

# **Enhancing Driver Experience in SAE Level 3 Automated Vehicles through Multimodal and Emotion-aware In-vehicle Agents**

XINGJIAN ZENG, Eindhoven University of Technology, The Netherlands

MD SHADAB ALAM\*, Eindhoven University of Technology, The Netherlands

PAVLO BAZILINSKYY, Eindhoven University of Technology, The Netherlands

In-vehicle agents (IVAs) are emerging as transformative innovations in intelligent transportation systems, particularly in automated driving contexts. This paper integrates two complementary studies on the development and evaluation of robot-like IVAs equipped with multimodal interaction and emotional feedback for SAE Level 3 automated vehicles. The first study introduced a robot-like IVA capable of communicating through gestures and facial expressions. An experiment with 12 participants showed that both modalities reduced workload (NASA TLX mean scores: baseline = 33%, facial expressions = 23%, gestures = 18%) and enhanced perceived usefulness and satisfaction. Seven participants preferred gestures for practicality and anticipatory cues, while five preferred facial expressions for their emotional and aesthetic qualities. The second study (N=12) developed a prototype with emotional feedback via facial emotion detection. The results revealed that emotional feedback and working status did not significantly affect overall workload or acceptance, although feedback influenced physical and temporal demands, and its interaction with working status significantly affected workload. Voice communication remained the primary way of interaction, while challenges included the accuracy of emotion detection and accounting for physical conditions such as fatigue and stress.

Additional Key Words and Phrases: In-Vehicle Agent, Robot-Like Agent, Emotional feedback, Emotion detection, Emotive driving

## **ACM Reference Format:**

Xingjian Zeng, Md Shadab Alam, and Pavlo Bazilinsky. 2025. Enhancing Driver Experience in SAE Level 3 Automated Vehicles through Multimodal and Emotion-aware In-vehicle Agents. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 22 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## **1 Introduction**

With advances in automotive technology (AT) and artificial intelligence (AI), in-vehicle agents (IVAs) have emerged as notable innovations within intelligent transportation systems (ITS) [19]. IVAs are typically embodied as driving assistants integrated into vehicle systems and can be broadly classified into voice agents (implemented through computer-generated sound or human voice recordings), virtual agents (developed on visual displays with face-only or full-body avatars) and physical-embodied agents (based on a movable robotic interface with crude features) [26]. Their primary objective is to support driving tasks and thus enhance the overall driving experience. Among these

---

\*Corresponding Author

Authors' Contact Information: Xingjian Zeng, esse.zeng009@gmail.com, Eindhoven University of Technology, Eindhoven, The Netherlands; Md Shadab Alam, m.s.alam@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands; Pavlo Bazilinsky, p.bazilinsky@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

types, physical IVAs have recently gained increasing attention due to their intuitive interactive capabilities, potential for stronger emotional connection, and deep integration with smart cockpit technologies such as NIO’s NOMI [32].

The first physical IVA, the affective intelligent driving agent (AIDA), was introduced by the MIT Media Lab in 2013 [48]. AIDA displayed facial expressions on a screen and provided navigation assistance through interactive communication. The Carvatar concept developed by Zihlsler et al. [49] also employed facial expressions to improve trust, while the robot human-machine interface (RHMI) [42] incorporated dynamic cues such as eye colours and body movements to deliver takeover warnings. More recently, Srivatsan et al. [8] demonstrated that robotic gestures significantly improve participant trust ratings and safety-related experience in automated vehicles (AV).

The commercial adoption of physical IVAs has notably advanced in eastern Asia, with successful products such as NIO’s NOMI [31] and Baidu’s Xiaodu [4] in China, Mochi in Japan [9]. These typically feature geometric designs with digital facial expressions, but still rely heavily on voice or virtual interaction, with limited use of gestures such as small angle rotation. As shown in Figure 1, these systems provide inspiration for new prototypes with richer multimodal capabilities.

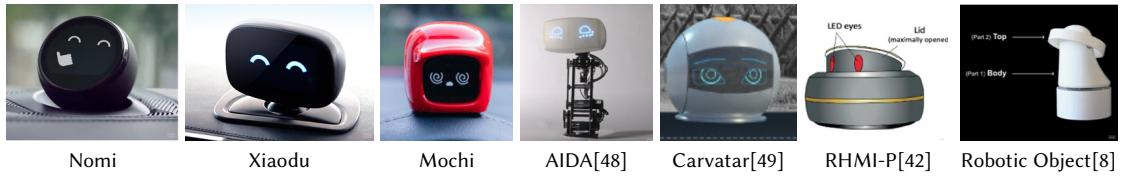


Fig. 1. Benchmark of existing and proposed IVAs.

In manual driving contexts, IVAs can provide vehicle-to-vehicle communication [15], entertain or comfort children to reduce distractions [12], decrease fatigue through social conversations [25], and mitigate negative emotions through positive speech feedback [27]. In automated driving (AD), IVAs can help drivers understand the functions and features of the system, while explaining the status of the system and the intention of AV driving behaviours [17, 23, 35], thus improving trust and acceptance. Physical agents in particular have been shown to enrich driver behaviour and experience, with facial expressions and gestures significantly improving trust. The robotic object developed by Srivatsan et al. [8], could greet the passenger, indicate that the vehicle was attentive to its surroundings, and indicated that the drive was about to begin. The study indicated that the trust ratings and safety-related experience of the participants were higher than those of a baseline group that did not interact with the robot after evaluation [49]. Cavatar can translate the state of the car into human behaviour and expressions which can be intuitively interpreted by the driver, through social signals and anthropomorphism. The driver is therefore more aware of the situation and might gain more trust in the system [8]. This study was especially focused on SAE Level 3 AV because automation handles most driving, yet drivers still need to take over when requested. This creates meaningful and safety-critical interactions in which a physical agent can support trust, attention, and smooth handovers. It also aligns closely with near-term industry deployments, making the findings highly relevant.

Voice interaction remains the most common modality in SAE Level 3 AD [36] due to low visual distraction [46], but research shows that no single voice suits all contexts [21]. Although conversational interfaces are often more trusted and anthropomorphised [35], physical robots add gestures and embodiment, increasing competence and reducing workload [45]. Gestures, as in RHMI [42], can signal urgency or intentions through dynamic movement, while small talk and social cues can improve trust and perceived warmth [24].

Beyond explicit voice input or tactile input by pressing the buttons, recent work highlights the potential of implicit interactions, where IVA adapts to user states without direct input [22, 40]. Facial emotion detection, supported by advances in deep learning frameworks such as DeepFace [38], is especially promising. Facial expressions convey more information (55%) than tone (38%) or words (7%) [29] and can be captured continuously in driving scenarios. Ekman and Dacher [37] further argue that facial expressions are universal, making them suitable for cross-cultural IVA interactions.

### 1.1 Aim of Study

The aim of this research is to investigate how physical IVAs can enhance user experience and driver support in SAE Level 3 AD scenarios through multimodal interaction and emotional feedback. First, *Study1* examines the integration of gestures and facial expressions with voice interaction to determine their relative advantages and limitations in reducing workload, increasing usefulness, and fostering trust in AD. Second, *Study2* explores implicit interactions by equipping the IVA with facial emotion detection and corresponding emotional feedback, thereby assessing its influence on driver workload, alertness, and acceptance under varying working conditions. Through these complementary objectives, the study seeks to define how robot-like IVAs can balance functional effectiveness with emotional involvement, ultimately contributing to safer, more intuitive, and human-centred AD experiences.

## 2 Method

Two different prototypes were developed for two complementary studies on the development and evaluation of robot-like IVAs equipped with multimodal interaction (*Study1*) and emotional feedback capabilities (*Study2*) in a Level 3 SAE AV. The prototypes were evaluated through the corresponding experiments to answer the two proposed research questions, as mentioned in subsection 1.1. The study was approved by the Ethics Review Board of the Eindhoven University of Technology and the participants gave their informed consent to use their data.

The scenarios videos were recorded in the GTA V video game [11] running on a Windows PC according to Table 4, 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 [14], allowing the camera inside the vehicle to get the driver's perspective, and (2) Enhanced Native Trainer Mod [47], which makes characters invisible (i.e., no hands holding the steering wheel were visible, providing a sense of driving in an AV). The videos were then edited together into one video (see supplementary material).

There are four types of raw data collected during *Study1* and *Study2* from each participant: basic information survey (including Driving Behaviour Questionnaire), NASA TLX Scale on mobile app [16, 30], online acceptance scale [44], and a semi-structured interview (see supplementary material). The experiment interview was analysed through thematic analysis [7]. The themes were generated after coding the data in the transcription.

Furthermore, for *Study2*, SPSS can only accept a wide format of CSV to perform the repeated measures ANOVA test [18] with the value alpha being 0.05. In this case, NASA TLX results [16] and acceptance scale [44] were adjusted from long format to wide format.

## 3 Study 1

### 3.1 Prototype Development

Figure 2a presents three modalities, and the middle was selected for further development. By adopting this configuration, the mechanical architecture is both simplified and stabilised, minimising potential sources of error or failure. The IVA

features facial expressions and body rotation gestures (Table 1) were designed for seven highway scenarios (greeting passengers, enter highway, speed limit and speed report, overtaking, lane changing, congestion, exit highway) inspired by the study of Beggiato et al. [6]. Figure 2b shows the 3D model created in Rhino 8 (for STL files check section 7) printed using an Ultimaker 2+ Connect 3D printer and contains a round 1.28-inch IPS-TFT display (240 x 240 pixels, IPS GC9A01) inside the round head ( $r=31\text{mm}$ ) connected to ESP32 (Figure 2c: Circuit ESP32-TFT display). The gestures are driven by an SG90 servo motor inside the stand connected to Arduino Uno R3 (Figure 2d). 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 2e shows the entire prototype.

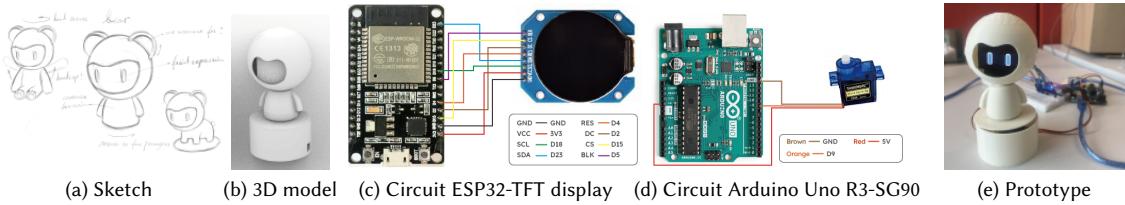


Fig. 2. The prototype development of the robot-like IVA.

The TFT display was connected to an ESP32 board and controlled by the Arduino IDE [1] (v.2.3.2) on the Apple Macbook A2442. Five facial expressions (normal, smile, excited, realising, sad) were designed, shown in Table 1. To enable the Arduino IDE to run on ESP32, the Arduino core [41] was installed for ESP32. Libraries Adafruit GC9A01A (v.1.1.1) [2], Adafruit GFX (v.1.11.9) [20], and TFT\_eSPI (v.2.5.43) [3] were installed in the Arduino IDE to run the code on the TFT display.

### 3.2 Method

The experiment was carried out 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 [34] and was edited as another soundtrack in a 270-second video recorded in GTA V.

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 3) 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 (55mm) ensuring visibility, placed at the right front of the participant to mimic dashboard positioning. Participants received a brief introduction to SAE Level 3 AD. Each participant experienced three task groups (B always first, FV and GV alternated). During tasks, they chose and 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 [16] assessing workload, and an acceptance scale [44] 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.

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

Scenarios	IVA gestures (GV)	Dialogues (FV & GV)	Facial expressions (FV)	Dialogues (B)
greeting passengers	Greeting (gesture)	“Welcome! My name is Eva. Shall we start our trip?” (Driver: Yes) “Here we go!”	 smile excited	
Enter highway	Situation reporting	“We will enter the highway ahead.”	 normal	“Enter the highway ahead.”
Speed limit and speed report		“The speed limit is 90, and right now we are at 87.”	 smile	“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!”	 smile excited	
Lane changing (construction)	Situation reporting	“Seems there is a construction ahead, we need to change lane.”	 realising	
Congestion	Situation reporting	“Seems there is a traffic jam, we need to slow down.”	 sad	
Exit highway	Situation reporting	“We will exit the highway ahead.”	 normal	“Exit the highway ahead.”

### 3.3 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 3. 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.



Fig. 3. Experimental setup.

Table 2. Results from the NASA TLX scale [16].

	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.

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.

#### 4 Study 2

##### 4.1 Prototype Development

The final concept sketch is shown in Figure 4. The soft LED (model: HRP-2064) was chosen to display the IVA's facial expressions. A gear transmission was used instead of the rubber wheel transmission because the gear transmission

Manuscript submitted to ACM

Table 3. Results from the acceptance scale [44].

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.

was more stable and easy to assemble. Due to safety concerns, the final prototype can be easily locked to prevent secondary damage in car accidents. On the other hand, the behaviour designed for the IVA robot was updated into two parts: the first part is to help with driving tasks and explain the current state (Table 4), and the other part is emotional feedback (Table 5). Table 4 was based on the table of IVA behaviour in seven highway scenarios; they have the same structure, while Table 4 adds a shutdown scenario, which means the robot will say goodbye to passengers and shut itself down when arriving at the destination, to complete the overall user experience and new gestures. Table 5 is based on the driver's emotions detected by the camera (Raspberry Pi Camera Module 2) and analysed through DeepFace [38]. Although OpenFace [5]performs better, it was not chosen due to the unmatched video stream formats. There are seven emotions in the DeepFace library, including angry, fear, neutral, sad, disgust, happy, and surprise, which is quite similar to the 7 basic emotions [13](except for neutral replacing contempt). For each emotion detected, the IVA robot has the corresponding feedback (including dialogue, facial expressions, and gestures).

**4.1.1 Materials.** A Raspberry Pi 5 (model: B, 4GB RAM, 32GB SD, Raspberry Pi OS (64bit): Bookworm 12.8); a PWM-PCA9685(16-channel I2C PWM-Servo Controller - PCA9685); two servomotors (MG996R Servo - 10kg - Continuous); a camera (Raspberry Pi Camera Module 2); a soft LED (model: HRP-2064); an AC-DC external power supply for servomotors (model: AED45US05, 5V, 6.0A); 3D print (see section 7); laser cut medium density fibreboard (MDF) board (see section 7).

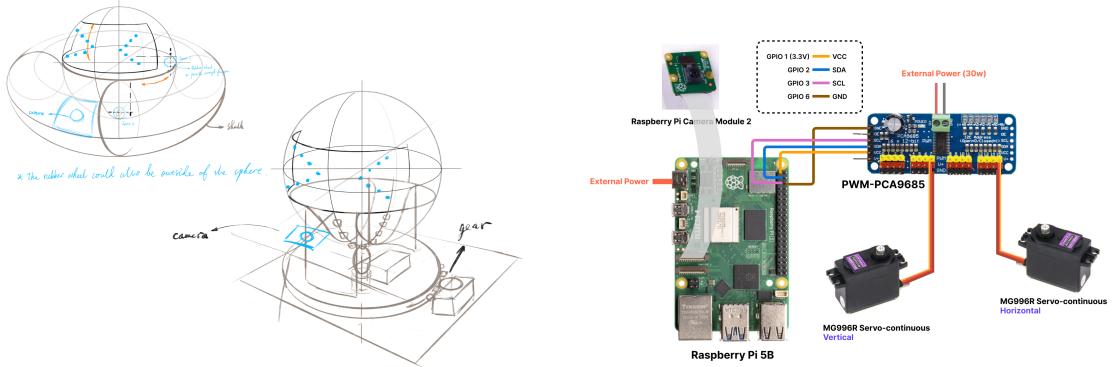


Fig. 4. Sketch of the final concept (left) and circuit (right).

Scenarios	Dialogue	Facial expressions (5)	Gestures H & V
Wake up	– “Welcome! My name is Eva. Shall we start our trip?” –Yes – “Here we go!”		Up to $5\pi/18$ (4 rad/s)
Enter highway	“We will enter highway ahead.”		Left to $\pi/3$ , wait 1s, right to $\pi/3$ (4 rad/s)
Speed limit and speed report	“The speed limit is 90km/h, and right now we are at 87km/h.”		Left to $\pi/3$ , wait 1s, right to $\pi/3$ (4 rad/s)
Overtaking	– “The front car is driving too slow, we gonna overtake it! wow Nice!”		Left to $4\pi/9$ , right to $4\pi/9$ , left to $4\pi/9$ , right to $4\pi/9$ (8 rad/s)
Lane changing (construction)	“Seems there is a construction ahead, we need to change lane.”		Up to $5\pi/18$ , wait 0.5s, left to $\pi/3$ , wait 1s, right to $\pi/3$ , wait 0.5s, down to $5\pi/18$ (4 rad/s)
Congestion	“Seems there is traffic jam ahead, we need to slow down.”		Down to $5\pi/18$ , wait 0.5s, left to $\pi/3$ , right to $\pi/3$ , wait 0.5s, up to $5\pi/18$ (4 rad/s)
Exit highway	“We will exit the highway ahead.”		Left to $\pi/3$ , wait 1s, right to $\pi/3$ (4 rad/s)
Shut down	“Thanks for your participation, and wish you a lovely day!”		down to $5\pi/18$ (4 rad/s)

Table 4. Updated IVA behaviour (gestures, facial expressions, and dialogues) in eight highway scenarios.

Emotion detected (7)	Dialogue	Facial expressions	Gestures H & V
Happy	"It's great to see you in such a good mood! Would you like to share what's making you happy?"	 Happy	Left to $\pi/3$ , right to $\pi/3$ (4 rad/s)
Neutral	(None)	 Neutral	(None)
Surprise	"Everything is under control. I'm here if you need support."	 Surprise	Up to $5\pi/18$ , left to $\pi/3$ , right to $\pi/3$ , down to $5\pi/18$ (4 rad/s)
Sad	"It seems like you're feeling a bit down. Would talking or some relaxing music help?"	 Neutral	Left to $5\pi/18$ , right to $5\pi/18$ (4 rad/s)
Angry	"Traffic can be tough. Don't worry, the system is optimized to handle it. How about some calming music?"	 Happy	Left to $\pi/3$ , right to $\pi/3$ (4 rad/s)
Fear	"It seems like you're feeling a bit uneasy. I'll monitor everything and react to keep us safe. Is there anything I can do for you?"	 Happy	Left to $\pi/3$ , right to $\pi/3$ (4 rad/s)
Disgust	"Is there something unpleasant I can help with? Perhaps a window adjustment or a different temperature setting?"	 Neutral	Left to $\pi/3$ , right to $\pi/3$ (4 rad/s)

Table 5. IVA behaviour of emotional feedback.

**4.1.2 Hardware.** The assembly began with the integration of the electronic components (see Figure 4). Two servomotors were connected to the PWM-PCA9685 via Dupont wires, using pin 7 for pitch and pin 15 for yaw control. Communication between the Raspberry Pi and the PCA9685 was established through the following GPIO pins: GPIO 1 - VCC, GPIO 2 - SDA, GPIO 3 - SCL, and GPIO 6 - GND. The camera module was connected through the Raspberry Pi CAM 1 port. External power was supplied to the Raspberry Pi, the soft LED module, and the PWM-PCA9685 (Note: an external power source is required for the PCA9685 if the total current of the servomotors exceeds 500 mA).

After completion of the electronics setup, the mechanical structure was assembled (see Figure 5). The assembly begins with a wooden base, a rotating platform, and a servo-driven gear system (1–3), forming the mechanical foundation. A transparent, 3D-printed ring structure and a soft LED module were added (4–5), and then fixed on the rail in (3). Electronic components, including a Raspberry Pi, a servo-control board, and a camera module, are integrated (6). The board with electronic components is fixed upside down. Finally, a semi-transparent dome enclosure completes the system (8).

**4.1.3 Software.** The software environment was based on the Raspberry Pi OS (64-bit, Bookworm 12.8). Source code is available in supplementary material. All emotional feedback is set according to Table 5, including voice, facial expressions,

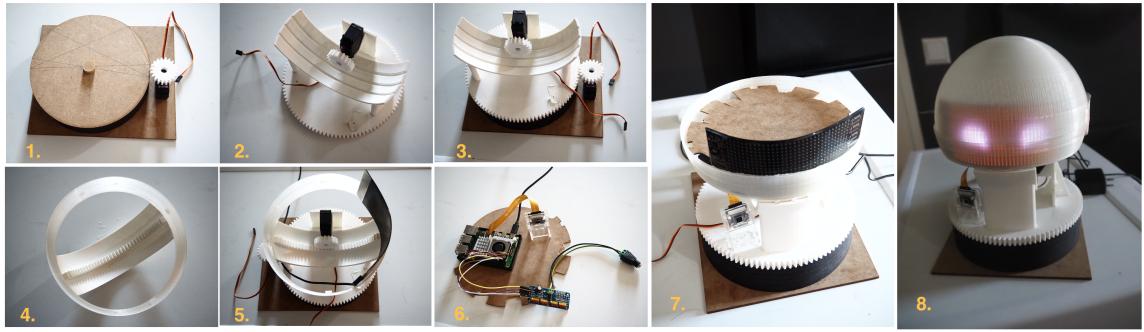


Fig. 5. Hardware assembly process (from 1 – 8). 1. wooden base; 2. rotating platform; 3. servo-driven gear system; 4. ring structure; 5. soft LED module; 6. electronics; 7. putting electronics upside down; 8. fixing the dome

and gestures. All audio was generated from PlayHT [34] (see supplementary material). However, sometimes the audio will just skip around 1 second when playing with Bluetooth connection, so all audio of basic behaviour is edited to add 1 second of silence at the beginning, using FFmpeg [10]).

The facial expressions of the IVA robot are actually uploaded through Bluetooth, since the soft LED cannot connect directly to the GPIO of Raspberry Pi, which means the soft LED cannot be controlled directly by Raspberry Pi. However, it has a mobile phone app and uses Bluetooth to send data, which makes it possible for a laptop to receive data sent from the mobile app to the soft LED. First, connect the mobile phone and laptop and the soft LED. Then send the expressions from the mobile app, and the data is a long string of numbers and letters. Next, connect the soft LED with the Raspberry Pi and use the Raspberry Pi to send the same string of numbers and letters; the expressions will eventually be shown on the soft LED.

## 4.2 Method

For *Study2*, an experiment was conducted to evaluate the final concept design. The experiment had two groups: Group *W* (working) and Group *N* (no working). Since SAE Level 3 AD is conditionally automated, working here means the participants in the Group *W* has a secondary task. In order to control the variables, participants in Group *W* need to play the Poly Bridge game [28] on an iPad (model: A2229) during the tasks, while participants in Group *N* cannot use any personal products on digital screens (including smart phones). Each participant in Group *W* and Group *N* needs to complete two tasks: Task *B* (basic) and Task *E* (emotional feedback). Both tasks are combined with a five-minute driving simulation and two surveys to fill out. The behaviour of the IVA robot in Task *B* is shown in Table 4. Task *E* is based on Task *B*, adding an emotional feedback function according to Table 5.

**4.2.1 Setup and Preparation.** Figure 6 shows the experimental setup. A JBL GO 3 (model: JBLG03BLK) speaker was connected to the Raspberry Pi inside the prototype, and the Raspberry Pi was remotely controlled [43] by the laptop (Apple Macbook A2442). An RCA 32' LED TV (model: RS32F3) was also connected to the laptop via an HDMI cable. The detailed parameters of the layout are shown in Figure 6. The participants were asked to sit in front of the centre of the RCA TV, and the prototype should face the participants with an angle of to  $\pi/8$  so that the camera could detect the emotions of the participants. The position of the prototype is settled on the front right of the participant, corresponding to the position above the dashboard in a real car.

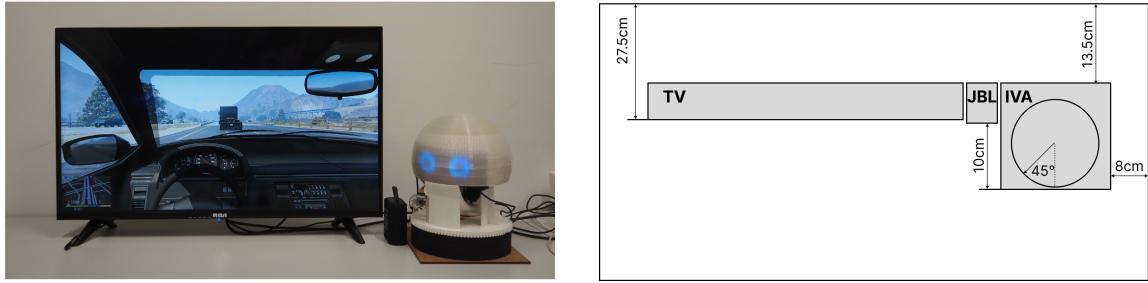


Fig. 6. Scenario setup (left) and layout of the equipment (right).

In Task *E*, the prototype will not only execute the behaviour in Table 4 but also the emotional feedback in Table 5. In case two actions interfere with each other and avoid frequent actions, emotional detection and feedback were only enabled after the following scenarios: welcome, speed report, construction, and traffic jam. Feedback can only be triggered once for each scenario. Each behaviour in Task *B* was matched to the video.

Furthermore, the servomotor (MG996R Servo—10kg—Continuous) can only use speed and time as parameters. Even when using mathematical methods to set the angle as the parameter, errors still occurred due to the startup delay. Consequently, the prototype cannot return to the exact starting position after each task. Thus, the position of the prototype needs to be adjusted before each task according to Figure 6.

**4.2.2 Participants.** A total of 12 participants (age:  $M = 26.33$ ,  $SD = 2.50$ ; 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 incentives were offered for the user test. All participants were over 18 years of age and had a driver's licence. Three participants had experience driving an automated vehicle (with automatic lane changing / automatic turning, automatic overtaking / automatic navigation, or higher functions, such as Tesla, Waymo, Apollo, etc.). One has been a passenger of an automated vehicle.

**4.2.3 Procedures.** First, participants signed the consent form and completed an online survey to collect basic information (see supplementary material). The slides then briefly introduced background information about the project to the participants (see supplementary material). The participants then took a seat and had two tasks to complete: Task *B* and Task *E*. To reduce the error, the sequence of Task *E* and Task *B* is changed for each half of the participants in Group *W* and Group *N*. The prototype was controlled by the Raspberry Pi, which was remote controlled by the author on the laptop during the experiment. After each task, participants were asked to fill in the NASA Task Load Index scale [16] on the official application [30] to measure workload and the acceptance scale [44] (integrated in the online survey) to measure the overall experience on the iPad, since Ju & Leifer [22] suggested that implicit interaction could reduce cognitive load and enhance user experience. Finally, a semi-structured interview was conducted to collect the experiment experience. During each task, participants were asked to imagine themselves acting as drivers in a SAE Level 3 AD (which means the driver can use AD conditionally and still needs to be responsible for driving duration) and to have an AD experience with the prototype. Participants in Group *W* will be asked to play the Poly Bridge game [28] on an iPad (model: A2229) during the tasks, while participants in Group *N* cannot use any personal products on digital

Table 6. Results from the NASA TLX Scale [16].

	<i>W</i>		<i>N</i>		Repeated-measures ANOVA
	Task <i>E</i>	Task <i>B</i>	Task <i>E</i>	Task <i>B</i>	
	M(SD)	M(SD)	M(SD)	M(SD)	
Mental Demand (%)	49 (32)	55 (31)	52 (20)	33 (20)	
Physical Demand (%)	16 (27)	21 (25)	22 (27)	16 (19)	
Temporal Demand (%)	66 (30)	48 (30)	43 (29)	25 (19)	
Performance (%)	45 (27)	68 (24)	14 (12)	13 (13)	
Effort (%)	50 (31)	41 (17)	22 (17)	20 (13)	
Frustration (%)	33 (14)	39 (20)	21 (15)	16 (15)	
Average (%)	47 (26)	47 (23)	31 (22)	22 (17)	p=0.139      p=0.057      p=0.037

Note: *W*=working group, *N*=no-working group.

screens (including smart phones). All participants were allowed to look up and check the situation at any time. If they felt that they wanted to take over control immediately, they were asked to inform the author about it.

### 4.3 Results

**4.3.1 NASA TLX.** Table 6 shows the mean and standard deviation values of each workload dimension score and the weighted average score of the overall workload of different groups (*W* or *N*) and different tasks (*E* or *B*). Since the standard deviations shown in Table 6 were too large, a statistical analysis (repeated measures ANOVA) is needed in this case. Table 6 also shows the analysis that examined the effects of different working status (*W/N*) and emotional feedback (*E/B*) on average weighted workload measures using the NASA Task Load Index [16]. There were two independent variables in the experiment: emotional feedback (*E / B*) as a variable within groups, and working status (*W / N*) as a variable between groups, which means that two-way ANOVA should be replaced by repeated measures ANOVA in this case. Thus, the repeated measures ANOVA test was applied with SPSS (see supplementary material). The main results are as follows: (1) The work status factor between subjects (*W / N*) did not have a significant effect on the average weighted workload measured by NASA TLX (p=0.139); (2) The within-subjects factor emotional feedback (*E/B*) also did not have a significant effect on the average weighted workload measured by NASA TLX (p=0.057); (3) The interaction effect between emotional feedback and working status was significant for the average weighted workload measured by NASA TLX (p=0.037).

**4.3.2 Acceptance Scale.** The results of the acceptance scale are shown in Table 8. Similarly to NASA TLX, the standard deviations shown in Table 8 were too large, and statistical analysis is needed. Table 8 also shows that the analysis examined the effects of different working status (*W/N*) and emotional feedback (*E/B*) on overall usefulness and satisfaction measures using the Van der Laan Acceptance Scale [44]. There were two independent variables in the experiment: emotional feedback (*E / B*) as a variable within groups, and working status (*W / N*) as a variable between groups, which means that two-way ANOVA should be replaced by repeated measures ANOVA in this case. Thus, the repeated measures ANOVA test was applied with SPSS (see supplementary material). The results suggest that emotional

Table 7. Results from the Acceptance Scale [44].

Negative (-2)	Positive (+2)	W		N	
		Task E	Task B	Task E	Task B
		M(SD)	M(SD)	M(SD)	M(SD)
Useless	Useful	0.67 (1.03)	1.00 (1.10)	1.17 (0.75)	1.33 (0.82)
Unpleasant	Pleasant	0.83 (0.98)	0.67 (1.03)	1.00 (0.89)	1.33 (0.52)
Bad	Good	0.83 (0.75)	1.00 (0.63)	1.17 (0.75)	1.33 (0.52)
Annoying	Nice	0.83 (0.75)	0.83 (0.75)	1.17 (0.75)	1.33 (0.52)
Superfluous	Effective	0.83 (0.75)	0.67 (0.82)	0.83 (0.75)	1.00 (1.26)
Irritating	Likeable	0.67 (0.82)	0.50 (0.84)	1.17 (0.75)	1.33 (0.52)
Worthless	Assisting	1.17 (0.41)	1.33 (0.82)	1.33 (0.52)	1.33 (0.52)
Undesirable	Desirable	0.67 (1.03)	0.17 (1.17)	1.33 (0.52)	1.33 (0.52)
Sleep-inducing	Raising Alertness	0.33 (1.21)	0.00 (1.55)	1.00 (0.89)	0.67 (0.52)
Repeated-measures ANOVA					
		W/N	E/B	W/N*E/B	
Overall usefulness score	0.77 (0.64)	0.80 (0.63)	1.10 (0.53)	1.13 (0.47)	p=0.299 p=0.801 p=1.000
Overall satisfaction score	0.75 (0.79)	0.54 (0.80)	1.17 (0.61)	1.33 (0.44)	p=0.090 p=0.926 p=0.412

Note: W=working group, N=no-working group.

feedback (E/B) and working status did not significantly impact the perception of usefulness or satisfaction. Furthermore, there was no significant interaction between emotional feedback and working status. These findings suggest that within the given experimental conditions, emotional feedback and working status did not play a significant role in shaping user acceptance levels.

**4.3.3 Interview.** The experiment interview was analysed through thematic analysis. The themes were generated after coding the data in the transcription. Details can be found in the supplementary material. The first column contains the participants' group and the number of participants who contributed to each code. There are five themes (technology concern, efficiency, attitude, prototype, and experiment) according to 16 codes extracted from the interview transcription (see the supplementary material).

The results indicate that most of the participants trust the robot (P1, P3, P4, P5, P6, P7, P8, P9, P10, P12) while others have doubts about AD technology (P2, P5, P11) and accuracy of facial detection (P1, P2, P3, P4, P5, P7, P8, P10, P12). Furthermore, some suggested that other emotions (physical conditions) such as tiredness and stress should also be taken into account (P6, P9, P11). Although IVA robot behaviour is designed for basic highway situations, and emotional feedback can encourage communication between the driver and the IVA robot (P2, P6, P7, P11), it can also make the participant annoyed (P2, P5, P7, P9) and distracted (P2, P5, P7). Even the IVA robot itself can be a good reminder of the situations (P1, P4, P7), some of the dialogue would be time-consuming (P2, P3) or cause misunderstanding (P4, P11). Therefore, some participants prefer direct suggestions (P3, P6) rather than open questions. In addition, some participants complained about the noise (P1, P2, P5, P9) and scenarios (P6, P8, P10). Other participants consider the meaning of the behaviour (P5, P7, P9) and the ideal size of the prototype (P1, P10).

Table 8. Results from the Acceptance Scale [44].

Negative (-2)	Positive (+2)	W		N	
		Task E	Task B	Task E	Task B
		M(SD)	M(SD)	M(SD)	M(SD)
Useless	Useful	0.67 (1.03)	1.00 (1.10)	1.17 (0.75)	1.33 (0.82)
Unpleasant	Pleasant	0.83 (0.98)	0.67 (1.03)	1.00 (0.89)	1.33 (0.52)
Bad	Good	0.83 (0.75)	1.00 (0.63)	1.17 (0.75)	1.33 (0.52)
Annoying	Nice	0.83 (0.75)	0.83 (0.75)	1.17 (0.75)	1.33 (0.52)
Superfluous	Effective	0.83 (0.75)	0.67 (0.82)	0.83 (0.75)	1.00 (1.26)
Irritating	Likeable	0.67 (0.82)	0.50 (0.84)	1.17 (0.75)	1.33 (0.52)
Worthless	Assisting	1.17 (0.41)	1.33 (0.82)	1.33 (0.52)	1.33 (0.52)
Undesirable	Desirable	0.67 (1.03)	0.17 (1.17)	1.33 (0.52)	1.33 (0.52)
Sleep-inducing	Raising Alertness	0.33 (1.21)	0.00 (1.55)	1.00 (0.89)	0.67 (0.52)
Overall usefulness score		0.77 (0.64)	0.80 (0.63)	1.10 (0.53)	1.13 (0.47)
Overall satisfaction score		0.75 (0.79)	0.54 (0.80)	1.17 (0.61)	1.33 (0.44)

Note: W=working group, N=no-working group.

Table 9. Statistical Analysis of Results from the Acceptance Scale

Source	Measure	F	p
Working Status	Overall usefulness score	1.200	0.299
	Overall satisfaction score	3.525	0.090
Emotional Feedback	Overall usefulness score	0.067	0.801
	Overall satisfaction score	0.009	0.926
working status * Emotional Feedback	Overall usefulness score	0.000	1.000
	Overall satisfaction score	0.732	0.412

Table 10. Thematic analysis of the interview.

Group	Codes	Themes	Participants	Support
W (3) & N (3)	Automated Driving	Technology Concern	P1, P2, P5, P6, P11, P12	P1: "I think it drives better than me." P2: "I think I just don't trust the technology behind it." P5: "probably I don't fully trust the AD technology" P6: "It drives more smoothly than me". P11: "I do not trust AD technology". P12: "And I have been a passenger of an automated vehicle so I guess I check the situation much less than others".

<b>Group</b>	<b>Codes</b>	<b>Themes</b>	<b>Participants</b>	<b>Support</b>
<i>W (5) &amp; N (4)</i>	Facial Detection	Technology Concern	P1, P2, P3, P4, P5, P7, P8, P10, P12	<p>P1: "It seems that emotion detection gives the wrong output. I never feel fear".</p> <p>P2: "I don't feel so angry or happy, just a bit down at the beginning".</p> <p>P3: "It said I feel uneasy. I just yawned".</p> <p>P4: "But I'm not angry, I was relaxed".</p> <p>P5: "But I don't have the happy or sad emotion, more like curiosity".</p> <p>P7: "Emotion detection through facial expressions has low accuracy, so that's the part I don't trust. Sometimes a bug can be dangerous".</p> <p>P8: "What if the driver was always in neutral emotion or the face of the driver wasn't detected for a long time?"</p> <p>P10: "I remember I was talking when it said I was angry."</p> <p>P12: "But I didn't feel angry or fear."</p>
<i>W (1) &amp; N (2)</i>	Other Emotions	Technology Concern	P6, P9, P11	<p>P6: "And I will not be so tired and stressed with it."</p> <p>P9: "I think talking can prevent me from falling asleep when I am tired. The feedback can make me sober."</p> <p>P11: "...and will make me feel more stressed...I was a little shocked when the big truck appeared next to me in the traffic jam scenario."</p>
<i>W (2) &amp; N (1)</i>	Distraction	Efficiency	P2, P5, P7	<p>P2: "I don't want to be distracted by it."</p> <p>P5: "The movement is so complex and need to pay attention to the road situation, which makes me distracted."</p> <p>P7: "Indeed it is a little annoying during my work period, and it made me distracted...gave some information not relative to driving, which made me distracted."</p>
<i>W (2) &amp; N (1)</i>	First Notice	Efficiency	P2, P4, P10	<p>P2: "I didn't really notice it during the second video."</p> <p>P4: "I haven't noticed it."</p> <p>P10: "I just noticed the voice."</p>

<b>Group</b>	<b>Codes</b>	<b>Themes</b>	<b>Participants</b>	<b>Support</b>
<i>W (1) &amp; N (1)</i>	Specific Suggestions	Efficiency	P3, P6	P3: "I hope it can provide specific suggestions." P6: "But I think playing some music is a nice idea."
<i>W (2)</i>	Time-consuming	Efficiency	P2, P3	P2: "think immersive work would increase the reaction time." P3: "Opening questions would waste a lot of time."
<i>W (1) &amp; N (1)</i>	Misunderstanding	Efficiency	P4, P11	P4: "But the sentence 'traffic can be tough...' might cause misunderstanding." P11: "There is a misunderstanding in 'traffic can be tough...', I'm not sure if I should take over control."
<i>W (1) &amp; N (2)</i>	Reminder	Efficiency	P1, P4, P7	P1: "But it could also be a good reminder." P4: "It will remind me of every situation I should pay attention to." P7: "It reminded me about the important situations"
<i>W (2) &amp; N (1)</i>	Annoyed	Attitude	P2, P5, P7, P9	P2: "I stayed up late last night and was doing work and it always asked me questions, which would be annoying." P5: "But if it disturbs the small talk or background music, that would be a little annoying." P7: "Indeed it is a little annoying during my work period, and it made me distracted." P9: "If I'm dealing with my work, it would be annoying."

<b>Group</b>	<b>Codes</b>	<b>Themes</b>	<b>Participants</b>	<b>Support</b>
W (6) & N (6)	Trust	Attitude	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12	<p>P1: "I will trust the robot."</p> <p>P2: "Maybe if I spend more time with it, I'll trust it."</p> <p>P3: "Of course. I think it's pretty good."</p> <p>P4: "And I think that's very good."</p> <p>P5: "But I trust the robot, the information or the action."</p> <p>P6: "Yes, it drives more smoothly than me."</p> <p>P7: "I thought it was reliable."</p> <p>P8: "Always provided clear explanations ahead, which made it reliable."</p> <p>P9: "The time could be saved and I can do some make-up or reply to emails on the way."</p> <p>P10: "Of course...as long as I can get the information of status."</p> <p>P11: "Not so highly trust actually."</p> <p>P12: "Yes."</p>
W (3) & N (2)	Willingness to Chat	Attitude	P1, P2, P6, P7, P11	<p>P1: "I don't think it's an intelligent living creature like us, so I don't desire to communicate with it."</p> <p>P2: "But if I'm tired or bored, it might be a good idea."</p> <p>P6: "And if I have nothing to do, I would like to chat with it."</p> <p>P7: "I hope to talk with it if I have nothing to do."</p> <p>P11: "I don't think I would have time to respond."</p>
W (1) & N (3)	Noise	Prototype	P1, P2, P5, P9	<p>P1: "Maybe you should balance the noise and range of gestures."</p> <p>P2: "I don't like the noise when overtaking."</p> <p>P5: "The noise of movement in overtaking."</p> <p>P9: "And the noise makes me uncomfortable."</p>
W (1) & N (1)	Appearance	Prototype	P1, P10	<p>P1: "is this the ideal size? And should it be fully transparent?"</p> <p>P10: "And the size of the prototype seems a little bigger."</p>

<b>Group</b>	<b>Codes</b>	<b>Themes</b>	<b>Participants</b>	<b>Support</b>
<i>W (2) &amp; N (2)</i>	Interactions	Experiment	P5, P7, P9, P12	P5: "Curious about what it will talk about and the meaning of movement...The movement is so complex and need to pay attention to the road situation, which makes me distracted." P7: "It will first tell me what happened and then take action. And all of the explanations are clear...The movement in overtaking is really interesting, can made me feel that it will accelerate soon." P9: "I cannot understand the facial expressions of the robot, but I can tell they are different somehow. So I don't know what it was doing." P12: "But I hope it can provide more information."
<i>W (1) &amp; N (2)</i>	Scenarios	Experiment	P6, P8, P10	P6: "But there's one time I want to takeover control and overtake the truck. I'm afraid of it." P8: "But sometimes it follows the front car too close, which is contradictory to my driving habits." P10: "I think the position of this robot is beyond my perspective scope."

## 5 Discussion

The results of the study 1 demonstrated notable findings: (1) Both facial expressions and gestures effectively reduced driver workload at 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 signals, causing participants to split their attention, resulting in higher Temporal demand scores; (2) 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; (3) Interviews revealed that participants who preferred gestures found them helpful as pre-alerts 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; (4) 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 *Study1*, 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.

In *Study2*, here are notable results: (1) Although the within-subjects factor emotional feedback (E/B) did not have a significant effect on the average weighted workload measured by NASA TLX ( $P=0.057$ ), the statistical analysis (see supplementary material) indicated that the emotional feedback factor showed a significant influence on specific measures, particularly the dimension of physical demand ( $F(1,10) = 5.052$ ,  $p = 0.048$ ) and temporal demand ( $F(1,10) = 5.213$ ,  $p = 0.046$ ), but their impact varies depending on the nature of the task and the working status. Additionally, the interaction between emotional feedback and working status highlights the potential combined effects on task load, underscoring the importance of considering both factors in SAE Level 3 AD scenarios. However, due to the small sample size ( $N=12$ ), the study results may be more susceptible to individual variability, reducing the stability of the data. (2) Within the given experimental conditions, emotional feedback and working status did not play a significant role in shaping user acceptance levels. Interestingly, Task *E* of both groups had higher scores in the dimension of Sleep-inducing - Raising Alertness than those of Task *B*, which means that emotional feedback can make the driver more alert and sober, also supported by P9 in the interview. This may apply to some specific situations. For example, the driver is tired or stressed, but AD cannot be enabled. (3) Responses and attitudes to emotional feedback are different due to different situations and individual preferences. Most felt annoyed (P2, P5, P7, P9) when working while others would like to chat with it when available (P2, P6, P7, P11). (4) The primary way to convey information to the driver is still through voice interaction; Most participants could clearly remember the dialogue, but could hardly tell the behaviour of the IVA robot. Some participants admitted this idea (P2, P4, P10). So, the emotional connection could also be through the intonation of the voice and spoken words [29]. (5) Tiredness and stress are rather physical conditions than emotions, so it would be difficult to detect and we need other methods to measure (since the precision of facial expression detection is quite low during the experiment). However, these are typical physical reactions that indicate that the driver is not suitable to drive, which should be taken into consideration. (6) Even though deepface [38] was a powerful framework for recognising facial expressions, it did not execute the right output most of the time. Possible reasons are inappropriate parameters that cannot recognise complex facial expressions like yawns, or different colours of participants' clothes that interfere with the background colour (since sometimes no one is in front of the camera but it still produces fear or angry). Also, if facial recognition detects emotions, which are strongly influenced by light, it probably doesn't work when driving at night. A possible solution might be to combine different ways of emotion detection [39] or input modalities [40].

### 5.1 Limitations and Future Work

In both *Study1* and *Study2*, the RCA TV was used to display the video to simulate driving scenarios, which could have caused errors, as participants could not feel the acceleration and deceleration of the vehicle. Furthermore, both of the experiment process is Wizard of Oz, which means that the video would not change if the participants tried to take over the control, and the driving route is the same.

The limitations of *Study1* included the use of video simulations instead of real driving scenarios and the variability in participation in secondary tasks between participants, which affected IVA perception.

When asked if they trusted the IVA robot during the experiment of *Study2*, P2 and P8 mentioned that they were not in the real car and knew they were safe, so they trusted the IVA robot (see the supplementary material). In addition, the prototype is larger than the ideal size (head diameter around 15cm), which might cause errors in the evaluation. Due

to the size of the soft LED and the two servomotors, if the size is smaller, the different parts will interfere with each other. However, for industrial production, all components could be customised so that the mechanical structure could be smaller but more meticulous, just like the Nomi [31]. The gear transmission could also be replaced by magnetic levitation and momentum wheels. Furthermore, the emotive driving concept is cool but abstract, and the feedback of the IVA (vehicle system) is based on what data it received rather than what happened in reality, which is the reason why all the dialogues of emotional feedback only provide general suggestions or comforting words.

Future research of *Study1* 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.

Future studies of *Study2* could improve the reliability and external validity of the research results by increasing the sample size, testing in real AD scenarios, or conducting a larger controlled trial. Besides, there are only positive or neutral expressions for emotional feedback because when drivers' negative emotions are detected, it is not reasonable for the IVA robot to show a negative expression. As P9 mentioned, sometimes a bug can be dangerous. But maybe the IVA could also act like a weak robot [33] to ask for help in order to meet everyone's need to care for someone else, which might be interesting to research in the future. Furthermore, based on the feedback of being shocked by the big truck or contradictory driving habits (P6, P8) and the reaction of participants during the experiment, emotion detection can be combined with machine learning to form a customised IVA robot (AD system) according to their own driving habits. For example, if the driver had a surprised expression each time driving parallel to a truck, it would prevent them from getting close to the truck in the future. The same idea can be applied to the speed limit and the following distance. Finally, there will be a unique IVA robot for each customer, which is exactly the profile part of the input modalities used to recognise the states and intentions of the driver or passenger [40].

## 6 Conclusions

In summary, in this paper, two complementary studies were conducted on the development and evaluation of robot-like IVAs equipped with multimodal interaction and emotional feedback capabilities. Two prototypes were developed, and the prototypes were evaluated through the corresponding experiments to answer the two proposed research questions. For the first research question, it indicated that 7 of 12 participants preferred gestures due to their practicality and earlier signal timing, while the remaining 5 participants for their emotional and aesthetic appeal. For the second research question, results showed that neither emotional feedback nor work status has a significant impact on the average weighted workload in AD scenarios. Emotional feedback might have an impact on the dimension of Physical and Temporal Demand, and the interaction effect between emotional feedback and work status was significant for the average weighted workload. However, more experimental samples are still needed to confirm and refine the above conclusions. However, voice is still the main interaction between the IVA robot and the driver, especially when the driver is working. The results of the experiment also revealed the challenges of using facial recognition to detect emotions in the driving scenario, other physical conditions like tiredness or stress that need to be taken into consideration, and participants' perspectives towards the IVA robot.

## 7 Supplementary Material

Surveys, STL files, analysis, materials used in the experiment, and raw data can be found at: <https://www.dropbox.com/scl/fo/8xz3ok1s4zsagf7nytky5/AJQPeMbzQAZ8ncz3LqjfQ?rlkey=ge4ro37wsj6f21v19fty95ydv&st=22886l0r&dl=0>.

The maintained source code for the hardware with detailed operational instructions is available at <https://github.com/esse009/emotion-face>.

## References

- [1] Arduino. 2025. *Software – Arduino*. <https://www.arduino.cc/en/software> Accessed: 2025-09-22; current version as of IDE 2.3.6 :contentReference[oaicite:0]index=0.
- [2] Arduino. n.d. *Adafruit GC9A01A – Arduino Reference*. <https://www.arduino.cc/reference/en/libraries/adafruit-gc9a01a/> Accessed: 2025-09-22.
- [3] Arduino. n.d. *TFT\_eSPI – Arduino Reference*. [https://www.arduino.cc/reference/en/libraries/tft\\_espi/](https://www.arduino.cc/reference/en/libraries/tft_espi/) Accessed: 2025-09-22.
- [4] Apollo (Baidu). 2020. *Xiaodu in-vehicle OS*. [https://developer.apollo.auto/platform/dueros\\_cn.html](https://developer.apollo.auto/platform/dueros_cn.html) 2025-09-22.
- [5] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace: A Facial Behavior Analysis Toolkit. GitHub repository. <https://github.com/TadasBaltrušaitis/OpenFace> Released 2018; accessed: 2025-09-22.
- [6] Matthias Beggiamo, 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*. doi:10.13140/RG.2.1.2462.6007
- [7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa
- [8] 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. doi:10.1145/3610977.3634998
- [9] Dasai. 2025. *Mochi3*. <https://dasai.com.au/pages/mochi3> Accessed: 2025-09-22.
- [10] FFmpeg Developers. n.d. *FFmpeg - Official Multimedia Framework*. <https://www.ffmpeg.org/> Accessed: 2025-09-22.
- [11] Rockstar Games. n.d. *Grand Theft Auto V*. <https://www.rockstargames.com/zh/gta-v> Accessed: 2025-09-22.
- [12] 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* (Boston, Massachusetts) (IDC '15). Association for Computing Machinery, New York, NY, USA, 359–362. doi:10.1145/2771839.2771916
- [13] Paul Ekman Group. n.d. *Universal Emotions: What are Emotions?* <https://www.paulekman.com/universal-emotions/> Accessed: 2025-09-22.
- [14] V3ry H1gh. 2023. How To Install Dynamic Vehicle First Person Camera - GTA V Mods. YouTube video. <https://www.youtube.com/watch?v=jwxgmAHtwIY> Accessed: 2025-09-22.
- [15] 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* (Winnipeg, Manitoba, Canada) (AH '20). Association for Computing Machinery, New York, NY, USA, Article 10, 8 pages. doi:10.1145/3396339.3396404
- [16] 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. doi:10.1016/S0166-4115(08)62386-9
- [17] Renate Häuslschmid, Max von Bülow, Bastian Pfleging, and Andreas Butz. 2017. SupportingTrust in Autonomous Driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) (IUI '17). Association for Computing Machinery, New York, NY, USA, 319–329. doi:10.1145/3025171.3025198
- [18] IBM Documentation Help. 2024. IBM SPSS Statistics 30.0.0. <https://www.ibm.com/docs/en/spss-statistics/30.0.0>
- [19] Md Naeem Hossain. 2024. Artificial Intelligence Revolutionising the Automotive Sector: A Comprehensive Review of Current Insights, Challenges, and Future Scope. *Challenges, and Future Scope (December 02, 2024)* (2024). doi:10.32604/cmc.2025.061749
- [20] Adafruit Industries. 2012. Adafruit GFX Library. GitHub repository / Adafruit Learn Guide. <https://github.com/adafruit/Adafruit-GFX-Library> “Overview” published July 29, 2012; latest release 1.12.2 on Sep 16, 2025 :contentReference[oaicite:0]index=0.
- [21] A.K. Jonsson and N. Dahlbäck. 2009. Impact of Voice Variation in Speech-Based In-Vehicle Systems on Attitude and Driving Behaviour. In *Proceedings of the European Conference on Human Factors in Transport*. doi:10.11470/jsaprev.220409
- [22] Wendy Ju and Larry Leifer. 2008. The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious. *Design Issues* 24, 3 (2008), 72–84. doi:10.1162/desi.2008.24.3.72
- [23] 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. doi:10.1007/s12008-014-0227-2
- [24] 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. doi:10.1145/3004323.3004338
- [25] 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. doi:10.1049/iet-its.2017.0201
- [26] 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. doi:10.1016/j.ijhcs.2022.102864

- [27] 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. doi:10.1016/j.apergo.2020.103098
- [28] Dry Cactus Limited. 2017. Poly Bridge (App). App Store for iOS. <https://apps.apple.com/us/app/poly-bridge/id1197552569> Requires iOS 9.0 or later; accessed: 2025-09-22.
- [29] Albert Mehrabian. 2017. Communication without Words. In *Communication Theory*. Routledge, 193–200. doi:10.4324/9781315080918-15
- [30] NASA. 2018. NASA TLX (App). App Store for iOS. <https://apps.apple.com/us/app/nasa-tlx/id1168110608> Requires iOS 10.0 or later; accessed: 2025-09-22.
- [31] NIO. 2017. NOMI. YouTube video. <https://www.youtube.com/watch?v=SAZ2Dd9lrVc> Accessed: 2025-09-22.
- [32] NIO. 2021. *Hi, NOMI Speaks English*. <https://www.nio.com/blog/hi-nomi-speaks-english> Blog post by NIO.
- [33] Michio Okada. 2022. Weak robots. *JSAP Review* 2022 (2022), 220409. doi:10.11470/jsaprev.220409
- [34] Inc. Play.ht. 2019. *Play.ht – AI Voice Generator & Text-to-Speech Platform*. <https://play.ht/> Founded in 2019; Accessed: 2025-09-22; Headquarters: Palo Alto, California; Offers AI voices, text-to-speech API, voice cloning etc..
- [35] 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. doi:10.3390/mti2040062
- [36] 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](https://www.sae.org/standards/content/j3016_202104)
- [37] Ulrica Segerstråle and Peter Molnár (Eds.). 1997. *Nonverbal Communication: Where Nature Meets Culture*. Lawrence Erlbaum Associates, Mahwah, NJ. doi:10.4324/9781351243131
- [38] Sefik Ilkin Serengil and Alper Ozpinar. 2025. *deepface: A Lightweight Face Recognition and Facial Attribute Analysis Library for Python*. <https://github.com/serengil/deepface> Version 0.0.95, MIT License; accessed: 2025-09-22.
- [39] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. 2015. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing* 7, 1 (2015), 17–28. doi:10.1109/TAFFC.2015.2436926
- [40] Annika Stampf, Mark Colley, and Enrico Rukzio. 2022. Towards Implicit Interaction in Highly Automated Vehicles-A Systematic Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–21. doi:10.1145/3546726
- [41] Espressif Systems. 2025. Arduino core for the ESP32. GitHub repository. <https://github.com/espressif/arduino-esp32> Latest stable release 3.3.1 (as of Sep 16, 2025); LGPL-2.1 license; accessed: 2025-09-22.
- [42] Hiroko Tanabe, Yuki Yoshihara, Nihan Karatas, Kazuhiko Fujikake, Takahiro Tanaka, Shuhei Takeuchi, Tsumeyuki 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. doi:10.1007/978-3-031-06086-1\_9
- [43] Raspberry Pi Documentation Team. n.d.. Remote Access — Raspberry Pi Documentation (Computers / Remote Access). GitHub repository, documentation branch. <https://github.com/raspberrypi/documentation/tree/develop/documentation/asciidoc/computers/remote-access> Accessed: 2025-09-22.
- [44] 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. doi:10.1016/S0968-090X(96)00025-3
- [45] Manhua Wang, Seul Chan Lee, Harsh Kamlesh 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. doi:10.1145/3409118.3475142
- [46] 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. doi:10.1145/3543174.3546830
- [47] Willez. 2020. GTA 5 - How To Install Enhanced Native Trainer (2020). YouTube video. <https://www.youtube.com/watch?v=UHHXTh0Xdow> Accessed: 2025-09-22.
- [48] 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. doi:10.1109/IVS.2013.6629497
- [49] Jens Zihsler, Philipp Hock, Marcel Walch, Kirill Dzuba, Denis Schwager, Patrick Szauer, and Enrico Rukzio. 2016. Caravatar: 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. doi:10.1145/3004323.3004354