The Impact of In-Vehicle Voice Interaction System on Driving Safety

Jun Ma, Yining Ding*
School of Automotive Studies, Tongji University, Shanghai, China

*Corresponding author: 1833450@tongji.edu.cn

Abstract. The development of intelligent connected vehicle (ICV) makes voice interaction an important interaction method in the car. The voice interaction system supports natural speech recognition and can complete a large number of secondary tasks. This paper is based on the Chinese market and 5 ICVs are screened out. Application scenarios for voice interaction systems are constructed through presurvey. Using simulated driving platform which can be connected with mass-produced vehicle, data on visual demand can be obtained. From the visual demand aspects, the corresponding indicators are selected to analyze the impact of in-vehicle voice interaction system on driving safety.

1. Introduction
In the car, the In-Vehicle Information System (IVIS) plays an important role. When driving, drivers often need to listen to the audio, answer the phone, find the point of interest (POI) and etc., which will make the driver's hand away from the steering wheel, eyes away from the road and can't focus on the driving. Among the road traffic accidents caused by drivers in 2016, 54.18% were caused by perception errors, 35.88% by judgment and decision-making errors, and 9.15% by operation and decision-making errors [1].

Voice interaction is being applied to cockpit more and more because of its high degree of naturalness and rich realizable functions. In the US, 114 million adults use voice assistants in their cars, almost twice as many as smart speakers (57.8 million). With more and more users, we need to study the impact of intelligent voice interaction systems on driving safety from the perspective of users.

2. Methods
Considering the complexity of the in-vehicle voice interaction system and environment, the driving simulation test was finally selected. It can ensure the safety, restore the driving scenario and provide accurate data. The experimental platform can be divided into three modules: driving platform, environment and eye tracker. The simulation environment is designed based on the Unity 3D platform. The definition and design of all roads are based on real Chinese road scenes. In this paper, the experiment chose a 13.8km route, including urban roads, elevated roads and suburban roads.

80% information a driver gets while driving comes from vision [2]. It is a common evaluation method to collect eye movement signals by eye tracker.

There are 3 types of eye movements when distracted driving occurs: fixations, pursuit and saccades. Existing studies have shown that the number and duration of gaze are the most direct indicators that visual distraction affects driver's main task. Previous studies have found that if the time interval of ea
ch task is 3s, the driver will feel more secure when the fixation time is less than 0.8s. However, once a certain driving task requires more than 5 fixations, the driver will be forced to compress the fixation in terval, which will have a greater impact on safety [3].

When the drivers use the voice interaction system to complete the secondary task, they need to review the front interior area to confirm whether the task is correctly performed. The time of sight in different areas reflects the distribution of the driver's vision. In the simulated driving environment, there is no need to obtain the environmental information through the rearview mirror and the car window, so in this paper, the windshield (Area 1) and the front interior (Area 2) are designated as areas of interest.

3. Pre-survey and experiment

Currently, there have been many mass-produced ICVs equipped with voice interaction systems. Michael Braun [4] studied that the different visual schemes that affect the safety performance. Finally, 5 ICVs with typical voice interaction system were selected, which have network and natural voice interaction ability, as shown in the Fig.1 below.

Fig 1. Interior of 5 tested cars: from left to right: RR Model, GB Model, LK model, WE Model and DA Model.

Before the experiment, a pre-survey was conducted for the owners of 5 models, including quantitati ve questionnaires through the owner forum, targeted connection and etc., and qualitative interviews wi th owners of 5 models. In the pre-survey questionnaire, the common driving secondary tasks are listed and the owner will choice the secondary tasks finished by the voice interaction system usually. At the s ame time, the owner will be asked in the interview if there are common tasks to add. Finally, through t he pre-survey, the following Table.1 test tasks are determined:

Table 1. Selected task.

| Function     | Secondary task                                      |
|--------------|-----------------------------------------------------|
| Air conditioner | Turn air conditioning on / off                      |
|              | Turn temperature up / down                         |
|              | Increase / decrease air volume                      |
| Radio/music  | Previous / Next                                     |
|              | Switch music mode (radio, music, Bluetooth music, etc.) |
|              | Search music                                       |
| Navigation   | Navigate to "Tongji Technology Plaza"               |
|              | Navigate to the POI "Nearest Gas Station"           |
| Telephone    | Call "Mr. Wu"                                      |
|              | Call "10086"                                        |

When designing the testing scenario, the user's usage habits, the operability of the experiment, and t he scientific result were considered. In the final voice interaction scenario in this paper, the steering wh eel buttons were used as the wakeup method during the task execution. And the drivers can complete t he relevant task by their own expression habits.

After the tested driver arrives in the laboratory, the experimenter first introduces the principle and b asic process. The formal experiment is divided into 5 stages: adaptive practice, interactive practice, no
n-task driving, tasked driving, and subjective evaluation. Throughout the experiment, the experimenter is located in the co-pilot position, and the camera assumed to be in the sunroof is used to record the entire course and recording of the experiment. A total of 12 people, 60 people participated in the experimental process.

4. Results and analysis
The data of the simulated driving platform is collected by eye tracker and matched with the task-record, including the real-time sight deviation duration (SDD), revisit number (RN), fixation count (FC). In this paper, the time segment, from the time when the driver receives the task instruction to the time when the task is completed, is used as the cut-off point to divide the data segment.

This paper adopts the pauta criterion method for data cleaning, which is suitable for the case where the sample data is relatively large [5]. The calculation formula is:

\[
|\text{value} - \text{mean}| > 3\sigma \tag{1}
\]

Where value refers to the actual data, mean refers to the mean of the sample data, and \(\sigma\) refers to the standard deviation of the sample. According to the normal distribution law of random variables, after repeated tests, the test data are 99.73% in the range of \(\text{mean} - 3\sigma, \text{mean} + 3\sigma\), so the data not in this range is considered as small probability event, which is determined as abnormal data.

Driving performance and visual demand are two key factors for evaluating safety during finishing secondary tasks. This paper will evaluate the driving safety from the aspect of visual demand when drivers finish secondary tasks using voice interaction.

4.1. Sight deviation duration
As Fig. 2 shown below, using BeGaze analysis software to delineate the area of interest, we can obtain the duration of the sight in different areas during the selected time segment. SDD represents the time the driver's line of sight is in the car's interior area. The longer the SDD, the more serious the distraction and the greater the impact on driving safety.

Fig 2. Sight duration in different areas.

The following Fig.3 shows the comparison results of SDD of different models when performing voice interaction secondary tasks with the normal driving.
Compared with touch screen operation, voice interaction can effectively reduce drivers’ eye movement to the car. Through video analysis and interview, drivers will shift and keep their sight to the car for the following reasons:

1) Find the position of the button when waking up by the button;
2) Check the voice state, such as whether it is woken up or in processing;
3) Confirm whether the content of the voice expression is correctly recognized;
4) View the options, such as quickly viewing the listed addresses when navigating;
5) Check whether the task is executed correctly, such as the temperature of the air conditioner.

It can be found that to reduce the impact of voice interaction on SDD, clear feedback should be given, like telling the status of the system through a “di” sound, or to clarify the result after the task is completed, such as after adjusting the air conditioner, actively inform "the temperature of the air conditioner has been adjusted to 26°C", etc. WE model’s average SDD is 198.32ms and less than other models, an important reason is that by the NOMI voice assistant robot above the center console, the driver does not need to frequently look down to confirm the status of the system.

Significance tests are performed on the SDD during normal driving and when finishing secondary tasks using the voice interaction system for different vehicle models. The results are shown in Table 2 below:

**Table 2. Significance tests of sight deviation duration.**

|        | RR model | GB model | LK model | WE model | DA model |
|--------|----------|----------|----------|----------|----------|
| SDD    | 0.000    | 0.000    | 0.028    | 0.023    | 0.000    |

The result shows that the difference between the SDD of drivers using different voice interaction systems and the normal driving state is significant (p<0.05).

4.2. Revisit number
During driving, the driver does not shift his eyes to the car once until the task is completed but obtains the information in the car through several short revisits. If the driver needs to shift his eyes frequently, it means that the driver cannot focus, and his eye movement is seriously distracted. Revisit numbers (RN) represents the number of times that the driver's eyes shift from the road area, i.e. Area 1, to the...
interior area, i.e. Area 2. Zhao Xincan and others pointed out that the RN is an important indicator of search efficiency in usability [6]. The more the RN, the unreasonable layout of the relevant area. The RN is related to the visualization of the voice interaction system.

The following Fig. 4 shows the comparison results of RN of different models when performing voice interaction secondary tasks with the normal driving.

![Box-plot of comparison results of RN.](image)

**Fig 4.** Box-plot of comparison results of RN.

During normal driving, the average RN is 0.31, which means that the driver will also shift his sight into Area 2 from time to time and constantly turn his eyes to scan the surroundings during normal driving. The sight may be shifted to Area 2 during the longitudinal scanning process, and there are also situations in which the IVIS or other things in the car attract the driver’s attention. However, it can be seen from the SDD that the time in this case will be very short, and it will not affect driving safety.

Significance tests are performed on the RN during normal driving and when finishing secondary tasks using the voice interaction system for different vehicle models. The results are shown in Table. 3 below:

**Table 3.** Significance tests of revisit number.

|          | RR model | GB model | LK model | WE model | DA model |
|----------|----------|----------|----------|----------|----------|
| RN       | 0.004    | 0.000    | 0.054    | 0.207    | 0.000    |

The test results of RN shows that the difference of 3 models (RR, GB, DA) between using voice interaction systems and the normal driving state is significant (p<0.05). However, the p values of LK model and WE model are 0.054 and 0.207, respectively, indicating that there is no significant difference from normal driving.

4.3. Fixation count
When the drivers are gazing, they aim the central pupil of the eye at an object for a period of time. A gaze point will be formed on the object. When driving and completing secondary tasks, the driver’s eyeballs are not still, but continue to change the gaze point, and at the same time it will be accompanied by high-frequency eye tremor, slow drift and micro-jump [7]. BeGaze software can record the driver’s
gaze point and can also record the gaze point’s movement trajectory in the specified area within time segment after dividing the time. According to the research of Zhao Xincan and others, the number of gaze points is related to the efficiency of finding information during driving [8]. If the driver can quickly search for the required information, the number of gaze points in the car area will be less; on the contrary, if the driver is difficult to search for his own target information, he will repeatedly adjust his gaze point in the car area, this process will bring serious visual distraction, which will affect driving safety.

Fixation count (FC) refers to the number of gaze points the driver has in the car interior the task. The more FC, the more difficult and less efficient the driver finds the target information. The following Fig.5 shows the comparison results of FC of different models when performing voice interaction secondary tasks with the normal driving.

![Fig 5. Box-plot of comparison results of fixation count.](image)

The average FC during normal driving is 1.02. The FC of RR model, GB model, LK model, WE model and DA model when completing driving tasks by the voice interaction system are 3.13 times and 7.32 times, 2.15 times, 2.56 times and 7.48 times. It shows that when completing the secondary task by the voice interaction system, the driver's FC are significantly increased, and the driver will focus on finding the required information in the car. Significance tests are performed on the FC during normal driving and when finishing secondary tasks using the voice interaction system for different vehicle models. The results are shown in Table.4 below:

|               | RR model | GB model | LK model | WE model | DA model |
|---------------|----------|----------|----------|----------|----------|
| FC            | 0.000    | 0.000    | 0.008    | 0.111    | 0.000    |

The test result shows that, except WE model, the difference between FC and the normal driving state of the driver using different voice interaction systems to complete the secondary task is significantly different from the normal driving state (p<0.05). The p-value of the WE model is 0.111, which is not significantly different. It can be seen from the interview with the tested driver that the NOMI robot of the WE model can provide feedback status, and the robot’s position is above the center console. It is
flush with the ground and belongs to the road area (Area 1) in the area division of this paper, so the difference between the data and the normal driving state will be reduced.

5. Conclusion
This paper studies the voice interaction system of ICV in China market, and analyzes its impact on driving safety. Through the data, it is found that the use of voice interaction system to complete secondary tasks will have an impact on visual demand, cause distraction and then affect driving safety. For Sight deviation duration, the value under voice interaction task is longer than when normal driving. The revisit number and the fixation count are significantly more than the normal driving state.

Reference
[1] Wang Wuhong. Vehicle human-machine interaction safety and assisted driving (in Chinese) [M]. Beijing People's Communications Press, 2012.
[2] Dewar R E. In-vehicle information and driver overload [J]. International Journal of Vehicle Design, 1988, 9(4-5): 557-564.
[3] Wang Ying, Zhang Wei, Wu Su, et al. Overview of research on vehicle information system and Driving Safety (in Chinese) [J]. Science & Technology Review, 2009, 27(13):105-110.
[4] Braun M, Broy N, Pfleging B, et al. Visualizing natural language interaction for conversational in-vehicle information systems to minimize driver distraction[J]. Journal on Multimodal User Interfaces, 2019, 13: 71-88.
[5] Zhang Wentong. SPSS 11 statistical analysis tutorial. Advanced (in Chinese) [M]. Beijing: Hope Electronic Press, 2002.
[6] Zhao Xincan, Zuo Hongfu, Ren Yongjun. Overview of eye tracker and gaze tracking technology (in Chinese) [J]. Computer Engineering and Applications, 2006(12): 122-124+144.
[7] Meng Ni. Analysis of driver's gaze behavior in different road traffic environments (in Chinese) [D]. Changan University, 2009.
[8] Zhao Xincan, Zuo Hongfu, Ren Yongjun. Overview of eye tracker and gaze tracking technology (in Chinese) [J]. Computer Engineering and Applications, 2006(12): 122-124+144.