



Do Drivers Prefer Female Voice for Guidance? An Interaction Design About Information Type and Speaker Gender for Autonomous Driving Car

Wang Ji¹(✉), Ruijie Liu², and SeungHee Lee¹

¹ University of Tsukuba, Tsukuba 305-8577, Japan
wangjige@foxmail.com

² Shaanxi Normal University, Xi'an 710062, China

Abstract. This research is to get a better understanding of what kind of intention is needed to communicate between drives and autonomous car. What types of communication methods should be used to offer a better user experience to drivers and passengers. In this study, we examined the effects of information type (“why” message, “how” message) and speaker gender (female voice, male voice) voice messages that provide communication about the car operations in advance before the autonomous car’s action. 30 participants took part in this experiment. Results showed that female voice was more trustworthy, acceptable and pleasure than the male voice, because participants were more familiar with female voice as it was widely used in smart device and navigation system. Also, the female voice is easier to hear than the male voice in the noisy condition. More than half of the participants changed their preference for the speaker gender, which reminds researchers to be cautious about the results of an online survey of drivers’ attitude to speaker gender. How message is preferred by participants. It is considered to be more intelligent and interesting than why message as it described what automated activity it is undertaking directly, participants didn’t need to take part in the decision process which may add their cognitive load. Additionally, how message uses the first person word like “I” “we” or “us” to communicate with the drivers increase participants’ trust and satisfaction.

Keywords: Voice interaction design · Autonomous car · Speaker gender · Information type · Usability evaluation · User experience

1 Introduction

1.1 Background

The autonomous car has been predicted in the science fiction and discussed in popular science media for many years. Recently, with the development of new technology and the need for the new transportation revolution, autonomous cars have been an essential strategy of the motor corporations. Most of them and some internet companies have announced their plans to begin selling such cars in a few years. In the new century, it is

widely believed that technology will continue to increase the efficiency and safety of transportation. One of the major breakthroughs will be the autonomous car. It is expected to alleviate or completely solve all these serious problems the traditional car faces. By automating the vehicles, human error can be significantly reduced since the intelligent system will never get drunk or be distracted, and it can be designed to execute often appropriate maneuvers that avert the crash entirely. Also, smaller headway and better driving behaviors can be achieved, which will bring out less traffic congestion. Autonomous car will also benefit specific groups of people who are not able to drive by themselves for any reason, they will enjoy the mobility facilitated by the new technology. In summary, it is probable that the autonomous car will likely have a profound impact on society.

1.2 Levels of Automation

Fully autonomous car and semi-autonomous car are the two main type of intelligent cars. Fully autonomous cars are completely controlled by their automated systems such as artificial intelligence, machine learning, GPS, deep learning and land changing technology. Semi-autonomous cars need drivers control together with some automatic functions such as automated parking, cruise control and so on. Multiple classification systems are published to understand better and direct the development of autonomous car technology. Recently, the NHTSA updated its standardization to aid clarity and consistency in the federal automated vehicles policy⁶ [1]. It adopts the SAE International (Society of Automotive Engineers) definitions for levels of automation. SAE definitions divide vehicles into levels based on “who does what and when” [2].

1.3 Manufactures

This market for autonomous vehicles is expected to be quite large in the coming decades, which attracts many companies invest a lot of money and human resources to start their research in this sunrise industry.

Automakers are the main driving engines behind the research and development of autonomous car technology in the major car-producing countries. These manufacturers are comprised of two main subgroups, traditional car manufacturers, and new technology innovators. Those tech giants play the role of innovators to develop new strategies and models for automation and single occupancy cars. But they still need to cooperate with the traditional automakers as they do not have the equipment and factory to produce automobiles. According to a report from the Brookings Institution, several key challenges will be faced as intelligent cars emerge in order to make all the great dreams come true [3]. These crucial challenges include technical challenges or societal action. Each of these matters poses problems for autonomous vehicles and their success in the market.

1.4 Public Acceptance

Whatever the future translation will be, the core part will never change—humans. No matter how autonomous the car is and what kinds of automaker the company is, all

those new technology and products are developed for supporting a better life for humans, and the company needs to make a benefit from their customers. Automakers and researchers must consider the psychological dimensions of technological design. Even the greatest technology, such as vehicles that drive themselves, is of little benefit if consumers are unwilling to use it.

Ultimately, the public must feel comfortable with the autonomous car for this market to develop. As with any emerging technologies, it takes a long time for individuals to accept new models of travel. According to a survey from AAA [4], More than 75% American drivers feeling “afraid” to ride in a self-driving car. In new emerging market countries like China, people appear more open to vehicular experimentation. The view was echoed in a separate survey undertaken by the Roland Berger consulting firm. It found that “96% of Chinese would consider an autonomous vehicle for almost all everyday driving, compared with 58% of American and Germans” [5]. Because Chinese people do not have a long history to own a private car, their emotional relationship related to the car is not so positive as the Americans, Germans or Japanese who have their unique profound car culture, so they are more amenable to autonomous vehicles. Before drivers are ready to switch their control to an autonomous car, the automaker and new technology companies have to take this question into considering: how to design the car, especially the interaction system between a car and a driver, what can help to establish and increase drivers’ trust in their autonomous car.

According to a public opinion survey in the US, many people prefer to use voice or sound to interact with their autonomous car. When asked about their preferences, 36.2% of Americans said they prefer to tell their self-driving vehicles their route or destination by voice command. More than 90% of Americans prefer to be notified by a combined warning mode that includes sound (sound + vibration + visual 59.4%, sound + visual 19.4%, sound + vibration 9.7%) or only sound warning (7.9%) when they need to take control of their partially self-driving vehicle [6]. The voice has a long history in Human-Computer Interaction, often as a supporting function in the user interface as a way to notify the user about their actions. In the traditional human-vehicle interactions, the main function of the voice message is to alert the driver to the actions of the car and the changes in the surrounding environment. The researcher found that professional drivers preferred auditory signals since it could provide a warning as well as arouse the drivers [7]. Thus, in a traditional in-vehicle warning system, auditory signals are often used to notify drivers about impending danger, operational feedback, and malfunction warnings. With existing high demands on a driver’s visual attention, many on-board computer systems utilize speech technologies. While speech-based interfaces may seem to be a safer option than graphical user interfaces in cars, there remains a concern that the cognitive demands associated with richer communication will damage driving safety [8]. Natural language dialog may become the preferred mode to inform drivers of control issues, as well as changing road and environmental conditions, to help give an explanation for the decision that the vehicle makes as it is in a natural way people communicate with each other.

Speaker Gender. Speaker gender is the first impression a user gets about the products and company [9]. A combination of real world mishaps and controlled experimental studies has shown that several factors significantly affect driver responses to voice

interfaces in cars, including perceived speaker gender, emotion and even age. The BMW5-series released in Germany included a voice-based navigation system, featuring a computer-generated voice with female characteristics [10]. Although these drivers were well-aware that the voice was computer-generated, they reacted with gender-stereotyped responses, ultimately rejecting the female voice and demanding a product recall.

From our life experience, we can find that most of the in-vehicle information systems, such as navigation systems are using a female voice. Also, most internet companies develop the voice assistant as a female, such as Apple Siri, Microsoft Cortana, Google and Amazon Alexa. Voice is also a brand and copyright of the company. Using voice can reinforce the company's brand identity. The study found that listeners can encode speakers specific voice attributes in memory along with information about the specific word [11]. Also, another study reported that if participants' familiarity with a talker's voice will lead to higher intelligibility of speech enunciated by that speaker, suggesting a memory link between listener sensitivity to speaker-specific voice and overall speech intelligibility [12]. So voice branding can be a carrier for conveying a memorable message to targeted consumers, taking advantage of the powerful memory sense of sound. Sound or voice used in the in-vehicle system can help the car companies to increase their brand awareness, distinguish themselves from their competitors and connect with their users on a deeper level. Taking Apple Siri as an example, Siri is available in all the ecosystems of Apple products. Users can talk with their Siri on their iPhone, iPad, Apple Watch or Macbook, they don't need to learn how to interact with their "voice assistant" in a new device as they are so familiar with the voice, just like talking with an old friend. Gender is one of the most psychologically powerful social categories.

The Intelligibility also affect people's impression of voice. Intelligibility of the female and male voice is equivalent in most normal conditions. Because of the small differences between female and male acoustic speech signals, the intelligibility of the voice one can be different in certain situations such as in the condition with high levels of noise. A study of aircraft cockpits suggested that the lessened interference between cockpit noise and female voices made female voices easier to be heard in the cockpit than male voices [13].

Little is known about the effects of the in-vehicle information system of the autonomous vehicle on the driver, as it is a sunrise industry. Autopilot is a new function for the automobile industry, but in the aviation industry, pilots have worked with autopilot from the 1930s. The autopilot in aircraft is a system used to control the trajectory, direction and altitude without a pilot handling the wheel. Autopilot assists the pilot in controlling the aircraft, allowing them to pay attention to broader aspects of the operation and can significantly reduce workload during critical phases of flights [14]. So the methods of studies on the cockpit information system are very valuable.

More recent researches, however, have indicated that the original popular hypothesis may be unreliable since more females have been employed as pilots and air traffic controllers. General human factors now indicate largely that, either due to current culture or changing attitudes, an automated female voice is no more or less effective than a male voice [15].

One important characteristic used to distinguish between gender is the pitch. The pitch is associated with the physical sensation of frequency. In humans, the auditory frequency is positively correlated with the perceived pitch. Higher voice pitches are usually associated with the female voice, and they also tend to elicit higher perceived urgency [16].

A study found that both acoustic and no acoustic differences between male and female speakers are negligible [17]. Therefore, they recommended, the choice of the speaker should depend on the overlap of noise and speech spectra. The female voice did, however, appear to have an advantage in that they could portray a greater range of urgencies because of their usually higher pitch and pitch range. They reported an experiment showing that knowledge about the gender of a speaker has no effect on judgments of perceived urgency, with acoustic variables accounting for such differences. A study from Defense Research and Development Canada in Toronto [18], found that with simulated cockpit background radio traffic, a male voice rather than a female voice, in a monotone or urgent annunciation style, resulted in the largest proportion of correct and fastest identification response times to verbal warnings, regardless of the gender of the listeners.

The study of social psychology has demonstrated that gender makes a difference on a lot of human-human interaction [19], this gender stereotype also appeared in human-machine interaction as computers are social actors. A computer-generated speech study showed that the male-voiced computer exerted greater influence on the user's decision than the female-voiced computer and was perceived to be more socially attractive and trustworthy [20]. More strikingly, gendered synthesized speech triggered social identification processes, such that female subjects conformed more to the female-voiced computer, while males conformed more to the male-voiced. Similar identification effects were found on social attractiveness and trustworthiness of the computer. As gender can influence behavior, it is important to choose a proper voice for the in-vehicle information system of the autonomous car.

Voice Content. One of the essential factors for the success of the autonomous car is providing feedback to the driver. Norman pointed out that the problem with automation is that its condition is inadequately communicated to the drivers [21], even this kind of communication can assist the drivers into the control loop. He noted that the inappropriate feedback had made to some serious incidents in the aviation domain. The results of some previous studies [22] also suggested that providing feedback to make the driver understand the car's operations and the surrounding situation is a key feature of autonomous driver assistance systems. Moreover, the context, integration and abstraction should be proper. Otherwise, the driver may not understand the feedback [23].

If we consider the driving task as teamwork, driver and the car are teammates who collaborate to ensure the driving activity will be safely and efficiently. How to get a better understanding of what kinds of voices need to be used and what types of messages should be sent. Hence, the study and design of an appropriate feedback model is a foundation for increasing the drivers' trust in the autonomous car.

A study from Koo [24] in the USA claimed that the nature of feedback is to tell drivers the outcome of the system's operation, and deliver the operation information after the event. However, in the condition of autonomous driving, the information

about the operations and situations should be provided to drivers ahead of the event, he named this kind of information: “feed-forward” information. As drivers transmit their control power to the car, how drivers perceive and accept the autonomous function of the car become increasingly. For instance, when there is a red light ahead and the car is about to stop automatically. The automated system could provide the information about how the car is going to act or supply message regarding why the car is going to conduct that activity. That’s the two main types of information the drivers are interested in. In Koo’s study. They explored those two types of feed forward information: How message (what automated activity it is undertaking) and Why message (reason the car is acting that way) in the study to explore which types of information can allow the drivers to respond appropriately to the situation and gain trust the autonomous car is taking control for good reason. In their study, they tested three different design settings. This study found that the driver preferred receiving the message only with why information, which created the least anxiety and high trust. This study took place in the US and explored how the context of semi-autonomous affect driver’s attitude. Is there any difference in the condition of a fully autonomous car as the driver does not need to operate? Whether there is a cross-cultural difference between the western and eastern drivers will be an important issue, as the previous study showed that East Asians are holistic thinkers [25]. Holistic thinking encourages people to have a worldview in which all kinds of events and phenomena are interrelated and perpetually changing. This implies that East Asians are more attentive to the behavior or flow of objects. In contrast, Americans are more analytic. They pay attention primarily to the object and the categories to which it belongs and using rules, including formal logic, to understand its behavior. So the East Asian drivers may have a different preference on the type of feedforward information. The current auto industry is a global business, offering the product suited to local users will bring great benefit to the automakers.

1.5 Present Study

An autonomous vehicle is a new model of future transport; few studies were done on the in-vehicle information system. For vehicles the most important part is safety, not only at the technology level but also giving the driver a feeling of safety and trust, this is also a very significant field that the researcher should pay attention to. The present study examines speaker gender as well as voice content; the former study only focused on voice content or speaker gender. Our study explores two main questions about the in-vehicle information system: (a) whether participants have a preference for speaker gender; (b) whether participants have a preference for voice content.

2 Method

2.1 Study Design

A 2 (male voice, female voice) \times 2 (Why message, How message) factorial design was used.

2.2 Participants

30 participants (15 males and 15 females) aged between 21 years and 35 years (mean = 24.6 years, standard deviation = 2.58) took part in this experiment. 24 participants are Chinese (12 males and 12 females), and the other 6 participants are Japanese (3 male and 3 females). All of the 30 participants have a good English ability, and they can understand all the English prompts used in the experiment. They all had a driver license for more than one year, and none had any obvious hearing abnormalities. Before the experiment, each participant was briefly introduced about the purpose of the experiment, and a short interview was conducted to get the information about their driving experience, personal preference about speaker gender of voice assistant (Including the voice assistant of their smart device and in-vehicle information system such as navigation system). Each participant was paid 500 Japanese yen (500 yen = \$4.59) for the total test duration of around 35 min.

2.3 Apparatus

In this experiment, we use TTS (Text to Speech) technology to make the voice message. Text-to-Speech, abbreviated as TTS, is a technology that converts digital text into the spoken voice output. Text-to-Speech systems were first developed to aid the visually impaired. They are nowadays ubiquitous, having an extremely broad field of application ranging from voices giving directions on navigation devices to voices for public announcement systems and virtual assistants [26]. We use the TTS provided by Ivona (<https://www.ivona.com/us/>) and chose the British English Brian as the speaker of the male voice message and American English Kimberly as the female voice. Both of the voice with no particular inflections or moods.

Table 1 Voice prompts used in the experiment. The same line is the same car action. For example, when there is a red light in front of the car and the car is going to brake to stop. In why message experiment setting, the voice content is “Red light ahead”, in how message experiment setting, the voice content is “Car is braking.”

Table 1. Voice prompts used in the experiment.

Why message	How message
“Autopilot start”	“I will take the control”
“Red light ahead”	“Car is braking”
“Road clear”	“Speed up”
“Vehicle in front is too slow”	“We will pass the vehicle in front of us”
“Intersection ahead”	“Car is barking”
“Roadworks ahead”	“Turning around”
“Autopilot ending”	“Car is stopping”

The video used in the experiment is from an autopilot demonstration of a Tesla Model SP85D. The 4 min 6 s video includes highway, urban, and suburban road situations with stop signs, road work sign, and traffic signals. The car in the video

conducts the following operations: start the autopilot model, brake, speed up, overtake, cross intersection, steering, turn around, and ending the autopilot model. In the experimental conditions, whenever the car changed its operation, the voice message was generated. All the voice messages were sent 1~2 s before the car conducts the operation. Such a voice message ahead of the operation allows the participants to understand the current situation. To minimize effects caused by the lack of sound feedback that exists in the real driving situation, we added the background sound of driving in the video, an engine sound was simultaneously provided as a cue of an accelerating action every time the car was speeding up or starting. All the voice prompts used in the video are in English as participants are from different countries, but all of them have a good English ability.

The experiment videos was run on lab computer—A Macbook Air 13.3 2015 was used to play the videos of all experiment setting. The videos were projected using an ASUS DLP S1 projector, the biggest brightness is 200 lm. Speaker in the projector was used to output the audio for all videos.

2.4 Questionnaire

Attitudinal measures were based on self-reported data on adjective items in the post-drive questionnaire. The questionnaire was adapted from a published model from the CHIME Lab at Stanford University that is used to measure driver experience [27, 28]. Participants were asked to rank each item on a nine-point Likert scale ranging from “Describes very poorly (= 1)” to “Describes very well (= 9)” The questionnaire combined an English and Japanese version on one paper. Four same questionnaire papers were given to the participants, and required them to finish one questionnaire just after they watched the video, the time to finish was no limited.

2.5 Procedure

This experiment was conducted in a small quiet room without light and all the participants were tested individually. After they had came to the room, they received a packet consisting of an approved human subject consent form, four questionnaires and a pencil. The experimenter asked them to read the consent form and sign the form if they agree to participate in the experiment (See Appendix A and B). Then, the experimenter gave a brief introduction of the whole experiment. An interview was conducted with the participants to know their preferences regarding the speaker gender of their smart device voice assistant and in-vehicle navigation system. Also they were asked to explain the reason of their preferences. Three questions were be asked as follows:

Q1 Do you have a driver license? How many years have you driven a car?

Q2 In the future, if the self-driving car is available what kind of voice system do you hope it to be, in terns of kind of information or voice type? And why would you like this?

Q3 As we know most navigation system sin the car now are female, which do you prefer in the future self-driving car? Could you tell me the reason why you like.....

If they preferred a female voice, the order was Female-Why—Female-How—Male-Why—Male-How, otherwise the order was Male-Why—Male-How—Female-Why—Female-How. For the convenience of data collected, we use “WF” for Why message Female voice, “HF” for How message Female voice, “WM” for Why message Male voice and “HM” for How message Male voice. The details about participants’ order and preference are shown in Appendix D.

Participants were asked to sit in the middle of the room, about 2 m away from the screen, they can take the seat and adjust their gesture. The videos were projected onto a wall directly in front of the participants. The size of projected area was about 1.5 m diagonal; the participants had about a 60-degree field of view. The brightness of the display is 80% of the biggest brightness, about 160 lm. The intensity levels of test voice ranged between 55–75 dB measured by sound level meter depending on the comfortable threshold of participants. When they were ready to begin the experiment, they should motion to the experimenter by hand. Then experimenter turns off the light in the room and played the video and projected the frame on the screen. After watching the video, participants filled out the questionnaire that assessed their overall experience. They filled in the first questionnaire consisting of general information such as gender and age in addition to nationality. The same as the second, third and fourth trial.

After all the four trials had been finished, a short post-experiment interview was conducted to explore whether participants have a new attitude towards the speaker gender, voice content and explain the reason they change their preference. Three questions were asked:

Q1 After you watched the four videos, you still prefer your choice at the beginning or did anything change?

Q2 About the information types which do you prefer? Why?

Q3 Is there any place you think the voice system can be improved?

Afterwards, the experimenter thanked participants for their participant and gave 500 Japanese yen (500 yen = \$4.59) as the gift. All the participants were debriefed at the end of the experiment.

This experiment had received the permission from the research ethics review committee of School of Art and Design, University of Tsukuba, research ethics number 150422038.

3 Results

3.1 Descriptive Statistics

Figures 1, 2, 3 show the distribution of 30 participants’ preference for speaker gender and voice content. Interesting data is that 57% participants changed their speaker gender preference after the experiment. For the female participants, 2/3 of them changed from male voice to female voice, and 1/3 from female voice to male voice. Most of the male participants (7 out of 8) made a change on the speaker gender, from female voice to male voice.

■ male voice ■ female voice

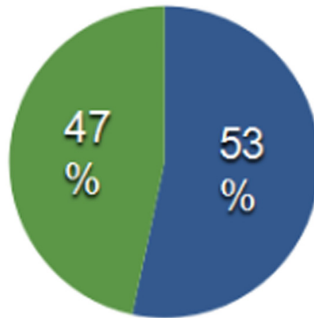


Fig. 1. The distribution of 30 participants' preference for speaker gender before experiment. 57% participants said they prefer female voice as the voice used in the future in-vehicle information system autonomous car while 43% participant would like to choose male voice.

■ male voice ■ female voice

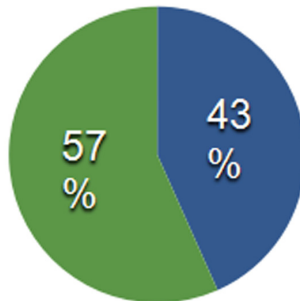


Fig. 2. The distribution of 30 participants' preference for speaker gender after experiment. 53% participants thought male voice is a better choice for the autonomous car voice system and 47% participant viewed the female voice is more suitable for voice assistant in autonomous car.

3.2 Chi-square Test

A Chi-square test was performed to examine the relation between a participant's gender and their preference for speaker gender before the experiment. The relation between these variables was significant, $X^2 = 6.652$, $p = 0.01 < 0.05$. Female participants were more likely to choose male voice than the male participants, and the male participants were more likely to choose female voice than the female participant.

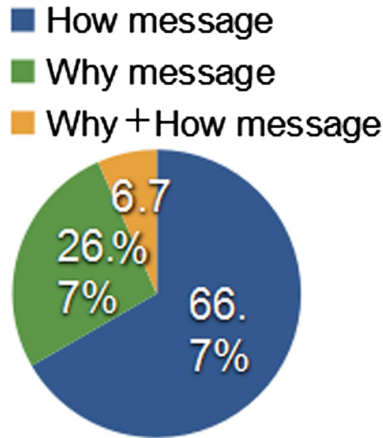


Fig. 3. The distribution of 30 participants’ preference for voice content. 66.7% participants said they like How message, 26.7% prefer Why message, other 2 participants thought Why + How message is the best choice for in-vehicle information system.

A Chi-square was conducted to examine the relation between participant’s gender and their preference for speaker gender after the experiment. There is no significant difference between these variables, as $X^2 = 0.536, p = 0.464 > 0.05$.

3.3 ANOVA

A speaker gender (female voice vs. male voice) x voice content (why message vs. how in message) repeated—measures ANOVA was conducted with the ratings of all 15 adjective words as dependent variable. The purpose of the analysis was to compare the effect of speaker gender and voice content on ratings of adjective words which were used to evaluate the participants’ experience. The summary results of ANOVA are shown in Appendix E.

There was no significant interactive effect on speaker gender and voice content in the rating of all the 15 adjective words, but it shown significant main effects on speaker effect gender and voice content as follow.

There was a significant main effect on speaker gender in the rating for “trustworthy”, the rating for female voice ($M = 6.133, SE = 0.221$) is higher than the rating for male voice ($M = 5.650, SE = 0.276$), $F(1,29) = 5.608, p = .025 < .05$. These results suggest that speaker gender really does have a effect on participants’ trust in the voice system. Specifically, the result suggests that a female voice is more “trustworthy” than a male voice.

A significant main effect of voice content is shown in the rating for “uninterested”, the rating for why message ($M = 4.183, SE = 0.310$) is higher than the rating for how message ($M = 3.583, SE = 0.256$), $F(1,29) = 8.534, p = .007 < .01$. These results suggest that voice content really does have an effect on participants’ interest in the voice system. The results show that the why message is less is “interesting” than the how message.

There was a significant main effect of voice content in the rating for “intelligent”, the rating for how message ($M = 6.550, SE = 0.260$) is higher than the rating for why message ($M = 5.983, SE = 0.215$), $F(1,29) = 6.227, p = .019 < .05$. These results suggest that voice content really does have an effect on participants’ view about the intelligent level of voice system. Specifically, the results suggest that the how message is more “intelligent” than the why message.

The ANOVA reveals that participants felt the how message is more “stimulating” than the why message. The rating of “stimulating” for how message ($M = 4.600, SE = 0.316$) is higher than the rating for why message ($M = 4.000, SE = 0.292$), $F(1,29) = 6.178, p = .019 < .05$. These results suggest that voice content really does have an effect on participants’ attitude to the stimulated level of voice system. The results suggest that the how message is more “stimulated” than the why message.

What’s more, there was a marginal significant main effect of speaker gender in the rating for “acceptable”, $F(1,29) = 3.955, p = .056 < 0.1$, female voice ($M = 6.350, SE = 0.212$) is higher than the male voice ($M = 5.950, SE = 0.263$), which means the female voice is more “acceptable” than male voice. Another marginal significant main effect of speaker gender is observed in the rating of “pleasure”, $F(1,29) = 3.754, p = .064 < 0.1$, female voice ($M = 5.450, SE = 0.248$) is higher than the male voice ($M = 4.983, SE = 0.271$).

4 Discussion

The major goal of this study is to find out whether drivers have a specific preference on the speaker gender and voice content in the future autonomous car driving. The results showed that participants did not have a preference on speaker gender, but the female voice had a significantly higher rating than the male voice at some adjective words. In the voice content, both results of the descriptive statistics and ANOVA showed that how information is more popular and welcomed by the participants. Those points will be discussed more thoroughly below.

According to the research from Cozby [29], in the repeated measure design, the order of presenting the treatments may have an effect on the dependent variable. The ANOVA found that the order of the experiment has no significant effect on participants’ preference for speaker gender and voice content in this study.

The result showed that the number of participants who prefers female voice is reduced after the experiment. In the pro-experiment gender voice choice, male participants preferred to choose female voice, 8 out of 12 reported that their voice assistant they used such as Siri, Google now or the navigation system in their vehicle are female voice, they hope the autonomous car can keep this design. A study of warning signals also suggested that when introducing a new signal or voice, it should not be too different from the existing ones [30]. Rather, Female participants prefer male voice, 7 out of 10 reported that the reason they like male voice was because their early memory about car and driving were always together with their male family members who drove them everywhere, taught and guided them the experience and skills about driving, so they reported that it would make them feel relaxed and trustworthy if the autonomous car uses a male voice to interact with them.

More than half of the participants changed their choice. Most of them claimed the voice was not what they expected. The autonomous car is not yet available to the public so the participants can only imagine they are driving an autonomous car, but there can be a huge gap between the real situation and their imagination. Even though in this study, we tried to provide a reliable method for manipulating the experiment; there is a fair amount of difference in fidelity between the experiment setting and real-life autonomous driving.

7 male participants who changed their preference reported that the female voice used in this experiment was a little annoying and shrill because of female high voice pitch. Meanwhile, 8 female participants switched to choose female voice after the experiment; they said the male voice was too vague to hear clearly. Edworthy and Waring's study also found that a quiet, low-pitched sound or a loud, high-pitch sound are less pleasurable than the sounds between these two extremes [31].

This finding can be explicated by the fact that auditory frequency is positively correlated with perceived pitch. Lower voice pitches are usually associated with the male voice, and the female voice is contacted with higher. Studies from neuroscience showed male and female have a significant difference in hearing sensitivity, that female show a greater sensitivity to sound than the male [32]. The results were consistent with the study of human-robot interaction [33], that male and female subjects both anthropomorphized the robot with the same-gender human voice more strongly compared to the robot with opposite gender human voice, they showed greater acceptance and felt psychologically closer to the robot which shared the same gender with them. Considering all of those facts, it can explain why the males and females changed their preference to the voice of the same gender with them.

The Chi-square test result showed no significant difference on participants' preference for the speaker gender after experiencing all four experiment stimuli. To investigate further, we conducted a repeated-measures ANOVA with the ratings of all 15 adjective words as the dependent variables to compare the effect of speaker gender.

Participants rated enunciated by the female as more trustworthy, acceptable and pleasurable than message enunciated by the male. In the theory of "Computer As Social Actors", it proposes that people actually engage in the same kinds of social response to the computer as they do with humans [34]. Nass's studies found that users treat the voice of a computer as the social response but not an unseen human "behind" the computer, usually a programmer [35]. While this theory approaches to human-vehicle interaction, it is a new way to look at the autonomous car as the source of information. The participants do not define the voice as a driver for the autonomous car who is taking control of their car, but as the voice assistant in the car. Almost all were the same as the navigation system they used in the car were female voice. Participants were wholly aware that the voice was a narrator for the autonomous car system (as opposed to some human "behind" the car). From their experience, they would rather trust a voice they are familiar with, that can be the explanation of this result. Also, 8 participants reported that the male voice is on a low pitch, the message is not clear, this result supported the study that female voices are easier to hear in the noisy condition than male voice [11], as we add the background sound in the experimental videos. The lower Intelligibility of the male voice in the noisy condition may also contribute to the result that participants rated male voice lower on "trustworthy", "acceptable" and "pleasurable".

Even though, studies about the cockpit warning voice show a male voice resulted in a larger proportion of correct and faster response times to verbal warning than a female voice [15], the role and function of speech systems in the cockpit and autonomous car are completely different. Warning in the cockpit is asking the pilot to take control of the aircraft and deal with the emergency situation. However, in the autonomous car, the only thing that drivers need to do is switching their control to the car. Moreover, vehicle drivers were not got the professional training for years to acquire a license as the aircraft pilots. So the lessons from aviation have limited application to autonomous car design due to the difference in operating methods for the different machines in different “road” condition [36].

Less than 1/3 of participants like the why message more than the how message and rated the why message as “uninterested”, not “intelligent” and not “stimulating”. In Endsley’s model of situation awareness (SA) [37], he explained the concept of SA. He defined the SA as the ability to perceive the related elements of the surrounding, to understand the current situation. However, firstly the participants need to be aware that the autonomous system already took control of the car and is driving. We can imagine one scenario: an autonomous car gives the driver a message “Stop sign ahead!”. At this moment, the driver will look for a stop sign, after he notices the stop sign the car is braking to stop, so the driver must process two types of information: the situational status and the car’s status. Then the driver can know what the car is doing and the reason it performs such an action. As a consequence, the driver’s cognitive resource is overloaded, as an autonomous car owner, he may think the car is not “Intelligent”, he also needs to take part in the decision process together with car. Moreover, sometimes the stop sign is not easy to find, which will easily cause the driver lose their patience to enjoy the interaction with the car. The next time, the voice message comes again, the driver will show no interest to listen carefully. In addition, the why message is not considerate to provide information that can help the driver know what to do next. Lee’s study showed that the content used in displaying a system’s operational status play a key role in building driver trust in the vehicle’s automation capabilities [38]. As how message described what automated activity it is undertaking directly, but why message informed the divers the reason the car is acting that way, the result is consistent with Lee’s finding.

The finding that participants prefer the how message can also be explained by the concept of anthropomorphism. A previous study showed the user who drove the anthropomorphized vehicle with enhanced humanlike features (name, gender, voice) reported trusting their vehicle even more [39]. Technology appears better able to perform its intended design when it seems to have a humanlike mind. How information uses the first person word like I, we or us (“I will take the control” “we will pass the car in front of us”) to interact with the driver, which make them like they are combined with the car, what the car does is just like what they do. Anthropomorphism of in-vehicle information system predicts trust in that car.

5 Conclusion

The female voice was found to be more trustworthy, acceptable and pleasurable than the male voice, because participants are more familiar with female voice as it is widely used in smart devices and navigation systems. In addition, the female voice is easier to hear than the male voice in the noisy condition. More than half of the participants changed their preference for the speaker gender, which reminds researchers to be cautious about the results of an online survey of drivers' attitude to speaker gender of the autonomous car, as there is a fair amount of difference in fidelity between real autonomous driving and participants' subjective imagination. Additionally, researchers should take the difference of autonomous car and aircraft into consideration when they apply the methods and lessons from aviation studies.

How a message is preferred by participants. They thought the how message is intelligent as it described what automated activity it is undertaking directly, they didn't need to take part in the decision process which may add their cognitive load. In addition, how message uses the first person word like "I" "we" or "us" to communicate with the drivers. Those anthropomorphized features made participants trust it more.

5.1 Limitation

One limitation of our study is that the sample group is limited demographically: participants are university students, most of them are less than 30 years old. A wider range of participants—including newer drivers as well as elderly drivers with longer experience but possibly slower perception and reaction times—could yield an opportunity to more broadly generalize our results or to produce different findings.

As the technology and time were limited, we only used two speaker voices, a wider range of speaker voices—including young as well as the middle age also with different accents and timbres—could yield an opportunity to produce different results. Therefore more study is needed to see if the speaker gender, or just the speaker's timbre, is affecting the participants' choice. Meanwhile, the study already showed that people like a natural voice more than a synthetic voice, this part also needs to be studied further.

5.2 Future Work

In this study, we used text-to-speech (TTS) voice, rather than pre-recorded human voice. Previous research has focused on speech quality, human voice versus TTS, and how it affects user's attitude change. Future studies should examine how speech quality interacts with speaker gender in the study of the in-vehicle information system.

Some participants reported that they did not like the why information, because the time between the voice and car action is too short to catch the point of what the car is going to do. We are going to set an experiment to test what timing is better for the interaction between the autonomous car and the driver, and whether the best timing is correlated with the on-going car speed.

References

1. Federal Automated Vehicles Policy—Accelerating the Next Revolution in Roadway Safety. <https://www.transportation.gov/AV>. Accessed 2016
2. Automated Driving Levels of Driving Automation Are Defined in New Sae International Standard J3016. https://www.sae.org/misc/pdfs/automated_driving.pdf. Accessed 2016
3. West, D.M.: Moving forward: self-driving vehicles in China, Europe, Japan, Korea, and the United States. Brooking (2016)
4. Three-Quarters of Americans Afraid to Ride in a Self-Driving Vehicle, 1 March 2016. <http://newsroom.aaa.com/2016/03/three-quarters-of-americans-afraid-to-ride-in-a-self-driving-vehicle/>. Accessed 2016
5. Schoettle, B., Sivak, M., Transportation, S.W.: Motorists' Preferences for Different Levels of Vehicle Automaton: University of Michigan Sustainable Worldwide Transportation (2016)
6. Chi, C.F., Dewi, R.S., Huang, M.H.: Psychophysical evaluation of auditory signals in passenger vehicles. *Appl. Ergon.* **59**, 153–164 (2017)
7. Mittal, R.: In-vehicle Multimodal Interaction. Doctoral dissertation, Arizona State University (2015)
8. Bem, S.L.: Gender schema theory: a cognitive account of sex typing. *Psychol. Rev.* **88**(4), 354 (1981)
9. Nass, C., Brave, S.: Wired for speech: how voice activates and enhances the human-computer relationship (2005)
10. Palmeri, T.J., Goldinger, S.D., Pisoni, D.B.: Episodic encoding of voice attributes and recognition memory for spoken words. *J. Exp. Psychol. Learn. Mem. Cogn.* **19**(2), 309 (1993)
11. Nygaard, L.C., Sommers, M.S., Pisoni, D.B.: Speech perception as a talker-contingent process. *Psychol. Sci.* **5**(1), 42–46 (1994)
12. Nixon, C., et al.: Female voice communications in high level aircraft cockpit noises—part II: vocoder and automatic speech recognition systems. *Aviat. Space Environ. Med.* **69**(11), 1087–1094 (1998)
13. “Automated Flight Controls” (PDF). [faa.gov](http://www.faa.gov). Federal Aviation Administration. Accessed 20 Feb 2014
14. Samuel Gibbs: China's Baidu could beat Google to self-driving car with BMW. 10 June 2015. <https://www.theguardian.com/technology/2015/jun/10/baidu-could-beat-google-self-driving-car-bmw>. Accessed 2016
15. Hellier, E., Edworthy, J., Weedon, B., Walters, K., Adams, A.: The perceived urgency of speech warnings: semantics versus acoustics. *Hum. Factors J. Hum. Factors Ergon. Soc.* **44**(1), 1–17 (2002)
16. Edworthy, J., Hellier, E., Rivers, J.: The use of male or female voices in warnings systems: a question of acoustics. *Noise Health* **6**(21), 39–50 (2003)
17. Arrabito, G.R.: Effects of talker sex and voice style of verbal cockpit warnings on performance. *Hum. Factors* **51**(1), 3–20 (2009)
18. Eagly, A.H.: Gender and social influence: a social psychological analysis. *Am. Psychol.* **38**, 971–981 (1983)
19. Lee, E.J., Nass, C., Brave, S.: Can computer-generated speech have gender? An experimental test of gender stereotype. In: *Extended Abstracts on Human Factors in Computing Systems, CHI 2000*, pp. 289–290. ACM, April 2000
20. Norman, D.A.: The “problem” with automation: inappropriate feedback and interaction, not “over-automation”. *Philos. Trans. Roy. Soc. London B Biol. Sci.* **327**(1241), 585–593 (1990)

21. Stanton, N.A., Young, M.S.: Vehicle automation and driving performance. *Ergonomics* **41**(7), 1014–1028 (1998)
22. Walker, G.H., Stanton, N.A., Young, M.S.: The ironies of vehicle feedback in car design. *Ergonomics* **49**(2), 161–179 (2006)
23. Lee, J.D., Seppelt, B.D.: Human factors in automation design. In: *Handbook of Automation*, pp. 417–436 (2009)
24. Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., Nass, C.: 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 (2015)
25. Nisbett, R., Peng, K., Choi, I., Norenzayan, A.: Culture and systems of thought: holistic versus analytic cognition. *Psychol. Rev.* **108**(2), 291–310 (2001)
26. What is Text-to-Speech? <https://www.ivona.com/us/about-us/text-to-speech/>. Accessed 2016
27. Takayama, L., Nass, C.: Assessing the effectiveness of interactive media in improving drowsy driver safety. *Hum. Factors J. Hum. Factors Ergon. Soc.* **50**(5), 772–781 (2008)
28. Takayama, L., Nass, C.: Driver safety and information from afar: an experimental driving simulator study of wireless vs. in-car information services. *Int. J. Hum.-Comput. Stud.* **66**(3), 173–184 (2008)
29. Mullennix, J.W., Stern, S.E., Wilson, S.J., Dyson, C.L.: Social perception of male and female computer synthesized speech. *Comput. Hum. Behav.* **19**(4), 407–424 (2003)
30. Lemaitre, G., Susini, P., Winsberg, S., McAdams, S., Letinturier, B.: The sound quality of car horns: designing new representative sounds. *Acta Acustica united with Acustica* **95**(2), 356–372 (2009)
31. Edworthy, J., Waring, H.: The effects of music tempo and loudness level on treadmill exercise. *Ergonomics* **49**(15), 1597–1610 (2006)
32. McFadden, D.: A speculation about the parallel ear asymmetries and sex differences in hearing sensitivity and otoacoustic emissions. *Hear. Res.* **68**(2), 143–151 (1993)
33. Eyssel, F., Kuchenbrandt, D., Bobinger, S., de Ruitter, L., Hegel, F.: ‘If you sound like me, you must be more human’: on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 125–126. ACM, March 2012
34. Reeves, B., Nass, C.: *How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications and Cambridge (1996)
35. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. *J. Soc. Issues* **56**(1), 81–103 (2000)
36. Koo, J., Shin, D., Steinert, M., Leifer, L.: Understanding driver responses to voice alerts of autonomous car operations. *Int. J. Veh. Des.* **70**(4), 377–392 (2016)
37. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. *Hum. Factors J. Hum. Factors Ergon. Soc.* **37**(1), 32–64 (1995)
38. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors J. Hum. Factors Ergon. Soc.* **46**(1), 50–80 (2004)
39. Waytz, A., Heafner, J., Epley, N.: The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* **52**, 113–117 (2014)