanced performance for any task, or class of stimuli, relying on the same pathways.

The information transmission and processing in the SBMI model can trigger the unconscious perceptual connection between the audio and visual information processing modules in the human brain and exert an unconscious fusion effect. SBMI model is more effective than single information stimulus or non-synesthetic audio-visual information stimulus not only in terms of information transmission efficiency and effect, but also in terms of output response/ action.

E. HYPOTHESIS

Most of the autonomous driving technologies are at L2-L3 level currently, which does not reach the level of complete automatic driving. The current autonomous driving system is more inclined to assist the automatic driving function, and users are generally required to take over the driving behavior functions while a special situation occurs, so as to improve the safety and reliability of the driving system. The implementation of takeover behavior is largely depends on users' trust in the machine system, which includes the authenticity, accuracy and reliability of the information conveyed by users to the system and the rationality, accuracy and effectiveness of the system's operation state prediction based on the data calculation results.

Furthermore, understandings of trust in AVs within HCI research had moved beyond questions of usability and perceived use to turn our attention towards how factors external to a momentary HCI impact on what makes people comfortable with a technology [30]. ''Synesthesia'' is one of the external factors which make people comfortable to some extent. Studies have shown that 97% of the information people receive comes from vision and hearing. In the context of AVs, speech-based dialogue, e.g. for programming the navigation system, has become quite popular. Pfleging *et al.* report that the combination of speech with other modalities has been the focus of recent research. Therefore, our study mainly focuses on audio-visual synesthesia [28].

Thus, according to Sun, X.'s experimental results on synesthesia, that is, there is synesthesia effect between high frequency sound and red color as well as low frequency speech and blue color, we propose the hypothesis that the multimodal interaction mode based on synesthesia can effectively improve the reception rate of system warning information in case of sudden or extreme events so as to enhance users' trust in the AVs system.

II. EXPERIMENT

A. PARTICIPANTS

82 participants (47 males, 35 females; M = 31.1 years old, SD = 9.78 were recruited using online advertisements and web posts for this experiment, questionnaires and interviews. We fully consider the age and driving experience of the participants, and all participants had a valid driver's license and at least two years of driving experience. The participants are from different professional backgrounds, including marketing, education, finance, and freelancing to increase the diversity of the participants so as to ensure the experimental conclusions more convincing. The participants were assigned into two groups: Group A (non-synesthesia) and Group B (synesthesia group). There are 25 males and 16 females in Group A and 22 males and 19 females in Group B. Each participant's experiment last for 10 minutes.

B. APPARATUS

The experiment is carried out on a stationary driving simulator, which consists of a steering wheel, pedals, seat and screen. We record the relevant road conditions encountered in the experiment by driving the car in advance. The speed is no more than 60 km/h for the entire drive. During the experiment, the driving scenario is displayed on the screens and participants seated in the driver's seat to watch the screens. They can participate in the automatic driving process by holding the steering wheel and other ways according to the tasks encountered by the simulated driving. The two groups' experiment uses the same scenario, lasting about 10 minutes.

Speech prompt the information of road conditions in the process of automatic driving will be played through the system's own speech system and ambient light will be displayed through LED lights set on the steering wheel, as shown in Figure 3. (a) and (b) show the LED lights displayed on the steering wheel during the experiment of group B, while (c) and (d) show the LED lights displayed on the steering wheel during the experiment of group A. Participants completed a secondary task and used their own mobile phones (to avoid the difference caused by different participants' familiarity with mobile phones). We would inform them of the reply contents by emails in advance and give them prompts for the start of secondary tasks during the experiment.

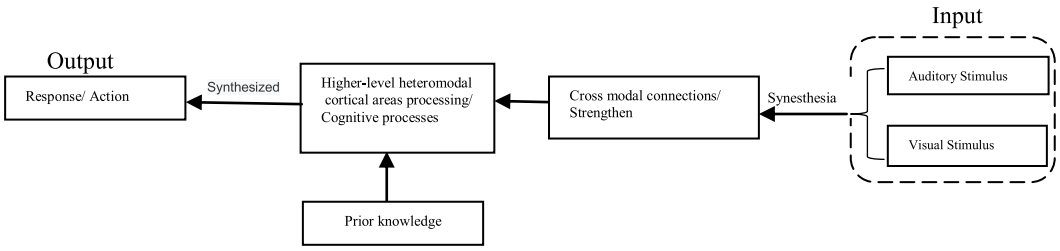


FIGURE 2. Synesthetic audio-visual information processing.

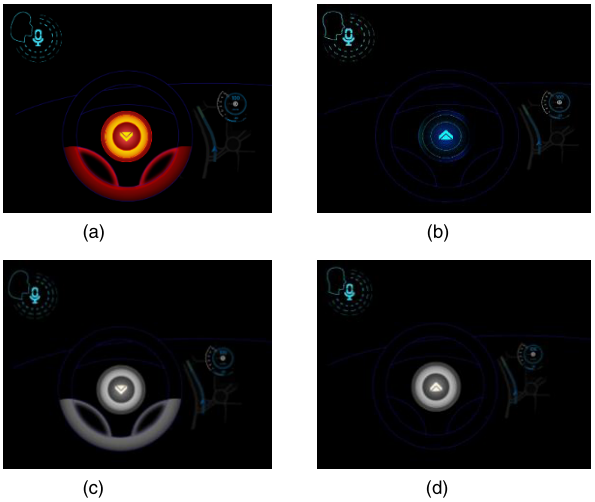


FIGURE 3. The ambient light cues on the steering wheel.

FIGURE 4. The screen shots of driving road conditions. a, b and c represent the three conditions of NTC; d, e and f represent the three conditions of ETC.

The speech prompt information in the experiment was provided by baidu broadcasting open platform. This is a standard text-to-speech software, which can effectively convert text into speech and directly generate speech files by inserting text. Speech output such as male speech or female speech can also be selected. In the experiment, we use 150HZ as low frequency speech (male) and 300HZ as high frequency speech (female).

C. DESIGN

The experiment is carried out by comparing between groups. Participants in Group A and Group B participated in the experiment under the automatic driving scenario respectively. Both of the two groups experienced NTC (turning, traffic lights, over and limit speed prompts) and ETC (sudden braking of the car in front, temporary lane change, pedestrian thrusting) in the automatic driving scenarios, Figure 4 shows screen shots of different driving road conditions

The difference of the experiment between two groups is the reminders way. In the experiment of group A, the traffic condition information was transmitted simultaneously through speech and international standard artificial daylight prompt. In Group A, an artificial standard light source with a color temperature of 6000K±100K (ISO 3664:2000: standard artificial daylight color temperature of 5000K-6500K) is used to simulate the bright blue sky and the average natural sunshine under sunshine. The reason we choose artificial standard light is that it is a neutral white light that we are familiar with and adapted to physically and psychologically. In addition, in the spectrum, artificial standard light is a kind of compound light, which can be dispersed into seven monochromatic lights through the prism effect, namely, red, orange, yellow, green, blue, indigo and purple. Hence, the artificial standard light is more inclined to be a comprehensive neutral light source, which is the collection of all monochromatic light. It has no clear cooling and warming characteristics, and has no strong tendency to induce people's emotions. At the same time, due to the universality and persistence of natural light signal stimulation, the human body's sensory system has no significant response to the stimulation. Therefore, artificial standard light is very suitable as the source of light source for the control group. In contrast, group B is reminded by synesthesia-based speech + ambient light which have synesthetic effect. (see Table 1). Meanwhile, the road condition prompts were also displayed on the display panel of the instrument in the whole process of experiment, so it was considered as a constant rather than a variable. At the same time, both groups A and B are required to complete a secondary task of using mobile phones to reply an email according to the prompts (the reply content is a paragraph of

TABLE 1. The way of road condition reminder during the experiment.

GROUP	Driving conditions	Speech	Ambient light	
A	NTC	Low frequency speech prompt	Standard daylight	artificial
A	ETC	High frequency speech prompt	standard daylight	artificial
B	NTC	Low frequency speech prompt	Blue color	
B	ETC	High frequency speech prompt	Red color	

Chinese characters), which was used to reflect the occupied situation of audiovisual channels of participants. The whole experiment is filmed to measure how often users held their hands on the steering wheel and looked up at the road.

D. PROCEDURE

Upon arrival, participants are required to read and sign an informed consent form and read the instructions for the experiment. Then they complete four questionnaires measuring their attitude to trust, technical competence, situation management, perceived ease of use. This will provide a baseline of their attitude to AVs. Then they seat before laboratory simulation driving device, began to be familiar with the experimental environment and simulator, etc.. The next stage is they began 5 minutes of driving simulation exercises, which could let them be familiar with the experiment simulator operation. Meanwhile it also stimulates the sense of reality, because if they start experimenting directly, they may feel they're looking at a screen, which will cause the lack of desire to participate in driving.

The driving scenarios in the experiment are divided into NTC and ETC. In NTC section, the AVs started on a straight stretch of a four-lane road, and stopped at a traffic light. Then the participants receive a prompt message to perform the secondary task of responding to the email. When the traffic light turns green, the car completes the task of turning left (Group A: male speech prompt 'Turn left ahead' + steering wheel standard artificial daylight prompt; Group B: male speech prompt 'Turn left ahead' + steering wheel blue light prompt). After going straight for a while, the AVs braking and slowing down at a traffic light just when it turned yellow(Group A: male speech prompt 'Please slow down at the junction ahead' + steering wheel standard artificial daylight prompt; Group B: male speech prompt 'Please slow down at the junction ahead' + steering wheel blue light prompt). Then, the vehicle goes straight after the traffic light turns green until it come up a 60km speed limit sign, and it slows down to 60km/h. (Group A: male speech prompt 'The speed limit ahead is 60km/h, please obey the traffic rules' + steering wheel standard artificial daylight prompt; Group B: male speech prompt 'The speed limit ahead is 60km/h, please obey the traffic rules' + steering wheel blue light prompt). During the process, participants are allowed to look up at the road or hold the steering wheel, or continue the secondary task of responding to an email without doing anything.

Similarly, during ETC section, AVs encounter sudden braking of the car, temporary lane change and pedestrian thrusting respectively. When the AVs encounter these three situations, it will have different prompts and methods. When it comes to the sudden braking of the car, the prompts are (Group A: female speech prompt 'The car in front stops abruptly, please slow down' + steering wheel standard artificial daylight prompt; Group B: female speech prompt 'The car in front stops abruptly, please slow down' + steering wheel red light prompt). When it comes to the temporary lane change, the prompts are (Group A: female speech prompt 'The car in front stops abruptly, please slow down' + steering wheel standard artificial daylight prompt; Group B: female speech prompt 'The car in front stops abruptly, please slow down' + steering wheel red light prompt). When it comes to the pedestrian thrusting, the prompts are(Group A: female speech prompt 'Please pay attention to the pedestrians' + steering wheel standard artificial daylight prompt; Group B: female speech prompt 'Please pay attention to the pedestrians' + steering wheel red light prompt). The same as NTC section, participants are allowed to look up at the road or hold the steering wheel, or continue the secondary task of responding to an email without doing anything during the ETC process.

Group A and Group B conducted experiments separately. After completing the experiment, all participants fill out a Likert scale questionnaire on trust of AVs in a randomized order to counteract sequential effects. Subsequently, participants finish a qualitative interview with four questions. The experimental procedure is depicted in Figure 5.

FIGURE 5. Flowchart of experimental procedure.

E. DATA COLLECTION

1) QUANTITATIVE

By collecting the data of participants' instinctive reaction in the process of experiment (holding the steering wheel, pay attention to the road), and the choices of different road conditions (normal and emergency), we analyze the selection differences between Group A and Group B, and verified whether the synesthesia effect had an impact on driver selection. After the experiment, the participants fill out questionnaires about their trust in the AVs. The questionnaire content includes four aspects: trust, technical competence, situation management and perceived ease of use.

① Trust

Three questions about trust are selected from Jian, Bisantz, and Drury's AVs trust 7-point Likert Scale (Cronbach's α = .798) [31] AVs is dependable; AVs is reliable; Overall, I can trust AVs.

② Technical competence

Technical competence is measured via a subcomponent of the Choi and Ji [32], 'Trust on adopting an AVs' questionnaire. It comprises three questions such as "I believe that AVs is free of error; I believe that I can depend and rely on AVs; I believe that AVs will consistently perform under a variety of circumstances. Participants rank their choices on a 7-point scale from 1 (strongly disagree) to 7 (strongly agree)

③ Situation management

Situation management, another subcomponent of the Choi and Ji questionnaire [32], comprised three 7-point scale questions. Those are, for example, "I believe that AVs provides alternative solutions; I believe that I can control the behavior of AVs; I believe that AVs will provide adequate, effective, and responsive help".

④ Perceived ease of use

Perceived ease of use, also an subcomponent of the Choi and Ji questionnaire [32], comprised of three 7-point scale questions, include "Learning to operate AVs would be easy for me;Interacting with AVs would not require a lot of my mental effort; AVs interactions can provide good feedback.".

2) QUALITATIVE

After experiment, qualitative opinions are also collected from the participants through interviews. We ask the participants some questions about the usefulness and convenience of the interactive information provided by the system during the experiment, as well as their attitudes and opinions on the way to match the speech prompts and ambient light prompts of the automatic driving system. It includes four questions: 1) Did the information provided by the AVs effectively remind you to withdraw from other tasks and participate in the driving task? 2) Were high-frequency or low-frequency speech more effective in prompting? 3) Would the combination of speech and ambient light have any influence on your choice whether to participate in automatic driving? 4) Which combination of speech and ambient light did you think is more suggestive?

F. RESULTS

1) QUANTITATIVE RESULTS

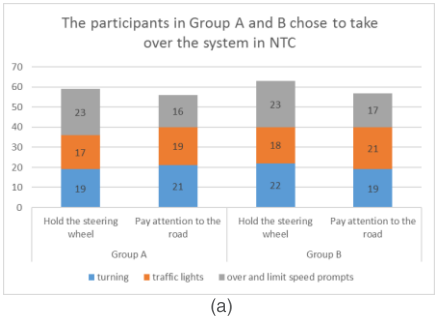
The results of the quantitative data include two parts. The first part is the results of the relevant reaction data of the experimental participants during the experimental driving. The second part is the data results of the questionnaire and qualitative interview after the experimental driving.

① Experimental selection data results

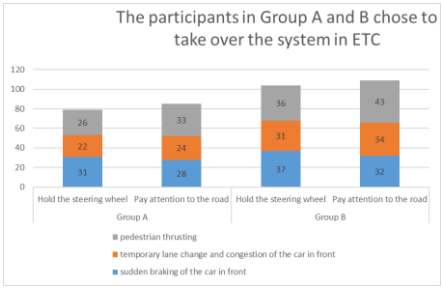
From the video recorded by the experiment, we count the performance of the instinctive reaction of the participants in Group A and Group B when different road conditions occurred during the experiment. The statistical results show that the frequency of the participants in Group A and B holding the steering wheel or pay attention to the road during the emergency (183 times holding the steering wheel; pay attention to the road: 194 times) significantly higher than normal road conditions (steering wheel: 122 times; Pay attention to the road: 113), see Table 2.

TABLE 2. The frequency of instinctive response of participants in different road conditions during driving.

GROUP	NTC	NTC	ETC	ETC
	Hold the steering wheel	Pay attention to the road	Hold the steering wheel	Pay attention to the road
A	59	56	79	85
B	63	57	104	109



(a)



(b)

FIGURE 6. (a) The participants in Group A and B chose to take over the system in NTC. (b) The participants in Group A and B chose to take over the system in ETC.

When NTC occurs, participants in Group A and Group B have similar frequency in choosing whether to hold the steering wheel or pay attention to the road, and there is no significant difference. In contrast, when ETC occurs, the frequency of Group B choosing to hold the steering wheel or pay attention to the road surface is significantly higher than that of Group A, and there is a significant difference between the two groups in choice.

Meanwhile, we also count the instinctive response of the participants in the normal and emergency road conditions as shown in Figure 6. It can be seen that there is no significant difference in the selection when participants in NTC. However, in case of ETC, participants in Group A and Group B choose to take over (hold the steering wheel or pay attention to the road) more frequently than the other two road conditions when confront with a pedestrian's sudden warning. Therefore, we speculate that information reminders related to passers-by during driving could attract more attention.

② Questionnaire results

Independent samples analyses of variance are conducted with experimental condition (non-synesthesia vs. synesthesia) as the independent variable and trust, technical

TABLE 3. Results of one-way ANOVA showing the effects.

Measured variables	*M* (Group A)	*M* (Group B)	*F*	*p*
Trust	3.35	4.37	.231	**.001**
Technical competence	4.08	5.13	8.786	**.004**
Situation management	4.27	5.08	5.790	**.018**
Perceived ease of use	3.42	3.83	.743	**.041**

Note: *M* Means, *F* values, *p* values effect sizes. Significant effects are shown in bold

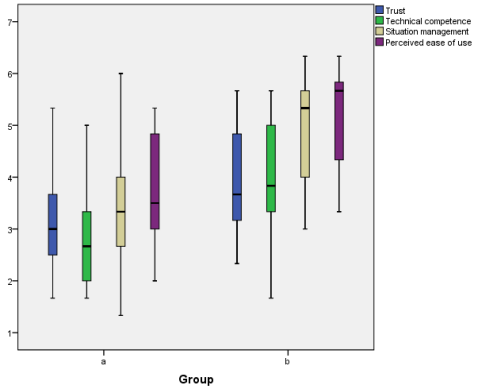


FIGURE 7. The distribution of questionnaire scores of participants in groups A and B.

competence, situation management and perceived ease of use were taken as dependent variables respectively.

The result shows that SBMI model would give people more confidence in AVS system (see Table 3). All of these four parameters in Group B are higher than that in Group A. The score distribution has been shown in Figure 7 which content mainly includes trust $F(1, 80) = 0.231, p = .001$, technical competence, $F(1, 80) = 8.786, p = .004$, situation management $F(1, 80) = 5.790, p = .018$, perceived ease of use $F(1, 80) = .743, p = .041$.

Furthermore, person's correlations were conducted between trust and technical competence, situation management, and perceived ease of use. Technical competence ($r = .633, p < .001$), situation management ($r = .552, p < .000$) and perceived ease of use ($r = .668, p < .001$) were positively correlated to trust. This proves that all three factors affect people's trust in AVs.

Table 4 shows that statistically significant differences were found between pretest and posttest (group A) for situation management. The post-test questionnaire score of situation management increased by 1.293 ($p = .012$) than pre-test. Meanwhile, it also shows that statistically significant differences were found between pretest and posttest (group B) for trust, technical competence and situation management. The post-test questionnaire scores of trust, technical competence and situation management improved by 1.715($p = .002$), 2.455($p = .001$) and 2.113($p = .025$), respectively.

TABLE 4. Paired-samples t-tests showing the pairs of conditions and the resulting t and p (2-tailed), df(40). Significant values are highlighted with * (p <.05) or ** (p < .01).

	posttest -pretest (group A)		posttest -pretest (group B)	
Trust	1.528	.354	1.715	.002**
Technical competence	.967	.203	2.455	.001**
Situation management	1.293	.012*	2.113	.025*
Perceived ease of use	.984	.149	1.114	.063

2) QUALITATIVE RESULTS

Subjective opinions were collected through interviews after the experiment. Four questions were asked to all participants: 1) Did the information provided by the AVs effectively remind you to withdraw from other tasks and participate in the driving task? 2) Were high-frequency or low-frequency speech more effective in prompting? 3) Would the combination of speech and ambient light have any influence on your choice whether to participate in automatic driving?4) Which combination of speech and ambient light did you think is more suggestive? The answers to the questions were grouped in terms of the possible answers (see Table 5).

The validity of AVs system alerts mentioned in question 1, nearly half of the participants think it is valid, while the other think it is invalid. There is no difference in the choice of the answer between group A and Group B. For the question of whether high-frequency or low-frequency speech was more effective as mentioned in Question 2, the majority of participants think that high-frequency speech is more effective, and 53 of the 82 participants choose high-frequency speech. Thus we can see that high-frequency speech prompts can be more effectively recognized by people in AVS. The Questions 3 and 4 covered the matching of speech and ambient light. ''Whether the combination of speech and ambient light influences driving participation '' mentioned in question 3, most of participants in groups A and B chose the answer'' Yes, it has influence.'' (A group: 28; B group: 33)They believe that the combination of speech and ambient light have obvious influence on their AVS driving. ''Which speech and color combination was more effective'' mentioned in question 4, 10 participants choose high-frequency speech + standard artificial daylight, 12 participants choose low -frequency speech + standard artificial daylight, and 19 choose no significant difference in group A. The results indicate that there is no significant difference in people's choice of non-synesthetic speech and color combination.

By comparison, participants in group B have different choice. 32 participants choose high-frequency speech + red ambient light, far more than the participants who choose low -frequency speech + blue ambient light. The results of question 4 reveal participants in group B have the highest recognition degree for high-frequency speech + red ambient

TABLE 5. Results of qualitative interview.

Question	Answer	Group A	Group B
Did the information provided by the AVs effectively remind you to withdraw from other tasks and participate in the driving task?	Yes, it was helpful.	17	20
	Yes, but reminders didn't work effectively	21	18
	No, the reminders were distractions in AVs	3	4
Were high-frequency or low-frequency speech more effective in prompting?	High-frequency speech was more effective in prompting	25	28
	Low-frequency speech was more effective in prompting	7	5
	The frequency does not affect the effectiveness of the prompt	9	8
Would the combination of speech and ambient light have any influence on your choice whether to participate in automatic driving?	Yes, it has influence.	28	33
	No, it hasn't an influence.	5	6
	It doesn't matter	8	3
Which combination of speech and ambient light did you think is more suggestive?	High-frequency speech + red ambient light	X	32
	Low-frequency speech + blue ambient light	X	6
	High-frequency speech + standard artificial daylight	10	X
	Low-frequency speech + standard artificial daylight	12	X
	It doesn't matter	19	3

light, which prove the role of synesthesia effect in enhancing participants' trust in AVs to some extent.

G. DISCUSSION

In this study, we investigate whether SBMIs can improve people's trust in AVs. We designed two groups (A and B) to participate in AVs driving experiments. During the experiment, participants experience both NTC and ETC, so as to test their instinctive reactions during the experiment. Furthermore, participants also complete questionnaires and qualitative interviews about trust, technical competence, situation management and perceived ease of use. Several statistical analysis methods were applied to explore the inner link between SBMI and AVs trust. The results reveal that SBMI can improve people's trust in AVs driving to some extent. The findings may be discussed in two parts.

1) SYNESTHESIA CAN IMPROVE THE EFFECT OF EMERGENCY ROAD CONDITION WARNING DURING AVS DRIVING

During the experiment, participants in group B choose to drive more often, 377 times (hold the steering wheel or pay attention to the road) while participants in Group A is only 235 times (hold the steering wheel or pay attention to the road). The results indicate that participants were more trust in AVs to handle normal road conditions. However, it would be also because participants believe that NTC are not dangerous psychologically. Chi-square test show that Group A and Group B have a great difference in choice when emergency road conditions occur, $x^2 = 6.354, p = .009$, which reveal that SBMI can effectively alert people to the occurrence of emergency road conditions during automatic driving.

In the process of AVs driving, the main driving task of human drivers is to shift from manual control to automatic supervision and control. However, the driver needs to be able to achieve the driving takeover timely and accurately when an emergency occurs, so as to avoid the occurrence of danger. Therefore, it is particularly important for drivers to get the prompt information effectively and achieve the driving takeover timely in the case of ETC.

Experiments show that SBMI prompts can more effectively prompt the driver to participate in the driving takeover (377 times). These findings are in line with the work presented by Politis *et al.* that voice commands in combination with other cues led to better driving performance after handover than voice commands alone [33].

Our results also demonstrate that the use of ambient light has a certain promoting effect on people's acceptance of the information transmitted by AVs. Standard artificial daylight is used in the comparative group, while red and blue ambient light are used in the synesthesia group. The results show that different combinations of ambient light and sound have different influences on drivers. Our finding is consisted with Hanneke Hooft *et al.*'s finding where they showed that the ambient light has a positive impact on the driving experience, but attitudes towards ambient light are highly personalized [34]. We use red and blue ambient light in the experiment, which indicates that red ambient light is more likely to prompt the driver to take over in ETC. Meanwhile, the results indicate that there is no significant difference between standard artificial daylight and red ambient light in the early period to warning participants, but the warning effect became worse and worse in the later period. This may be because people have adapted to standard artificial daylight.

In addition, the results of the qualitative questionnaire also verified this point of view. Participants in group B have the highest recognition of the alert effect of high-frequency speech + red ambient light, which indicate that people trust the judgment results of the automatic driving system, so they choose to participate in the automatic driving takeover when an emergency occurs.

2) SBMI CONTRIBUTES TO THE OVERALL TRUST IMPROVEMENT OF TH AVS

Multimodal interaction, especially the combination of audio-visual, plays an important role in AVs trust. Liu *et al.* conducted relevant experiment on navigation multimodal operation and concluded that multimodal interactive mode was more efficient than single visual mode in completing driving tasks [35]. We also adopt audio-visual multimodal interaction related traffic information to the participants in the experiment, and most participants indicate that the speech + ambient light prompts have a significant impact on their choice of AVs. Especially, participants in Group B comment that they find high frequency speech +red color prompts based on synesthetic is more effective in alerting them to take over the AVs in ETC. Therefore, synesthesia can be fully considered in the design of AVs system.

The questionnaire used in this study includes four dimensions: trust, technical competence, situation management and perceived ease of use in AVs. The four dimensions cover people's cognition and feelings of AVs. The quantitative data gathered from the questionnaire in this study show that, participants in Group B had higher experience and scores in three dimensions (technical competence, situation management, and perceived ease of use) than group A. For example, the average score of trust in AVs chosen by participants in group A is 3.35 while 4.37 in group B. Meanwhile, in terms of technical competence, situation management and perceived ease of use, the average score of participants in group A are 4.08, 4.27, and 3.42, while participants in group B are 5.13, 5.08, and 3.83. In conclusion, participants in the synesthesia group are more likely to identify with several dimensions related to the trust of AVs than those in the non-synesthesia group, which indicate that the synesthesia group was significantly better than the non-synesthesia group in improving people's trust in AVs. Meanwhile, it's also verified that there was a positive correlation between AVs trust and the three dimensions: technical competence, situation management, perceived ease of use. The correlation coefficient is respectively 0.633, 0.552 and 0.668.

We also use a combination of pre-test and post-test to compare the changes in the attitudes of participants in group A and group B towards AVs between before and after the experiment. Paired samples t-tests show that the post-test scores of trust, technical competence and situation management of AVs in group B which using the SBMI model are significantly improved with statistical significance compared with pre-test. The scores of trust, technical competence and situation management improved by 1.715($p = .002$), 2.455($p = .001$) and 2.113($p = .025$), respectively. In general, group B who use the SBMI model for the experiment have greater improvement with the confidence in the trust, technical ability, situation management and other elements of AVs after the experiment. This result also proves the SBMI model used in Group B had a significantly influence on participants' driving behavior. It enhances the confidence of participants to the AVs system, so as to improve the level of trust.

Similar to Choi *et al.*'s studies that in the case of trust construct, 47.4% of variance was explained by system transparency, technical competence, and situation management [32] and our study verified the positive correlation between technical competence, situation management and trust. However, the results on the correlation between perceived ease of use and trust are not consistent with Choi *et al.*' study. Our results show that there is a strong positive correlation between them, with a correlation coefficient of 0.668, indicating that ease of perception is an important factor affecting people's trust in AVs, while their study showed that perceived ease of use had only a slight effect on behavioral intention (whether to drive on AVs or not) [32]. This may be due to individual differences in participants. The participants in their experiments are already familiar with driving a vehicle, so it is not that hard to use autonomous vehicles.

Overall, these three factors have a positive effect on improving people's trust in AVs. Consequently, improvement of people's trust in AVs should be made from the aspects of technical competence, situation management and perceived ease of use.

III. RESEARCH LIMITATIONS

We acknowledge that the research presented a number of limitations which could have had implications for the findings, and they should be further deepened or addressed in future research. The study is a multimodal interaction based on speech-color synesthesia, mainly focusing on human vision and hearing. However, as a complete spatial system, AVs should meet the need of various sensory interactions; hence we may consider adding touch, taste and other synesthesia in the subsequent studies. Another limitation of the study is that it was conducted in a driving simulator, where there was no risk of harm. The indoor environment containing controlled safety features could have given a sense of security to our participants, which could affect the attitude judgment to the AVs. The following study should consider carrying out outdoor real driving. Finally, participants complete a secondary task of using mobile phones to reply an email according to the prompts (the reply content is a paragraph of Chinese characters), which is relatively simple in form. The purposes of AVs are to free people from driving and enable them to do other things. Subsequent studies should focus on the participants engaging in other tasks, such as watching entertainment programs and playing games during the process of autonomous driving. In such conditions, it needs to be investigated whether the current experimental results are still valid.

IV. CONCLUSION

The results of the study verified that SBMI can improve people's trust in AVs to a certain extent. Both quantitative data and qualitative feedback from studies have proved that

SBMI can more effectively remind people to participate in automatic driving in case of emergency, thus improving people's trust in the judgment of AVs. Quantitative data gathered from the questionnaires also verified that participants are more likely to identify with the four dimensions of trust, technical competence, situation management, perceived ease of use through SBMI. Multimodal technologies offer a great potential to reduce shortcomings of single modalities for interaction. Although quite some research on multimodality has been conducted and some general guidelines have been shaped no specific patterns or interaction styles for an appropriate integration of different modalities have emerged yet. We introduce synesthesia, especially the theory of audio-visual synesthesia, into the MMI design of AVs. Then, we design experiments to verify that SBMI can improve the trust of AVs, which provides a new idea for the interface design of AVs in the future. Further research will focus on exploring more synesthesia effect in the application of Multimodal human-computer interaction in AVs, improving the effect of human-computer interaction, and thus improving people's trust in AVs.

REFERENCES

[1] P. Wintersberger and A. Riener, "Trust in technology as a safety aspect in highly automated driving," *I-Com*, vol. 15, no. 3, pp. 297–310, Jan. 2016.

[2] M. Ghazizadeh, J. D. Lee, and L. N. Boyle, "Extending the technology acceptance model to assess automation," *Cognition, Technol. Work*, vol. 14, no. 1, pp. 39–49, Mar. 2012.

[3] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, p. 50, 2004.

[4] J. K. Rempel, J. G. Holmes, and M. D. Zanna, "Trust in close relationships," *J. Personality Social Psychol.*, vol. 49, p. 95, Jul. 1985.

[5] J. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, Oct. 1992.

[6] F. Ekman, M. Johansson, and J. Sochor, "Creating appropriate trust in automated vehicle systems: A framework for HMI design," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 1, pp. 95–101, Feb. 2018.

[7] R. Parasuraman and C. A. Miller, "Trust and etiquette in high-criticality automated systems," *Commun. ACM*, vol. 47, no. 4, pp. 51–55, Apr. 2004.

[8] P. Wintersberger, T. von Sawitzky, A.-K. Frison, and A. Riener, "Traffic augmentation as a means to increase trust in automated driving systems," in *Proc. 12th Biannual Conf. Italian SIGCHI Chapter*, Sep. 2017, pp. 1–7.

[9] J. Koo, J. Kwac, W. Ju, M. Steinert, L. Leifer, and C. Nass, "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance," *Int. J. Interact. Des. Manuf.*, vol. 9, no. 4, pp. 269–275, Nov. 2015.

[10] J. M. Kraus, J. Sturn, J. E. Reiser, and M. Baumann, "Anthropomorphic agents, transparent automation and driver personality: Towards an integrative multi-level model of determinants for effective driver-vehicle cooperation in highly automated vehicles," in *Proc. 7th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Sep. 2015, pp. 8–13.

[11] F. Ekman, M. Johansson, and J. Sochor, "Creating appropriate trust for autonomous vehicle systems: A framework for HMI design," in *Proc. 95th Annu. Meeting Transp. Res. Board*, 2016, pp. 3216–3268.

[12] M. R. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proc. IEEE Nat. Aerosp. Electron. Conf.*, May 1988, pp. 789–795.

[13] M. Beggiato and J. F. Krems, "The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 18, pp. 47–57, May 2013.

[14] O. L. Caivano, "Color and Sound: Physical and Psychophysical Relations," *Color Res. Appl.*, vol. 19, no. 2, pp. 125–133, 1994.

[15] H. Hagtvedt and S. A. Brasel, "Cross-modal communication: Sound frequency influences consumer responses to color lightness," *J. Marketing Res.*, vol. 53, no. 4, pp. 551–562, Aug. 2016.

[16] I. H. Bernstein, T. R. Eason, and D. L. Schurman, "Hue-tone sensory interaction: A negative result," *Perceptual Motor Skills*, vol. 33, no. 3, pp. 1327–1330, 1971.

[17] L. E. Marks, "On colored-hearing synesthesia: Cross-modal translations of sensory dimensions," *Psychol. Bull.*, vol. 82, pp. 303–331, Dec. 1975.

[18] K. Evans and A. Treisman, "Natural cross-modal mappings between visual and auditory features," *J. Vis.*, vol. 10, no. 1, p. 61, 2010.

[19] T. L. Hubbard, "Synesthesia-like mappings of lightness, pitch, and melodic interval," *Amer. J. Psychol.*, vol. 109, no. 2, p. 219, 1996.

[20] W. Jie, F. Weining, and L. Guangyan, "Mental workload evaluation method based on multi-resource theory model," *J. Beijing Jiaotong Univ.*, vol. 36, no. 4, pp. 108–112, 2010.

[21] R. E. Mayer, *Multimedia Learning*. Beijing, China: The commercial Press, 2006.

[22] C. Müller and G. Friedland, "Multimodal interfaces for automotive applications (MIAA)," in *Proc. 14th Int. Conf. Intell. user Interfaces*, Feb. 2009, pp. 493–494.

[23] S. Gaudin. (2012). *Autonomous Cars Will Arrive Within 10 Years*. [Online]. [Online]. Available: https://www.computerworld.com/article/2492744/emerging-technology/autonomous-cars-will-arrive-within-10-years–intel-cto-says.html

[24] L. P. Robert, "Monitoring and trust in virtual teams," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 245–259.

[25] B. Pfleging, M. Kienast, A. Schmidt, and T. Dring, "SpeeT: A multimodal interaction style combining speech and touch interaction in automotive environments," *AutomotiveUI*, vol. 11, p. 15, Nov./Dec. 2011.

[26] A. Meschtscherjakov, C. Döttlinger, and C. Rödel, "ChaseLight: Ambient LED stripes to control driving speed," in *Proc. 7th Int. Conf. Automot. Interfaces Interact. Veh. Appl.*, 2015, pp. 212–219.

[27] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people and PLA," Bibliovault OAI Repository, Univ. Chicago Press, Chicago, IL, USA, Tech. Rep., 1996, p. 128.

[28] B. Pfleging, S. Schneegass, and A. Schmidt, "Multimodal interaction in the car: Combining speech and gestures on the steering wheel," in *Proc. 4th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, 2012, pp. 155–162.

[29] S. Oviatt, R. Coulston, and R. Lunsford, "When do we interact multimodally? Cognitive load and multimodal communication patterns," in *Proc. 6th Int. Conf. Multimodal Interfaces*, Oct. 2004, pp. 129–136.

[30] M. Hassenzahl, *Experience Design: Technology for All the Right Reasons, Synthesis Lectures on Human-Centered Informatics*. San Rafael, CA, USA: Morgan & Claypool, 2010.

[31] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cognit. Ergonom.*, vol. 4, no. 1, pp. 53–71, Mar. 2000.

[32] J. K. Choi and Y. G. Ji, "Investigating the importance of trust on adopting an autonomous vehicle," *Int. J. Hum.-Comput. Interact.*, vol. 31, no. 10, pp. 692–702, Oct. 2015.

[33] I. Politis, S. Brewster, and F. Pollick, "Language-based multimodal displays for the handover of control in autonomous cars," in *Proc. 7th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, 2015, pp. 3–10.

[34] H. H. van Huysduynen, J. Terken, A. Meschtscherjakov, B. Eggen, and M. Tscheligi, "Ambient light and its influence on driving experience," in *Proc. 9th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Sep. 2017, pp. 293–301.

[35] Y.-C. Liu, "Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems," *Ergonomics*, vol. 44, no. 4, pp. 425–442, Mar. 2001.

XIAOFENG SUN received the M.S. degree from the School of Mechanical Engineering and Automation, Northeastern University, in 2007, where she is currently pursuing the Ph.D. degree. Her research interests include cognitive psychology, human–computer interaction, and user experience design.

YIMIN ZHANG received the M.S. and Ph.D. degrees in mechanical reliability from Jilin University, Changchun, China, in 1989 and 1995, respectively. He is currently a Full Professor with the Equipment Reliability Institute, Shenyang University of Chemical Technology, Shenyang, China. He has authored and coauthored more than 200 articles. His main research interests include mechanical reliability, mechanical dynamics, and optimization design.