Robot-Like In-Vehicle Agent for a Level 3 Automated Vehicle

XINGJIAN ZENG, Eindhoven University of Technology, The Netherlands

PAVLO BAZILINSKYY, Eindhoven University of Technology, The Netherlands

With the rapid development of automotive technology and artificial intelligence, in-vehicle agents have great potential to solve the challenges of explaining the status of the system and the intentions of an automated vehicle. A robot-like in-vehicle agent was designed and developed to explore the in-vehicle agent communicating through gestures and facial expressions with a driver in a SAE Level 3 automated vehicle. An experiment with 12 participants was conducted to evaluate the prototype. The results showed that both interactions of facial expressions and gestures can reduce workload and increase usefulness and satisfaction. However, gestures seem to be more functional and preferred by the driver while facial expressions seem to be more emotional and preferred by passengers. Furthermore, gestures are easier to notice but difficult to understand independently, and facial expressions are hard to notice but more attractive.

CCS Concepts: • **Human-centered computing → Empirical studies in interaction design**; **HCI theory, concepts and models**; **Empirical studies in HCI**; *Usability testing*; *Laboratory experiments.*

Additional Key Words and Phrases: In-Vehicle Agent, Robot-Like Agent, Gesture, Facial Expression, Voice Interaction

ACM Reference Format:

Xingjian Zeng and Pavlo Bazilinskyy. 2018. Robot-Like In-Vehicle Agent for a Level 3 Automated Vehicle. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

With the development of automotive technology and artificial intelligence, in-vehicle agents (IVAs) have emerged as a transformative innovation for intelligent transportation systems. These agents are often embodied as driving assistants and are integrated into the driving system. IVAs are classified as voice agents, virtual agents, and physical agents. The purpose of integrating IVAs of any type is to help the driver with driving tasks and improve the driving experience [29].

1.1 In-Vehicle Agents in Manual and Automated Vehicles

In the manual driving context, the IVAs can not only help with driving-related tasks like vehicle-to-vehicle communication (both vehicles need to install IVA) [16], or non-driving related tasks like comfort children to reduce distractions for the driver [13], but also minimize driver's distraction by decreasing the number of directed utterances with a set of robots [22], reduce driver's fatigue through social communication [26], and mitigate drivers' negative affective status through giving positive comments about the situation [31].

Authors' Contact Information: Xingjian Zeng, esse.zeng009@gmail.com, Eindhoven University of Technology, Eindhoven, The Netherlands; Pavlo Bazilinskyy, p.bazilinskyy@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands.

Manuscript submitted to ACM

IVAs can explain the system status and intentions of an automated vehicle (AV) [18, 24, 30, 36, 45]. The user interface (UI) of IVAs can be a voice UI [24], a visual UI [18], or a physical UI [9]. Lee and Jeon [29] suggest that physical agents aid in better driving behaviour and overall experience, especially in the context of automated driving (AD). Zihsler et al. [45] and Chakravarthi et al. [9] showed that physical agents with facial expressions and gestures, respectively, can increase trust in AVs.

In an AD context, IVAs perform better in improving overall experience [29], such as explaining the status of the system with animation of a chauffeur avatar and a world in miniature [18], or using the "How + Why message" to lead better driving performance [24]. On the other hand, an IVA can serve as a companion by adopting a conversational dialogue style, using emotional tones and first-person language, which fosters a 'human-agent relationship' with the driver [27], giving the driver a sense of a "human-agent relationship". Furthermore, IVAs can increase trust and acceptance in AD using social cues and anthropomorphism to translate the state of the vehicle into human behaviour and expressions, which can be intuitively interpreted by the driver [45].

Physical agents can be divided into consumer products on the market and prototypes in research. Physical agents do not appear to be popular in Europe. However, Chinese and Japanese companies have already published a few physical agent products on the market. Nomi [32], Xiaodu [6], and Mochi [11] all have a geometric appearance and digital screens for facial expressions. Nomi can access the CAN bus. However, these products do not use gestures and act as virtual agents to improve the driving experience. Intelligent Puppet [33] is a comfort robot for babies rather than helping drivers with driving tasks. Kirobo Mini [10], RoBoHoN [2], and NAO [19] are usually applied as humanoid agents in IVA studies [25, 30, 38, 40, 41]. However, Kirobo Mini and RoBoHoN are initially companion robots, and NAO is used for coding education, which means that all of these humanoid robots are not designed specifically for driving scenarios. The Affective Intelligent Driving Agent (AIDA) [43, 44] is the first physical agent designed especially for driving scenarios. AIDA can act as a human passenger, communicate with the driver, and help the driver with some tasks. The robot human-machine interface (RHMI) [38] can use eye colour and body movements to warn the driver of a take-over request 5 seconds before it is issued. Carvatar [45] is another physical agent aimed at AD scenarios, using facial expressions to convey information and improve trust.

1.2 Interaction with In-Vehicle Agents

Voice interaction is a common communication method for IVAs in SAE Level 3 AD vehicles due to its minimal visual distraction [41]. Research on IVA voice interaction, including speech emotion and gender, indicates that no single voice suits all listeners and situations [21]. Lee et al. found that voice agents aligning with social role stereotypes (informative male and social female) enhance perceived ease of use (PEU) and perceived usefulness (PU) [28]. Jeon et al. showed the effectiveness of an in-vehicle software agent in mitigating effects on driver situation awareness and performance [20]. Ruijten et al. showed that conversational interfaces are more trusted, liked, anthropomorphised, and perceived as more intelligent than graphical UIs [36].

As IVAs evolve from voice-only agents to physical agents, interactions become more complex. Both virtual and physical agents can engage in visual interactions, with virtual agents being 2D or 3D characters, and physical agents having a physical appearance and facial expressions [14, 18, 23, 38, 43, 45]. However, the interesting thing is that except for AIDA published in 2014, other concepts are all in the context of AD.

Gestures are a unique feature of physical agents compared to other agents. The robot developed by Srivatsan et al. [9] shows that robotic objects are a promising technology to improve passengers' experience in AVs. RHMI developed

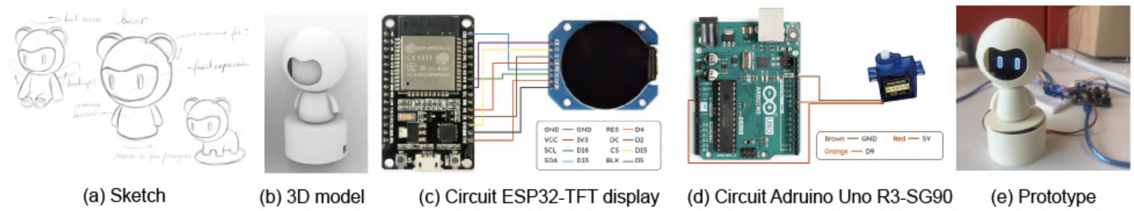


Fig. 1. The design concept of the robot-like IVA.

by Tanabe et al. [38] can adjust the turning angle, speed, and opening angle of the lid to inform different levels of emergency: normal state, unstable state and suspended state.

Social interactions, such as small talk, significantly increase driver trust compared to voice interactions alone [25]. Although robot agents can be visually distracting, yet increase trust, voice agents are preferred in low-speed situations [41]. Drivers have mixed attitudes towards conversational robot agents [30]. Both voice and robot agents improve likability and perceived warmth, with voice agents better at anthropomorphism, and robot agents offering greater competence and lower workload [40].

IVAs (especially physical IVAs) have significant potential to help with driving tasks and improve the driving experience, as well as a solution to the challenges raised in the context of AD. There is a research gap in exploring the advantages and challenges of combining facial expressions and gestures with voice interaction in physical IVAs. This project explores this area. Two research questions were defined: **RQ1:** *How to develop a robot-like IVA for the SAE Level 3 AD scenario?* and **RQ2:** *What are the advantages and challenges of comparing gestures and voice interaction with facial expressions and voice interaction in SAE Level 3 AD scenarios?* In the context of this work, AD is assumed to be SAE Level 3 [37]. In this project, a robot-like IVA was designed and developed to answer these two questions.

2 Interview and Design of In-Vehicle Agent

To understand attitudes and expectations about IVA and driving behaviour in Asian countries and Europe, five participants (5 males, M = 28.8, SD = 3.42) were invited to an interview (see supplementary materials/interview). Four of the participants had experience driving in Europe and one of the participants had experience driving in both Japan and Europe. The results showed that long-distance driving can be boring and can cause us to get distracted. Although only one of them had heard of IVA (Nomi of NIO), others were interested in the concept.

The sketch (Figure 1 (a)) presents three modalities, and the middle one was developed further. The face shows expressions and the body rotates to present different gestures according to seven highway scenarios [8] (Figure **??**). The IVA prototype is not a robot that acts independently, but a physical form of the whole driving assistant system [30]. Figure 1 (b) shows the 3D model created in Rhino 8 (STL files are in the supplementary material). The shell is 3D printed and contains a round 1.28 inch IPS-TFT display (240*240 pixels, IPS GC9A01) inside the round head (r=31mm) connected to ESP32 (Figure 1 (c)). The gestures are driven by an SG90 servo motor inside the stand connected to Arduino Uno R3 (Figure 1 (d)). No speaker was installed in the prototype because, in the real vehicle, the sound comes from the vehicle's audio system, rather than a physical robot. Figure 1 (e) shows the whole prototype.

The TFT display was connected to an ESP32 board and controlled by the Arduino IDE [3] (v.2.3.2) on the Apple Macbook A2442. See the supplementary material for the code. Five facial expressions (normal, smile, excited, realising, sad) were designed, shown in Table 1.

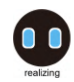
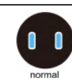
Scenarios	IVA gestures (GV)	Dialogues (FV & GV)	Facial expressions (FV)	Dialogues (B)
Greeting	Greeting (gesture)	"Welcome! My name is Eva. Shall we start our trip?" (Driver: Yes) "Here we go!"		
Enter highway	Situation reporting	"We will enter the highway ahead."		"Enter the highway ahead."
Speed limit and speed report		"The speed limit is 90, and right now we are at 87."		"The speed limit ahead is 90, the current speed is 87."
Overtaking	Situation reporting; overtaking (gesture)	"The front car is driving too slow, shall we overtake it?" (Driver: Yes) "Let's do this!" (After overtaking) "WOW, nice!"		
Line changing (construction)	Situation reporting	"Seems there is a construction ahead, we need to change line."		
Congestion	Situation reporting	"Seems there is a traffic jam, we need to slow down."		
Exit highway	Situation reporting	"We will exit the highway ahead."		"Exit the highway ahead."

Table 1. IVA behaviour (gestures, facial expressions, and dialogues) in seven highway scenarios.

To enable the Arduino IDE to run on ESP32, the Arduino core [12] for ESP32 was installed. Libraries Adafruit GC9A01A [1] (v.1.1.1), Adafruit GFX [5] (v.1.11.9), and TFT_eSPI [4] (v.2.5.43) were installed in the Arduino IDE to run the code on the TFT display.

The SG90 servo motor was connected to an Arduino Uno board and controlled by Arduino IDE on the laptop. The library Servo (v. 1.2.1) was installed in Arduino IDE to run the code on the servo motor. Three gestures were designed in this project according to the scenarios [8] in Table 1: (1) greeting: turn to the driver (starting position), then turn front (100/s, clockwise) to check the surroundings (66.7/s, clockwise and counterclockwise) and turn back to the driver (100/s, counterclockwise); (2) situation reporting: turn front (100/s, clockwise) and turn to the driver (100/s, counterclockwise); (3) overtaking after got permission: turn front and rotate to face the vehicle be overtaken (100/s, clockwise, only 30 degrees with SG90), then turn back to the driver (100/s, counterclockwise).

3 Method of Experiment

An experiment was conducted to evaluate the design. The experiment had three groups: Group *B* (baseline), Group *FV* (facial expressions and voice), and Group *GV* (gestures and voice). The behaviour of IVA in different groups is shown in

Fig. 2. Experimental setup.

Table 1. *B* as a baseline only had robotic voice interaction, making it sound like a conventional in-vehicle navigation system and convey limited information. Thus, *B* presented the Tesla Full Self-Driving (supervised) driving-assistance system. All audio was generated from PlayHT [34] and was edited as another soundtrack in a 4.5-minute video recorded in GTA V. The study was approved by the Ethical Review Board of Eindhoven University of Technology and the participants gave their informed consent to use their data.

The videos of scenarios were recorded in the GTA V video game [35] running on a Windows PC according to Table 1, and the highway route is chosen from downtown to Beeker's Garage. To get an inside view of AD, two mods were applied: (1) Dynamic Vehicle First Person Camera Mod [15], allowing the camera inside the vehicle to get the driver's perspective and (2) Enhanced Native Trainer Mod [42], which makes characters invisible (i.e., no hands holding the steering wheel were visible, providing a sense of driving in an AV).

A total of 12 participants (age: M = 27.42, SD = 2.11; 7 females and 5 males) from Eindhoven University of Technology joined the user test through the user test link posted on the social media platform. And no financial inducement was offered for the user test. All participants were over 18 years of age and had a driver's licence (issued in different countries). Three participants had experience of driving with Tesla autopilot. Figure 2 shows the experimental setup. A screen (RCA RS32F3), headphones (Sennheiser MOMENTUM 4), and the robot-like IVA prototype were connected to the laptop (Apple Macbook A2442). For each participant, the lowest point of the prototype was adjusted by stacking books (5.5 cm from the desk surface) until the participant could see the whole TFT display. The position of the robot-like IVA prototype is settled on the front right of the participant, corresponding to the position above the dashboard in a real car.

The author briefly introduced the background information about SAE Level 3 AD to the participants. The participants then took a seat and had three groups of tasks to complete: *B*, *FV* and *GV*. *B* was a baseline and was always first, but for

Table 2. Results from the NASA TLX scale [17].

	B M(SD)	FV M(SD)	GV M(SD)
Mental demand (%)	34 (23)	24 (26)	20 (18)
Physical demand (%)	33 (27)	28 (28)	11 (12)
Temporal demand (%)	21 (18)	21 (22)	15 (13)
Performance (%)	34 (27)	22 (24)	17 (13)
Effort (%)	28 (25)	22 (25)	25 (21)
Frustration (%)	48 (28)	19 (16)	19 (15)
Average (%)	33 (21)	23 (24)	18 (14)

Note: B=Baseline, FV=Facial expressions and voice, GV=Gestures and voice.

half of the participants, the sequence of *FV* and *GV* was switched. The prototype was controlled by the author during the experiment. After each group of tasks, participants were asked to fill in the NASA Task Load Index scale [17] to measure workload and the acceptance scale [39] to measure overall experience on an iPad. Finally, a semi-structured interview was conducted to collect the user test experience. During each group of tasks, participants were asked to imagine themselves in the SAE Level 3 AV and do their daily work as a secondary task (either replying to messages, watching videos on a mobile phone/iPad, or reading a book). They were allowed to look up and check the situation at any time. If they felt that they wanted to take over the control immediately, they were asked to inform the author about it.

The experiment interview was analysed through thematic analysis. The themes were generated after coding the data in the transcription.

4 Results of Experiment

The workload scores (Table 2) of *FV* (M=23, SD=24) and *GV* (M=18, SD=14) were both less than *B* (M=33, SD=21), and *GV* had the lowest workload score among the three groups. For the dimension of Physical demand, the workload score of *GV* (M=20, SD=18) was around half of *B* (M=34, SD=23). When comparing *FV* and *GV* separately, *GV* had lower workload scores than *FV* in the other five dimensions, except Effort, and lower standard deviation values in all dimensions. However, *GV* (M=25, SD=21) scored a higher workload score than *FV* (M=22, SD=25) in the Effort dimension. And *FV* had almost the same Temporal demand as *B*.

Table 3 shows the usefulness and satisfaction scores of each participant for *B*, *FV*, and *GV*. Both *FV* and GV had higher overall usefulness and satisfaction scores than *B*, and *FV* scored the highest both in usefulness and satisfaction among all. Except for Annoying-Nice, *FV* obtained higher or equal scores compared to *GV*, as well as lower or equal standard deviation for all dimensions. Furthermore, *GV* received scores far lower (over 0.5) than *FV* in the Unpleasant-Pleasant and Sleep-inducing-Raising Alertness dimensions.

According to the results of the experiment interview (see supplementary material), 7 participants preferred *GV* (gestures and voice), 5 participants preferred *FV* (facial expressions and voice) and no one preferred *B* (baseline). The reason for choosing gestures could be summarised as follows: (1) they have better perception than facial expressions (P1, P2, P3, P6, P8); (2) gestures move before the voice conveys information, providing more time to get out of the work and concentrate on the road situation (P1, P11); (3) facial expressions make people distracted (P8); (4) understanding facial expressions needs time (P1, P2, P8). The others who chose facial expressions suggested: (1) facial expressions

Table 3. Results from the acceptance scale [39].

Negative (-2)	Positive (+2)	B M(SD)	FV M(SD)	GV M(SD)
Useless	Useful	1.00 (1.04)	1.33 (0.78)	1.08 (1.16)
Unpleasant	Pleasant	0.67 (0.98)	1.17 (0.39)	1.08 (0.67)
Bad	Good	0.83 (0.94)	1.25 (0.45)	1.08 (0.79)
Annoying	Nice	1.08 (0.67)	1.25 (0.62)	1.33 (0.65)
Superfluous	Effective	0.92 (1.00)	1.25 (0.75)	1.25 (0.75)
Irritating	Likeable	0.67 (0.78)	1.08 (0.79)	0.83 (1.03)
Worthless	Assisting	1.00 (0.95)	1.17 (0.58)	0.92 (0.90)
Undesirable	Desirable	1.00 (0.60)	1.17 (0.83)	0.83 (1.19)
Sleep-inducing	Raising Alertness	-0.33 (0.89)	0.75 (0.75)	0.17 (1.27)
Overall usefulness score		0.68 (0.72)	1.15 (0.48)	0.90 (0.82)
Overall satisfaction score		0.85 (0.61)	1.17 (0.59)	1.02 (0.79)

Note: B=Baseline, FV=Facial expressions and voice, GV=Gestures and voice.

can provide more emotional support than gestures (P2, P4, P5, P10); (2) facial expressions do not have the noise of rotating (P7); (3) cannot understand the meaning of gestures (P12). Details of the interview thematic analysis are shown in Table4.pdf (see supplementary materials). In summary, four main themes were concluded: (1) perception (not just visual), efficiency, trust issues, and emotional support. Although *B* could be perceived by the user, it still needs to provide more information to explain the current status. This is also related to the trust of the system. (2) *FV* works better in emotional support than *GV*, however, it is difficult for users to notice and understand facial expressions in a short time. (3) *GV* has a higher perception than *FV* and *B*, but it is hard to understand the meaning and is a little boring. (4) All groups have trust concerns.

5 Discussion

We developed a robot-like IVA capable of voice interactions for an SAE Level 3 AV with five facial expressions and three gestures and evaluated it in an experiment.

There were notable results. (1) Both interactions of facial expressions and gestures can reduce workload in an SAE Level 3 AD scenario, and increase the usefulness and satisfaction of the driver. Both *FV* and *GV* were effective in reducing workload, and the effect of *GV* was better than *FV*. Furthermore, *GV* greatly reduced the workload of Physical demand. The reason may be that the gestures were always triggered before voice interaction and can be easily noticed by the participants, which leaves some time for them to get out of their work and focus on the road. The facial expression was shown at the same time as the voice was played, which may have caused the participants to check both road situations and expressions, resulting in a higher score of Temporal demand. As for the great reduction in Frustration, probably because IVA provide a sense of companionship, either through voice interaction, facial expressions or gestures. On the other hand, the acceptance scale shows better results in *FV* than GV, even though both groups can improve usefulness and satisfaction. This may indicate that gestures work better in reducing workload (functionally) and facial expressions work better in enhancing usefulness and satisfaction, providing more affective support (emotionally). The results of the thematic analysis of the interview showed that: (1) Participants who preferred gestures also indicated that gestures could remind them that something was going to happen before the voice informed them about it; while participants who preferred facial expressions argued that expressions were more intuitive (P5), comforting (P2), and cute (P4). (2) Although voice interaction is a more efficient way for an IVA to convey information, it is still not enough. The challenges could be missing information (P1) or needing more explanation (P3, P5, P9), and gestures can provide more time to get out of the secondary task and concentrate on the traffic situation. (3) Gestures appear to be more functional and preferred by the driver while facial expressions are more emotional and preferred by passengers. (4) Gestures are easier to notice but difficult to understand independently, and facial expressions are the opposite. (5) Users' concerns about physical IVA could be classified into four aspects: perception (not just visual), efficiency, trust issues, and emotional support.

After the experiment, two engineers from Nissan Co. were interviewed to discuss the project from the point of view of the vehicle manufacturer. They noted that installing physical IVAs in vehicles is challenging, especially if connected to the CAN bus. Privacy concerns arise if an IVA accesses vehicle functions, and the IVA's position must be considered to prevent injuries during airbag deployment.

The secondary task was not defined in the experiment because people have different driving habits. However, some people would look at the view outside while others read a paper. They had different levels of commitment to the secondary task, which may have led to errors. Different secondary tasks and different sitting postures also influence the participants' field of view. That is why some participants could easily notice the IVA, while others could not. Different driving modes of IVA could be defined to suit different workloads of secondary tasks. The experiment was conducted with a video, rather than in a real vehicle. Two participants (P10, P12) mentioned that they may have acted differently if they had felt the acceleration and deceleration of the vehicle. Since exposure to each group was only 4.5 minutes per participant, it is hard to predict if participants would get bored or fall asleep in case of extended duration. Some participants were curious about the video and always looked up in all groups. Even switching FV and GV to reduce the error, they knew what would happen when they tested FV and GV.

For future work, the combination of facial expressions and gestures could enhance the concept. Designing more intuitive gestures and 3D facial expressions is also recommended. Integrating IVA with other human-machine interfaces in the vehicle could make IVA the manager of all in-vehicle communication. Since the study method is Wizard of Oz, user tests in a working system are suggested to have more precise results. Different driving modes of IVA (when the user is working on laptop/reading a book/talking with others/driving alone, etc.) should be defined based on the feedback from the participants. Different appearances and sizes of IVA might also make a difference to the workload, satisfaction, and usability. Furthermore, different driving modes of IVA have different degrees of impact on the SAE level [37], which level of automation benefits the most would also be interesting to explore in the future. Additionally, exploring IVA's potential in interacting with vulnerable road users (VRUs) such as cyclists and pedestrians is suggested. The IVA was placed above the dashboard, where it can also be visible to people outside. So, it could communicate information to them by gestures and facial expressions, helping them in interactions with VRUs [7]. However, physical IVA will have a significant impact on safety in a car accident, so a protection chamber on the dashboard will be considered in future work.

6 Supplementary material

Interview, STL files, analysis and Arduino code, materials used in the experiment, and raw data can be found at: https://www.dropbox.com/scl/fo/8xz3ok1s4zsagf7nytky5/AJQPehMbzmQAZ8ncz3LqjfQ?rlkey=25dct1vyd3dzqyxyvihy34h4u&st=zu8ty1mn.

References

[1] 2024. *Adafruit GC9A01A*. https://www.arduino.cc/reference/en/libraries/adafruit-gc9a01a/

[2] 2024. *RoBoHoN*. https://robohon.com/

[3] 2024. *Software*. https://www.arduino.cc/en/software

[4] 2024. *TFT_eSPI*. https://www.arduino.cc/reference/en/libraries/tft_espi/

[5] Adafruit. 2024. *GitHub - ADAfruit/Adafruit-GFX-Library: Adafruit GFX Graphics Core Arduino Library, This is the "Core" class that all our other graphics libraries derive from.* https://github.com/adafruit/Adafruit-GFX-Library

[6] Apollo. 2020. *XiaoduOS*. https://developer.apollo.auto/platform/dueros_cn.html

[7] Jackie Ayoub, Feng Zhou, Shan Bao, and X Jessie Yang. 2019. From manual driving to automated driving: A review of 10 years of autoui. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*. 70–90.

[8] Matthias Beggiato, Franziska Hartwich, Katja Schleinitz, Josef Krems, Ina Othersen, and Ina Petermann-Stock. 2015. What would drivers like to know during automated driving? Information needs at different levels of automation.. In *7. Tagung Fahrerassistenzsysteme*.

[9] Srivatsan Chakravarthi Kumaran, Toam Bechor, and Hadas Erel. 2024. A Social Approach for Autonomous Vehicles: A Robotic Object to Enhance Passengers' Sense of Safety and Trust. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 86–95.

[10] Toyota Motor Corporation. 2024. *Toyota launches sales of "Kirobo Mini" nationwide | Toyota Motor Corporation official global website.* https://global.toyota/en/detail/19880995

[11] Dasai. 2024. *MEET MOCHI*. https://dasai.com.au/pages/meet-mochi

[12] espressif. 2024. *GitHub - espressif/arduino-esp32: Arduino core for the ESP32.* https://github.com/espressif/arduino-esp32

[13] Michal Gordon and Cynthia Breazeal. 2015. Designing a virtual assistant for in-car child entertainment. In *Proceedings of the 14th International Conference on Interaction Design and Children*. 359–362.

[14] Yijie Guo, Hanyang Hu, Zhuoran Zhai, Yao Lu, Xuezhu Wang, Yuanling Feng, Zhihao Yao, Zhenhan Huang, Yuan Yao, and Haipeng Mi. 2023. Designing Future In-Vehicle Assistants: Insights from User Imaginations and Experiences. In *Proceedings of the Eleventh International Symposium of Chinese CHI*. 508–520.

[15] V3ry H1gh. 2023. *How to install dynamic Vehicle First person Camera - GTA V Mods.* https://www.youtube.com/watch?v=jwxgmAHtwIY

[16] Toshiyuki Hagiya and Kazunari Nawa. 2020. Acceptability evaluation of inter-driver interaction system via a driving agent using vehicle-to-vehicle communication. In *Proceedings of the 11th Augmented Human International Conference*. 1–8.

[17] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[18] Renate Häuslschmid, Max von Buelow, Bastian Pfleging, and Andreas Butz. 2017. Supportingtrust in autonomous driving. In *Proceedings of the 22nd international conference on intelligent user interfaces*. 319–329.

[19] SoftBank Robotics America Inc. 2024. *NAO*. https://us.softbankrobotics.com/nao

[20] Myounghoon Jeon, Bruce N Walker, and Thomas M Gable. 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 (2015), 185–199.

[21] Marie Jonsson and Nils Dahlbäck. 2009. Impact of voice variation in speech-based in-vehicle systems on attitude and driving behaviour. *Human Factors and Ergonomics Society Europe Chapter (HFES)* (2009).

[22] Nihan Karatas, Shintaro Tamura, Momoko Fushiki, and Michio Okada. 2018. Multi-party conversation of driving agents: the effects of overhearing information on lifelikeness and distraction. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. 84–91.

[23] Nihan Karatas, Soshi Yoshikawa, Shintaro Tamura, Sho Otaki, Ryuji Funayama, and Michio Okada. 2017. Sociable driving agents to maintain driver's attention in autonomous driving. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 143–149.

[24] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 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 (2015), 269–275.

[25] Johannes Maria Kraus, Florian Nothdurft, Philipp Hock, David Scholz, Wolfgang Minker, and Martin Baumann. 2016. Human after all: Effects of mere presence and social interaction of a humanoid robot as a co-driver in automated driving. In *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*. 129–134.

[26] David R Large, Gary Burnett, Vicki Antrobus, and Lee Skrypchuk. 2018. Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task-related fatigue. *IET Intelligent Transport Systems* 12, 6 (2018), 420–426.

[27] David R Large, Gary Burnett, and Leigh Clark. 2019. 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*. 335–340.

[28] Sanguk Lee, Rabindra Ratan, and Taiwoo Park. 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 (2019), 20.

[29] Seul Chan Lee and Myounghoon Jeon. 2022. A systematic review of functions and design features of in-vehicle agents. *International Journal of Human-Computer Studies* 165 (2022), 102864.

[30] Seul Chan Lee, Harsh Sanghavi, Sangjin Ko, and Myounghoon Jeon. 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*. 209–214.

[31] Shuling Li, Tingru Zhang, Wei Zhang, Na Liu, and Gaoyan Lyu. 2020. Effects of speech-based intervention with positive comments on reduction of driver's anger state and perceived workload, and improvement of driving performance. *Applied Ergonomics* 86 (2020), 103098.

[32] NIO. 2017. *NOMI*. https://www.youtube.com/watch?v=SAZ2Dd9lrVc

[33] NISSAN. 2024. *INTELLIGENT PUPPET*. https://www.nissan.co.jp/SP/INTELLIGENTPUPPET/

[34] PlayHT. 2024. . https://play.ht/

[35] Games Rockstar. 2024. *Grand Theft Auto V - Rockstar Games*. https://www.rockstargames.com/zh/gta-v

[36] Peter AM Ruijten, Jacques MB Terken, and Sanjeev N Chandramouli. 2018. Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction* 2, 4 (2018), 62.

[37] SAE. 2021. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. https://www.sae.org/standards/content/j3016_202104

[38] Hiroko Tanabe, Yuki Yoshihara, Nihan Karatas, Kazuhiro Fujikake, Takahiro Tanaka, Shuhei Takeuchi, Tsuneyuki Yamamoto, Makoto Harazawa, and Naoki Kamiya. 2022. Effects of a Robot Human-Machine Interface on Emergency Steering Control and Prefrontal Cortex Activation in Automatic Driving. In *International Conference on Human-Computer Interaction*. Springer, 108–123.

[39] Jinke D Van Der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (1997), 1–10.

[40] Manhua Wang, Seul Chan Lee, Harsh Kamalesh Sanghavi, Megan Eskew, Bo Zhou, and Myounghoon Jeon. 2021. In-vehicle intelligent agents in fully autonomous driving: The effects of speech style and embodiment together and separately. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 247–254.

[41] Manhua Wang, Seul Chan Lee, Genevieve Montavon, Jiakang Qin, and Myounghoon Jeon. 2022. Conversational voice agents are preferred and Lead to better driving performance in conditionally automated vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 86–95.

[42] Willez. 2020. *GTA 5 - How to install Enhanced Native Trainer (2020)*. https://www.youtube.com/watch?v=UHHXTh0Xdow

[43] Kenton Williams, José Acevedo Flores, and Joshua Peters. 2014. Affective robot influence on driver adherence to safety, cognitive load reduction and sociability. In *Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications*. 1–8.

[44] Kenton J Williams, Joshua C Peters, and Cynthia L Breazeal. 2013. Towards leveraging the driver's mobile device for an intelligent, sociable in-car robotic assistant. In *2013 IEEE intelligent vehicles symposium (IV)*. IEEE, 369–376.

[45] Jens Zihsler, Philipp Hock, Marcel Walch, Kirill Dzuba, Denis Schwager, Patrick Szauer, and Enrico Rukzio. 2016. 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*. 9–14.