Robot like in-vehicle agent for a level 3 automated car

Xingjian Zeng, Md Shadab Alam*, Pavlo Bazilinskyy

Abstract— **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 facial expression and gesture interactions can reduce workload (NASA TLX mean scores: baseline = 33%, facial expressions = 23%, gestures = 18%) and increase usefulness and satisfaction. In general, gestures were preferred by 7 of 12 participants due to their practicality and earlier signal timing, while facial expressions were preferred by the remaining 5 participants for their emotional and aesthetic appeal. These findings highlight the distinct advantages of gesture-based interactions for functional communication and facial expressions for emotional connection in automated driving scenarios.**

I. INTRODUCTION

With the development of automotive technology and artificial intelligence (AI), 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 [1].

In the manual driving context, IVAs can not only help with driving-related tasks such as vehicle-to-vehicle communication [2], or non-driving-related tasks such as comforting children to reduce distractions [3]. These agents help minimise distraction from the driver by decreasing the need for direct verbal communication, reduce driver fatigue through social interactions [4], and alleviate negative emotions by providing positive comments [5]. In automated driving scenarios, IVAs are particularly effective in enhancing the user experience by clearly explaining the vehicle's status and intentions [6], [7], [8] through interfaces such as voice [9], visual [6], or physical embodiments [10]. Physical agents, especially those who display facial expressions and gestures, have demonstrated greater driver trust and a better overall driving experience [8].

Physical IVAs, though not yet widely adopted in Europe, are already commercially available from Chinese and Japanese companies. Products such as Nomi [11] use geometric designs and digital screens to show facial expressions but lack gestural interaction. Other physical IVAs, such as Intelligent Puppet, primarily focus on comfort rather than driving tasks. Humanoid robots such as Kirobo Mini [12] are used in IVA research [7], [13], although not originally developed for driving purposes. The Affective Intelligent Driving Agent (AIDA) [14], explicitly designed for driving scenarios, behaves as a human passenger and assists drivers with specific tasks, while the robot human-machine interface (RHMI) [13] employs gestures such as change of eye colour and body movements to warn drivers before critical takeover requests.

Voice interaction is common for IVAs in SAE Level 3 AD vehicles due to minimal visual distraction [15]. Research indicates that a single voice does not suit all listeners or situations [16]. Voice agents aligned with social stereotypes (informative male and female social) improve ease of use and usefulness [17]. Conversational interfaces are found to be more trusted, liked, and perceived as intelli-

*Corresponding Author
Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands; esse.zeng009@gmail.com, m.s.alam@tue.nl, p.bazilinskyy@tue.nl

gent compared to graphical UIs [18].

Interactions become more complex as IVAs evolve from voice-only to physical agents. Both virtual and physical agents can interact visually, with virtual agents as 2D or 3D characters, and physical agents that feature facial expressions and physical forms [6], [8], [13]. Except for AIDA (2014), most IVA concepts focus on automated driving contexts.

Gestures uniquely differentiate physical IVAs. Robotic objects have shown promise in enhancing passengers' experiences [10]. For instance, the RHMI developed by Tanabe et al. [13] communicates varying emergency levels by adjusting its turning angle, speed, and lid opening.

Social interactions, such as small talk, significantly increase driver trust compared to voice interactions [19]. Although robot agents can be visually distracting, yet increase trust, voice agents are preferred in low-speed situations [15]. Drivers have mixed attitudes towards conversational robot agents [7]. Both voice and robot agents improve likeability and perceived warmth, with voice agents better at anthropomorphism, and robot agents offering greater competence and lower workload [20].

A. *Aim of the study*

Physical IVAs offer significant potential to help drivers and improve driving experiences, particularly in AD scenarios. This study addresses the research gap on the integration of facial expressions and gestures with voice interactions in physical IVAs. Two primary research questions are explored: **RQ1:** How can a robot-like IVA be developed for SAE Level 3 AD [21] scenarios? **RQ2:** What are the comparative advantages and challenges of using gestures with voice interactions versus facial expressions with voice interactions in SAE Level 3 AD scenarios?

II. INTERVIEW AND DESIGN OF IN-VEHICLE AGENT

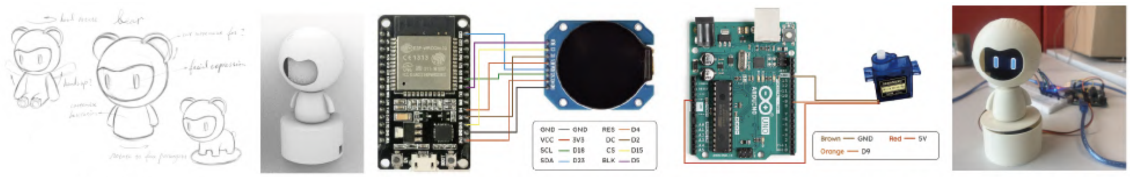
To understand attitudes about IVA and driving behaviour in Asia and Europe, five participants (all males, M = 28.8, SD = 3.42) were interviewed. Four participants had driving experience in Europe and one had driven in both Japan and Europe. The results showed that long-distance driving can be boring and can cause distraction. Although only one participant was familiar with IVA (Nomi of NIO), others showed interest in the concept.

The sketch (Figure 1 (a)) presents three modalities, and the middle was selected for further development. The IVA features facial expressions and body rotation gestures designed for seven highway scenarios [22] (Table I). Figure 1 (b) shows the 3D model created in Rhino 8 (for STL files check section VI) printed using a 3D printer and contains a round i.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 entire prototype.

The TFT display was connected to an ESP32 board and controlled by the Arduino IDE [23] (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 I.

To enable the Arduino IDE to run on ESP32, the Arduino core [24] was installed for ESP32. Libraries Adafruit GC9A01A [25] (v.1.1.1), Adafruit GFX [26] (v.1.11.9), and TFT_eSPI [27] (v.2.5.43) were installed in the Arduino IDE to run the code on the TFT display.

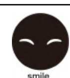
The SG90 servo motor was connected to an Arduino Uno board and controlled by Arduino IDE. The Servo library (v. 1.2.1) was used to execute three gestures designed according to the scenarios [22] in Table I: (1) greeting: turn to the driver (starting position), then turn front ($-5\pi/9$ rad/s) to check the surroundings ($\pm 667\pi/3600$ rad/s) and turn back to the driver ($5\pi/9$ rad/s); (2) situation reporting: turn front ($-5\pi/9$ rad/s) and turn to the driver ($5\pi/9$ rad/s); (3) overtaking after got permission: turn front and rotate to face the vehicle be overtaken ($-5\pi/9$ rad/s, only $\pi/6$ rad with SG90), then turn back to the driver ($5\pi/9$ rad/s).



(a) Sketch (b) 3D model (c) Circuit ESP32-TFT display (d) Circuit Adruino Uno R3-SG90 (e) Prototype

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

TABLE I: IVA behaviour (gestures, facial expressions, and dialogues) in seven highway scenarios.

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!"	 	
Lane changing (construction)	Situation reporting	"Seems there is a construction ahead, we need to change lane."		
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."

III. METHOD OF EXPERIMENT

An experiment was conducted with three groups: Baseline (B), Facial expressions and voice (FV), and Gestures and voice (GV). Group B, as a baseline, uses robotic voice-only interaction, simulating a conventional navigation system similar to Tesla's full self-driving mode. The audio was generated from PlayHT [28] and was edited as another soundtrack in a 270-second video recorded in GTA V. The study was approved by the Ethics 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 [29] running on a Windows PC according to Table I, and the highway route is chosen from the city centre to Beeker's Garage. To get an inside view of AD, two mods were applied: (1) Dynamic Vehicle First Person Camera Mod [30], allowing the camera inside the vehicle to get the driver's perspective and (2) Enhanced Native Trainer Mod [31], which makes characters invisible (i.e., no hands holding the steering wheel were visible, providing a sense of driving in an AV).

Twelve participants (7 females, 5 males, M=27.42, SD=2.11) were recruited via social media. All had valid driving licences, three of them experienced in Tesla autopilot driving. The experimental setup (see Figure 2) consisted of a laptop (Apple Macbook A2442) connected to a screen (RCA RS32F3), headphones (Sennheiser MOMENTUM 4), and the IVA prototype. The IVA's position was adjusted on stacked books (5.5 cm) ensuring visibility, placed at the right front of the participant to mimic dashboard positioning.

Participants received a brief introduction to SAE

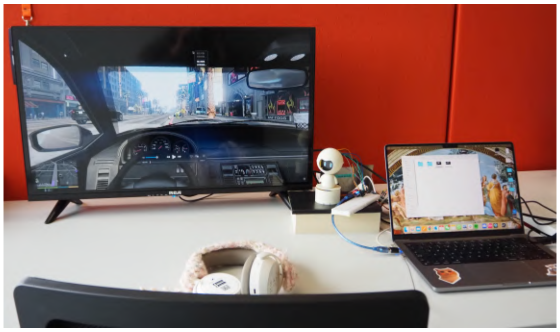


Fig. 2: Experimental setup.

TABLE II: Results from the NASA TLX scale [32].

	B Mean (SD)	FV Mean (SD)	GV Mean (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.

Level 3 automated driving. Each participant experienced three task groups (B always first, FV and GV alternated). During tasks, they engaged in typical secondary activities, such as texting, watching video or reading, simulating realistic driving distractions. They could look up and check the driving situation freely and request manual takeover at any time. After each scenario, participants completed the NASA Task Load Index scale [32] assessing workload, and an acceptance scale [33] evaluating usefulness and satisfaction, using an iPad. Finally, a semi-structured interview was conducted to collect qualitative feedback. The interview transcripts were subjected to thematic analysis, producing insights regarding participant preferences and effectiveness of the interaction.

IV. RESULTS

The workload scores for FV (M=23, SD=24) and GV (M=18, SD=14) were lower than B (M=33, SD=21), with GV showing the lowest score. GV significantly reduced the "Physical demand" workload (M=20, SD=18) compared to B (M=34, SD=23). GV had lower scores across dimensions than FV, except for the "Effort" category, where GV (M=25, SD=21) slightly exceeded FV (M=22, SD=25). FV had similar "Temporal demand" to B.

FV and GV outperformed B in usefulness and satisfaction ratings as shown in Table III. FV had the highest overall scores, except in the "Annoying-Nice" dimension, where GV performed better. GV notably scored lower than FV in the categories "Unpleasant-Pleasant" and "Sleep-inducing-Raising Alertness".

The interview results indicated a preference for GV among seven participants, with five favouring FV and none choosing B. GV was preferred for its clear perception and prevoice indication of information, allowing better concentration on driving tasks. In contrast, FV was preferred for emotional support, absence of mechanical noise, and intuitive understanding compared to gestures.

Thematic analysis identified four themes: perception, efficiency, trust, and emotional support. The participants noted that B lacked sufficient explanatory information, affecting trust. FV provided superior emotional support, but required additional cognitive effort to interpret expressions quickly. GV offered better initial perception but was harder to understand independently and some participants found it monotonous. Trust concerns emerged across all types of IVAs.

V. DISCUSSION AND FUTURE WORK

The results demonstrated notable findings. Both facial expressions and gestures effectively reduced the driver workload in SAE Level 3 AD, improving perceived usefulness and satisfaction. GV had a greater impact than FV, significantly reducing physical demand. Gestures were noticed before voice interactions, offering participants extra time to shift their attention to road conditions. In contrast, facial expressions appeared simultaneously with voice cues, causing participants to split their attention,

TABLE III: Results from the acceptance scale [33].

Negative (-2) Positive (+2)	B: Mean(SD)	FV: Mean(SD)	GV: Mean(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)
Usefulness score	0.68 (0.72)	1.15 (0.48)	0.90 (0.82)
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.

resulting in higher Temporal demand scores.

The acceptance scale showed that both FV and GV improved usefulness and satisfaction compared to baseline. FV was particularly strong in these areas, suggesting that facial expressions provide more emotional support. GV effectively reduced workload and improved functionality, but lacked the emotional engagement found with facial expressions.

The interviews revealed that participants who preferred gestures found them helpful as prealerts before voice notifications, helping to shift attention to road conditions. Those who prefer facial expressions described them as more intuitive, comforting, and appealing. Voice interaction alone, while efficient, lacked comprehensive situational details, highlighting the benefit of combining modalities.

Participants indicated that gestures were functionally preferable for drivers due to clearer perception, but acknowledged that they were sometimes difficult to interpret. In contrast, facial expressions provided emotional support and were preferred by passengers, but could be difficult to notice quickly. Concerns about trust in the system were consistent in all interaction modalities.

After the experiment, two Nissan engineers were interviewed to gain insight from the perspective of a vehicle manufacturer. They noted integration challenges, particularly with regard to connections to the vehicle's CAN bus, privacy concerns, and the critical issue of safely positioning IVAs to avoid injuries during airbag deployment.

The limitations of this study included the use of video simulations instead of real driving scenarios and variability in participation in secondary tasks between participants, which affected IVA perception. Future research should combine gestures and facial expressions for more intuitive interaction, define IVA driving modes tailored to secondary tasks, and explore interactions with vulnerable road users by positioning IVAs to communicate externally.

VI. SUPPLEMENTARY MATERIAL

The 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] S. C. Lee and M. Jeon, "A systematic review of functions and design features of in-vehicle agents," *International Journal of Human-Computer Studies*, vol. 165, p. 102864, 2022.

[2] T. Hagiya and K. Nawa, "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*, 2020, pp. 1–8.

[3] M. Gordon and C. Breazeal, "Designing a virtual assistant for in-car child entertainment," in *Proceedings of the 14th International Conference on Interaction Design and Children*, 2015, pp. 359–362.

[4] D. R. Large, G. Burnett, V. Antrobus, and L. Skrypchuk, "Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task-related fatigue," *IET Intelligent Transport Systems*, vol. 12, no. 6, pp. 420–426, 2018.

[5] S. Li, T. Zhang, W. Zhang, N. Liu, and G. Lyu, "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*, vol. 86, p. 103098, 2020.

[6] R. Häuslschmid, M. von Buelow, B. Pfleging, and A. Butz, "Supportingtrust in autonomous driving," in *Proceedings of the 22nd international conference on intelligent user interfaces*, 2017, pp. 319–329.

[7] S. C. Lee, H. Sanghavi, S. Ko, and M. Jeon, "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*, 2019, pp. 209–214.

[8] J. Zihsler, P. Hock, M. Walch, K. Dzuba, D. Schwager, P. Szauer, and E. Rukzio, "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*, 2016, pp. 9–14.

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

[10] S. Chakravarthi Kumaran, T. Bechor, and H. Erel, "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*, 2024, pp. 86–95.

[11] NIO. (2017) Nomi. [Online]. Available: https://www.youtube.com/watch?v=SAZ2Dd9lrVc

[12] T. M. Corporation. (2024) Toyota launches sales of "kirobo mini" nationwide — toyota motor corporation official global website. [Online]. Available: https://global.toyota/en/detail/19880995

[13] H. Tanabe, Y. Yoshihara, N. Karatas, K. Fujikake, T. Tanaka, S. Takeuchi, T. Yamamoto, M. Harazawa, and N. Kamiya, "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, 2022, pp. 108–123.

[14] K. Williams, J. A. Flores, and J. Peters, "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*, 2014, pp. 1–8.

[15] M. Wang, S. C. Lee, G. Montavon, J. Qin, and M. Jeon, "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*, 2022, pp. 86–95.

[16] M. Jonsson and N. Dahlbäck, "Impact of voice variation in speech-based in-vehicle systems on attitude and driving behaviour," *Human Factors and Ergonomics Society Europe Chapter (HFES)*, 2009.

[17] S. Lee, R. Ratan, and T. Park, "The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style," *Multimodal Technologies and Interaction*, vol. 3, no. 1, p. 20, 2019.

[18] P. A. Ruijten, J. M. Terken, and S. N. Chandramouli, "Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior," *Multimodal Technologies and Interaction*, vol. 2, no. 4, p. 62, 2018.

[19] J. M. Kraus, F. Nothdurft, P. Hock, D. Scholz, W. Minker, and M. Baumann, "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*, 2016, pp. 129–134.

[20] M. Wang, S. C. Lee, H. Kamalesh Sanghavi, M. Eskew, B. Zhou, and M. Jeon, "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*, 2021, pp. 247–254.

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

[22] M. Beggiato, F. Hartwich, K. Schleinitz, J. Krems, I. Othersen, and I. Petermann-Stock, "What would drivers like to know during automated driving? information needs at different levels of automation." in *7. Tagung Fahrerassistenzsysteme*, 2015.

[23] (2024) Software. [Online]. Available: https://www.arduino.cc/en/software

[24] espressif. (2024) Github - espressif/arduino-esp32: Arduino core for the esp32. [Online]. Available: https://github.com/espressif/arduino-esp32

[25] (2024) Adafruit gc9a01a. [Online]. Available: https://www.arduino.cc/reference/en/libraries/adafruit-gc9a01a/

[26] 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. [Online]. Available: https://github.com/adafruit/Adafruit-GFX-Library

[27] (2024) Tft_espi. [Online]. Available: https://www.arduino.cc/reference/en/libraries/tft_espi/

[28] PlayHT. (2024). [Online]. Available: https://play.ht/

[29] G. Rockstar. (2024) Grand theft auto v - rockstar games. [Online]. Available: https://www.rockstargames.com/zh/gta-v

[30] V. H1gh. (2023) How to install dynamic vehicle first person camera - gta v mods. [Online]. Available: https://www.youtube.com/watch?v=jwxgmAHtwIY

[31] Willez. (2020) Gta 5 - how to install enhanced native trainer (2020). [Online]. Available: https://www.youtube.com/watch?v=UHHXTh0Xdow

[32] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[33] J. D. Van Der Laan, A. Heino, and D. De Waard, "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 1, pp. 1–10, 1997.