Female Voice Agents in Fully Autonomous Vehicles Are Not Only More Likeable and Comfortable, But Also More Competent

Jiayuan Dong, Emily Lawson, Jack Olsen, and Myounghoon Jeon
Mind Music Machine Lab, Department of Industrial and Systems Engineering
Virginia Tech, Blacksburg, VA

Driving agents can provide an effective solution to improve drivers' trust in and to manage interactions with autonomous vehicles. Research has focused on voice-agents, while few have explored robot-agents or the comparison between the two. The present study tested two variables - voice gender and agent embodiment, using conversational scripts. Twenty participants experienced autonomous driving using the simulator for four agent conditions and filled out subjective questionnaires for their perception of each agent. Results showed that the participants perceived the voice only female agent as more likeable, more comfortable, and more competent than other conditions. Their final preference ranking also favored this agent over the others. Interestingly, eye-tracking data showed that embodied agents did not add more visual distractions than the voice only agents. The results are discussed with the traditional gender stereotype, uncanny valley, and participants' gender. This study can contribute to the design of in-vehicle agents in the autonomous vehicles and future studies are planned to further identify the underlying mechanisms of user perception on different agents.

INTRODUCTION

The evolution of technology ecosystems has led to the development of a wide range of intelligent agents in human-computer interaction. For level five fully autonomous vehicles, in which a driver is not driving, intelligent agents can provide vehicle and driving information to the drivers and passengers, increasing trust in the autonomous vehicle system. To make them more effective, the development of the agents requires traits to be established, such as appearance, gender, voice tone, and immersion. Current voice agents, including Alexa, Siri, and Google Home, use a sociable female voice that is immersed in the system, but the agents in autonomous vehicles may warrant different traits to increase their effectiveness (Lee, Ratan, & Park, 2019). The present study investigated the effects of the embodiment and gender of the in-vehicle agent on drivers' perception in a fully autonomous driving vehicle.

Related Work

Research has shown that the in-vehicle agents contributed to improving driving performance, increasing drivers' situation awareness, decreasing drivers' anger level, and decreasing perceived workload in manual driving (Jeon, Walker, & Gable, 2015). In the same line, a study with a semi-autonomous driving tested trust and perceived safety based on the time until the participant manually took over control of the driving. They concluded that the auditory feedback with or without a visual co-driver contributed to persuading drivers to maintain the autonomous mode longer, which implies that drivers feel more comfortable with an in-vehicle agent (Hock, Kraus, Walch, Lang, & Baumann, 2016).

Previous studies have shown that people tend to prefer a male voice through a straightforward medium with an informative speech, and female voice through a sociable medium (Nass & Moon, 2000). Therefore, choosing an appropriate speech style for different purposes was considered as one of the important design factors for agents. This is also applied to driving agents. For example, with equivalent driving conditions and voice used, Large and Burnett (2014) found that participants could recognize the personality attributes of the in-vehicle agent's voice, which influenced participant's preference and attitude towards the agents. They also showed that an interface with a human-human-like interaction is more intuitive and natural and concluded that future studies with a conversational interface could address specific concerns of driver workload, fatigue and distraction, potentially enhancing trust in the technology (Large, Burnett, & Clark, 2019). Williams and colleagues showed that their participants rated the embodied robot as friendlier than the disembodied voice, regardless of gender (Williams, Peters, & Breazeal, 2013). However, Crowelly, Villanoy, Scheutzz, and Schermerhornz (2009) found that the female voice to be more reliable for the disembodied voice condition.

Based on these findings, we hypothesized that the embodied robot agents would be more preferred over disembodied voice only agents. In addition, we hypothesized that the female disembodied voice only agent would be preferred over the male disembodied voice only agent. However, because driving is a safety-critical dynamic environment, there might be diverse perceptions on different characteristics of the in-vehicle agents.

To test these hypotheses, in the present study we had two variables, creating four different types of agents: embodied robot versus non-embodied robot, and male voice versus female voice. These four agents were assessed in a fully autonomous driving simulator. The robotic agent was a Nao robot setting up next to the participant, while the non-robotic agent was spoken voice clips through the driving simulation system. Each of these agents had scripts in a male and female voice and all of the scripts used human-like conversational speech as opposed to machine-like informative speech (Lee, Sanghavi, Ko, & Jeon, 2019). Koo and colleagues classified the information types as "How", "Why" or "How and Why" messages and showed that the combination of How and Why message such as "Car is breaking due to the obstacle ahead" is preferred by their participants (Koo, Kwac, Ju, Steinert, Leifer, & Nass, 2015). Therefore, we created our script following the How and Why message as seen in Table 1. Literature shows that perception and trust toward the in-vehicle agents are significantly affected by traffic conditions and the context of information given from the agent (Cramer, Evers, Kemper, & Wielinga, 2008). Thus, the traffic scenarios and information were held constant across the conditions to eliminate situational biases from the participants.

METHODS

Participants

Twenty college students were involved in this experiment as participants (6 female and 14 male) with a valid driver's license. Participants' age ranged from 18 to 22 years old (*M* = 20.5; *SD* =1.05). The mean driving year was 3.45 years with an average of 5.7 driving times per week. Only three out of the twenty participants have experienced autonomous driving previously to the present experiment.

Experimental Design

A two by two within-subjects design was established including four in-vehicle agent conditions (robotic male agent (RM), robotic female agent (RF), non-robotic voice only male agent (NM), and non-robotic voice only female agent (NF)). Two fully autonomous driving scenarios with the conversational speech prompts (Lee et al., 2019) were used in the experiment across all conditions. In-vehicle agents' voices provided participants with information about the driving situation during each scenario (Table 1). Each participant experienced different combinations of agent forms, genders, and driving scenarios. A Latin square design was used to randomize the order of the conditions for each participant.

Procedure

Before beginning the experiment, participants were asked to read and sign the consent form of the present study, which was approved by the Institutional Review Board (IRB) of Virginia Tech, to understand the purpose of the experiment. The demographic information of the participants was collected to determine their previous driving experience, including autonomous driving. Since level five autonomous driving was involved in the driving scenarios (Figure 1), each participant was informed that they could perform non-driving-related activities while experiencing each agent condition. Participants were asked to wear a pair of eye tracking glasses during experiment. The eye tracking data of each participant were recorded. Participants filled out the questionnaires after completing each condition using the Godspeed (five factors with 24 items, 5-point Likert scale) (Bartneck, Kulić, Croft, & Zoghbi, 2009), RoSAS (three factors with 18 items, 7-point Likert scale) (Carpinella, Wyman, Perez, & Stroessner, 2017), and social presence (five items, 10-point Likert scale) (Lee, Jung, Kim, & Kim, 2006). Participants' preference on those four conditions and explanations on the reason of their preference were recorded. Overall, the experiment took each participant around 45 minutes.

Equipment and Stimuli

A motion-based driving simulator (Nervtech™) was used in the experiment (Figure 1). The driving simulator was equipped with three 48" display monitors, a steering wheel, an adjustable seat, gas and brake pedals, and a surround sound dome. A humanoid robot, Nao (V6 standard edition, height: 22.6 in., width: 10.8 in.), was used as the robotic agent in two of the four conditions. Amazon Polly text-to-speech software was used to create female and male agents' speech clips (https://aws.amazon.com/polly, name: Sallie, gender: female voice, nationality: USA; name: Joey, gender: male voice, nationality: USA). Tobii Pro Glass 2, a wearable eye tracker, was wore by the participants during the experiment.

Table 1: Conversational speech style script

Script list
1. We are entering a new road
2. We are slowing down because of road construction
3. I am sorry for the sudden slow down. A car swerved into traffic
4. We are entering a tunnel
5. Are you okay? A man suddenly popped out onto the road green
6. We are waiting for the signal to turn
7. We are turning left / right
8.We've reached a four-way stop intersection. We are waiting for other cars to go first

Figure 1. Experimental setting (L) and the driving scenario (R)

RESULTS

Two-way repeated measures ANOVA and paired samples t-tests were used to analyze participants' responses to the questionnaires and eye tracking data through Minitab to determine if there are any statistically significant differences. The 0.05 alpha level was applied for the analysis.

The questionnaire results showed that there are significant differences in Likeability, Perceived safety, Competence, and Discomfort associated with the gender of the in-vehicle agent's voice. Figure 2 shows the overall mean rating scores on the Likeability, Perceived safety, Competence, and Discomfort scales for each condition. Results were analyzed with a 2 (Gender) x 2 (Agent Type) repeated measures analysis of variance (ANOVA), which revealed statistically significant differences between gender in the Likeability scale, $F(1, 76) = 5.32$, $p < .05$; the Perceived safety scale, $F(1, 76) = 23.67$, $p < .05$; the Competence scale, $F(1, 76) = 5.22$, $p < .05$; and the Discomfort scale, $F(1, 76) = 6.76$, $p < .05$.

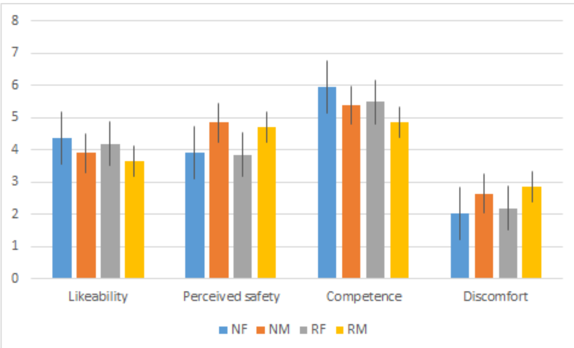

Figure 2. Subjective evaluation results in Likeability, Perceived safety, Competence, and Discomfort

Tables 2 and 3 show the statistical analysis of the responses. The participants rated the non-robotic voice only female agent (NF) as the highest in Likeability and Competence and the lowest in Discomfort (Table 2), which matches the results in the participants' ranking for each condition (Table 4). However, they rated the non-robotic voice only male agent (NM) as the highest in Perceived safety, followed by the robotic male agent (RM). The ranking score was calculated by multiplying the rank and the number of the participants in the rank, divided by the total number of the participants (e.g., (RM: 1*5 + 2*3 +3*6 + 4*6) / 20 = 2.65). Thus, the lower average score means the better rank. As expected, the non-robotic voice only female agent was most preferred, followed by the non-robotic voice only male agent, the robotic male agent, and the robotic female agent.

Table 2: Subjective evaluation results, ( ) = *SD*

Items	Agent Condition			
	NF	NM	RF	RM
Anthropomorphism	3.87 (1.51)	3.60 (1.33)	3.38 (1.21)	3.61 (1.36)
Animacy	3.88 (1.41)	3.50 (1.26)	3.35 (1.19)	3.55 (1.34)
Likeability	4.36 (1.01)	3.91 (0.89)	4.19 (0.92)	3.65 (1.18)
Perceived intelligence	4.50 (0.57)	4.19 (0.53)	4.19 (0.77)	3.85 (1.20)
Perceived safety	3.93 (0.84)	4.85 (0.99)	3.85 (0.86)	4.70 (1.20)
Competence	5.97 (0.87)	5.39 (1.12)	5.50 (1.07)	4.86 (1.92)
Warmth	4.06 (2.01)	3.58 (1.24)	4.18 (1.69)	3.66 (1.50)
Discomfort	2.03 (1.02)	2.65 (1.10)	2.19 (0.97)	2.85 (1.51)
Social Presence	6.46 (2.01)	5.72 (1.10)	6.47 (0.97)	5.61 (1.51)

Table 3: Subjective evaluation statistics

Items	Significance(*p*)		
	Agent condition	Gender	Agent condition X Gender
Anthropomorphism	0.411	0.945	0.392
Animacy	0.398	0.751	0.307
Likeability	0.319	0.024*	0.834
Perceived intelligence	0.062	0.062	0.942
Perceived safety	0.522	0.000*	0.855
Competence	0.064	0.025*	0.901
Warmth	0.784	0.172	0.964
Discomfort	0.46	0.011*	0.946
Social Presence	0.907	0.065	0.899

Table 4: Preference ranking for each agent condition

Preference	Agent Condition			
	RM	NM	RF	NF
1st	5	6	1	8
2nd	3	6	6	5
3rd	6	5	7	2
4th	6	3	6	5
Average Score	2.65	2.25	2.9	2.2
Standard Deviation	1.18	1.07	0.91	1.24

Note: Unit: numbers of participants

Participants' eye tracking data were collected through Tobii Pro Lab. All eye movements were categorized into three areas of interest (AOI): primary driving tasks, two-lane road, and Nao (in robotic agent conditions). Primary driving tasks were identified as glance at rear mirrors, front mirrors, and the instrument panel. The eye tracking data of each AOI's eye movement frequency and gaze duration were arranged for each trial. When participants performed non-driving-related tasks, such as looking away from the monitors and staring at the driving wheels, those eye movements were called distraction time (including Nao gaze duration). The total distraction time was calculated by the following equation: Total distraction time = Total driving time – ($G_{Primary} + G_{road}$), where G is the gaze duration for the AOI's in milliseconds (ms).

The total distraction times between robotic and non-robotic agent conditions were compared. Results showed no significant difference between robotic and non-robotic agent conditions in the total distraction time, $t(40) = 0.63$, $p = 0.530$. Each participant's eye movement frequency, gaze duration, and total distraction time for only robotic agent conditions were also compared, considering both the agent's voice gender and the participants' gender. Results also showed no significant differences based on the two types of gender (Table 5).

Table 5: Eye tracking results per gender in the robot conditions

Items	Nao (F) vs. Nao (M) Significance (*p-value*)		
	All conditions	Female Participants	Male Participants
Frequency	0.416	0.534	0.194
Gaze duration	0.643	0.620	0.188
Total distraction	0.533	0.743	0.531

DISCUSSION AND FUTURE WORK

To evaluate young drivers' perception and preference on in-vehicle agents' characteristics, we compared four agent conditions with two variables, agents' gender and embodiment in level five autonomous vehicles. The results show that the agents' gender plays a more significant role in the participants' perception on specific items, compared to the agents' embodiment. However, the agents' embodiment seems to contribute more to the overall preference ranking.

Our participants perceived the female agents more likeable and more comfortable, which is in line with literature (Nass & Moon, 2000). Interestingly, the participants also perceived significantly more competence from both female agents than from both male agents. Moreover, the voice only female agent's rating scores were numerically higher than the robotic female agent's scores in all of these scales. On the other hand, the male agents were rated higher than the female agents only in the Perceived safety scale. This seems to be related to the traditional stereotype, showing that the male voice is more appropriate for safety-critical environment (Nass & Moon, 2000). For the preference ranking results, the non-robotic voice only female agent was most preferred, followed by the non-robotic voice only male agent. Again, this rank shows that our participants preferred the non-robotic voice only agents over the robotic agents, regardless of their gender.

In a recent study (Lee et al., 2019) using the same robot and the driving simulator as the present study, the results showed that some participants liked the embodied robot agent the most, but other participants liked it the least. Our result seems to support that people still prefer disembodied agents in the vehicle context even though it was a fully autonomous vehicle and there was no concern about drivers' visual distraction. However, from the participants' explanation, we found that the young drivers' attitude towards the presence of the robotic agent was also somewhat contradictory as they thought that the robot is either "scary" or "likeable". Lack of experiences in autonomous driving and the robots per se might be the cause of this inconsistent results to the robotic agent conditions. Or this might stem from the uncanny valley (Mori, MacDorman, & Kageki, 2012) discussed in robotics research. Different people might have a different threshold of the uncanny valley, which can lead to different perceptions on the robot.

It is also interesting that the embodiment of the agent does not influence plausible questionnaire items, such as anthropomorphism, animacy, and social presence scales. It is not clear whether this is because the participants might perceive a sufficient amount of these characteristics even from the voice only conditions or this is because they might not perceive the robot (Nao) we used in our experiment to be sufficiently anthropomorphic. Contradictory to what we hypothesized, the participants thought that the robotic female was the least anthropomorphic and least animated, whereas the non-robotic female agent without a physical body was rated the highest in these categories, which requires further research.

Note that our participants' gender did not influence their preference to the gender of the agent's voice; for example, there was no trend indicated that female drivers would prefer a male/female voice or vice versa. The participants may prefer female voice over male voice in general, because they are already familiar with the female voice used in their in-vehicle navigation devices, as well as at-home voice agents. However, the robotic female voice agent was least preferred in the final ranking. Therefore, to unpack why people prefer the female voice over the male voice, more systematic research is still required.

Interestingly, most drivers addressed the gender factor more than the embodiment factor as the reason for their preference ranking in the present study. The male voice was thought to be "natural", "confident", "organic" but "aggressive", while the female voice was perceived to be "warm", "friendly", "trustworthy" but "artificial".

The presence of the in-vehicle agents' embodiment did not have a big impact on the total distraction time in the presented study, which might imply that having a robotic agent is unlikely to cause more visual distraction issues in driving. The percentage of the distraction time that participants were distracted by the robotic agent was only 1.034% in the robotic agent conditions. In the voice only conditions, drivers were also distracted by performing other activities, such as looking at their phones. The gender of the agent voice or participants also did not affect participants' eye gaze patterns. Most participants were driving carefully by constantly looking at the road and mirrors although they understood they were in fully autonomous driving scenarios, which might be the results of the lack of autonomous driving experience and trust in the autonomous system.

In the future, this experiment could be expanded to include more animated robot agents. In the present study, the robot agents stood on their feet and stayed completely still, but perhaps with hand gestures or facial expressions, the driver's trust might increase (Zihsler et al., 2016). Thus, we plan to conduct the next study to see whether the robots' movement with other characteristics can change the passengers' perceptions of an agent in the autonomous vehicle context.

ACKNOWLEDGEMENT

This work was supported by the "Northrop-Grumman Undergraduate Research Experience Award".

REFERENCES

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics, 1(1), 71-81.

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017, March). The Robotic Social Attributes Scale (RoSAS) Development and Validation. In Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction (pp. 254-262).

Cramer, H., Evers, V., Kemper, N., & Wielinga, B. (2008, December). Effects of autonomy, traffic conditions and driver personality traits on attitudes and trust towards in-vehicle agents. In 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Vol. 3, pp. 477-482). IEEE.

Crowelly, C. R., Villanoy, M., Scheutzz, M., & Schermerhornz, P. (2009, October). Gendered voice and robot entities: perceptions and reactions of male and female subjects. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 3735-3741). IEEE.

Hock, P., Kraus, J., Walch, M., Lang, N., & Baumann, M. (2016, October). Elaborating feedback strategies for maintaining automation in highly automated driving. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 105-112).

Jeon, M., Walker, B. N., & Gable, T. M. (2015). The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. Applied ergonomics, 50, 185-199.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. International Journal on Interactive Design and Manufacturing (IJIDeM), 9(4), 269-275.

Large, D. R., & Burnett, G. E. (2014). The effect of different navigation voices on trust and attention while using in-vehicle navigation systems. Journal of safety research, 49, 69-e1.

Large, D. R., Burnett, G., & Clark, L. (2019, September). Lessons from Oz: design guidelines for automotive conversational user interfaces. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (pp. 335-340).

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. International journal of human-computer studies, 64(10), 962-973.

Lee, S. C., Sanghavi, H., Ko, S., & Jeon, M. (2019). Autonomous driving with an agent: speech style and embodiment. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (pp. 209-214).

Lee, S., Ratan, R., & Park, T. (2019). The Voice Makes the Car: Enhancing Autonomous Vehicle Perceptions and Adoption Intention through Voice Agent Gender and Style. Multimodal Technologies and Interaction, 3(1), 20.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. IEEE Robotics & Automation Magazine, 19(2), 98-100.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of social issues, 56(1), 81-103.

Williams, K. J., Peters, J. C., & Breazeal, C. L. (2013, June). Towards leveraging the driver's mobile device for an intelligent, sociable in-car robotic assistant. In 2013 IEEE intelligent vehicles symposium (IV) (pp. 369-376). IEEE.

Zihsler, J., Hock, P., Walch, M., Dzuba, K., Schwager, D., Szauer, P., & Rukzio, E. (2016, October). Carvatar: increasing trust in highly-automated driving through social cues. In Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 9-14). ACM.