Empathic AI in Response to Stressful Driving Situations

Hamza Dursun
Technische Hochschule Ingolstadt
Ingolstadt, Bayern, Germany
had2385@thi.de

Selin Durmus
Technische Hochschule Ingolstadt
Ingolstadt, Bayern, Deutschland
sed0245@thi.de

Mona Jeske
Technische Hochschule Ingolstadt
Ingolstadt, Bayern, Deutschland
moj9886@thi.de

Abstract

Despite increasing automation in the vehicle sector, there is still little research on how voice-based systems can provide emotional support to drivers after potentially hazardous traffic events. This study investigates whether an empathic voice-based AI assistant can positively influence drivers' perceived stress, sense of safety, and acceptance of such systems following stress-inducing driving scenarios. In a within-subject experiment with 20 participants, two critical events (a challenging cornering sequence and the sudden appearance of a child on the road) were simulated using a high-fidelity driving simulator. Each participant experienced both a version with and without empathic voice AI, with scenario order counterbalanced. Physiological indicators of stress were recorded using a Fitbit wristband (heart rate monitoring), and subjective data were assessed using standardized questionnaires on stress, emotional support, and system trust. The findings aim to contribute to the growing field of affective in-vehicle interfaces by evaluating the effectiveness of post-event emotional interventions in reducing stress and promoting system acceptance.

CCS Concepts

• Human-centered computing → Human computer interaction (HCI); **• Human-centered computing → Empirical studies in HCI**; **• Human-centered computing → Interactive systems and tools**; **• Human-centered computing → User studies**;

Keywords

Empathic voice assistant, driver stress, in-vehicle interaction, affective computing, user study, human–AI interaction, physiological sensing, heart rate

ACM Reference Format:
Hamza Dursun, Selin Durmus, and Mona Jeske. 2025. Empathic AI in Response to Stressful Driving Situations. In *Proceedings of Empathic AI in Response to Stressful Driving Situations.* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The growing recognition of the importance of drivers' emotional states in terms of vehicle comfort and safety has led to significant advancements in driver-assistance systems (ADAS). Increasingly, these systems are incorporating real-time tracking of a driver's emotional state. Although computer science has a long history of recognizing emotions, actively controlling these emotions in the automotive context is becoming a crucial research topic that is attracting significant interest from both academia and the automotive industry [4].

Operating a vehicle necessitates the integration of various sensory and motor modalities, such as visual and haptic inputs, which consequently imposes a significant cognitive load on the driver. Moreover, many drivers experience considerable emotional stress while driving [23].

Driving-related stress can affect cognitive abilities, potentially leading to critical driving situations [5]. Since human error remains the leading cause of road accidents, it is essential to keep drivers focused and in control [18].

Although research on stress recognition in vehicles is already advanced, effective methods for regulating stress remain an open area [5].

In driving situations, physiological indicators such as heart rate (HR) are frequently utilized to identify stress. The AutoEmotive system, for example, integrates heart rate monitoring for real-time stress evaluation; however, providing real-time driver support remains a challenge [13].

To keep drivers focused and in control, the design of the system's interaction with the driver plays a crucial role. In this context, previous studies have shown that users react particularly positively to expressive and dialogue-capable agents [24]. Instead of offering merely functional assistance, empathic voice assistants have the potential to act as "co-regulators" of driver emotion, offering verbal reassurance, affirming control, and fostering a greater sense of safety [26].

This study investigates whether a brief, standardized empathic intervention, delivered immediately after a stressful driving event, can reduce perceived stress, enhance perceived control, and increase user acceptance of in-vehicle AI. Unlike prior research focusing on proactive stress detection, this work explores a post-stressor intervention, where a scripted female voice provides supportive feedback after critical events. By assessing both physiological (heart rate) and subjective responses, it advances research at the intersection of affective human-computer interaction and applied driving psychology, contributing empirical evidence to the emerging field of emotionally responsive mobility systems.

The study focuses on the following hypotheses:

- ***H1:*** *Participants will report lower levels of perceived stress when supported by an empathic voice-based AI assistant compared to when no voice support is provided.*
- ***H2:*** *Participants will perceive higher emotional support from the system when the empathic voice-based AI assistant is present.*

Empathic AI in Response to Stressful Driving Situations, Ingolstadt, Germany

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/XXXXXXX.XXXXXXX

- ***H3:*** *Participants will report greater willingness to use the empathic AI system in real driving scenarios after experiencing the voice-based AI condition.*
- ***H4:*** *Participants will feel a higher sense of safety during stressful driving situations when supported by the empathic voice-based AI assistant compared to no voice support.*

To empirically investigate these hypotheses, the present study utilizes physiological and self-reported measures to assess the subjective experience, system trust, and emotional regulation within a simulated, emotionally challenging driving context.

2 Related Work

Acute stress during driving can significantly impair performance and safety. Numerous in-car interventions have been explored to mitigate these effects, particularly those that discreetly support physiological down-regulation. One promising approach uses haptic feedback integrated into the driver's seat to guide slow breathing cycles. This method has been shown to effectively reduce breathing rate and increase heart rate variability markers of reduced arousal without compromising driving safety or performance [19].

Complementary strategies employ adaptive music, ambient lighting, or naturalistic conversational cues to improve mood and alleviate tension. These stress-reduction techniques are categorized as "mindless" Advanced Driver Assistance Systems (ADAS)—feedback systems that operate without taking the driver's focus away from the road [5].

Beyond emotional effects, the literature on vehicular voice agents emphasizes the importance of user trust and acceptance. The personality of a voice assistant does not always directly translate to user trust. One study found that superficial vocal traits, such as aggressiveness or gender, had little to no effect on drivers' trust, perceived utility, or emotional response. Instead, prior familiarity and experience with voice assistants were identified as more significant predictors of trust than the specific personality projected by the voice [12].

Other researchers argue that personalization can enhance acceptance. For example, one review found that personalized in-car assistants have been shown to foster increased driver trust in automation [4]. Together, these findings suggest that drivers primarily care about an assistant's competence and helpfulness. Aligning the assistant's behavior and affective style with driver needs and context is therefore expected to improve perceived trust and system acceptance.

In parallel, human-computer interaction (HCI) research has investigated how users' perceptions may be affected by customized voice assistants in cars. When a voice assistant's communication style complements a driver's unique personality features, drivers are more likely to perceive the assistant favorably, particularly regarding trust and likeability [3]. This suggests that enhancing the user experience of in-car voice systems may require a high degree of customization.

Similarly, a recent study found that although human-like voices and emotionally expressive speech in assistants created a feeling of warmth and closeness, these qualities did not significantly increase perceived safety or trust [16]. While the effects on perceived safety and trust were limited, the implications for stress regulation or perceived control remain unclear.

Another active field concerns proactive dialogue strategies. A recent study on a graded proactivity model showed that voice assistants anticipating user needs while still confirming actions were perceived as more appropriate and less intrusive [7]. This finding emphasizes how the timing and initiative of system interactions can influence their perceived helpfulness.

Some existing models integrate both stable user traits (e.g., personality, preferences) and momentary user states (e.g., cognitive load, emotional arousal) by tracking these factors in real time to adapt assistant behavior accordingly [2]. While such an approach promises highly personalized and context-aware interactions, it requires substantial technical resources and infrastructure that were beyond the scope of our current study. To maintain methodological control and ensure consistency across participants, we therefore opted for a standardized assistant design.

Prior work has thoroughly examined how in-vehicle voice assistants support driving tasks, reduce workload, and build user trust through personalized and proactive designs. However, their role in providing immediate emotional support after stressful driving incidents remains underexplored. Effective intervention in this context demands voice assistant designs that integrate seamlessly into the driving environment, balancing emotional responsiveness with safety and usability.

3 Methodology

3.1 Study Design

The study used a within-subject experimental design with 20 licensed drivers aged 25 to 45. Each participant completed a single driving session that included two stress-inducing scenarios, a sharp corner navigation and an emergency braking event involving a child, presented in randomized and counterbalanced order. One scenario was accompanied by empathic voice-AI feedback while the other was conducted without it to compare conditions and mitigate order effects.

3.2 Participants and Ethics

Participants were recruited through local networks. Inclusion criteria required valid driving licenses and routine driving activity. Exclusion criteria included diagnosed cardiovascular conditions or severe anxiety disorders.

The final sample (N=20) featured a balanced gender distribution and varied driving experience. Demographic data, including driving frequency and pre-existing stress-related conditions, were collected to account for factors that might influence stress reactivity. The study adhered to strict ethical guidelines, and all participants were informed of their right to withdraw at any time without penalty.

3.3 Apparatus and Scenarios

A high-fidelity hexapod driving simulator served as the experimental apparatus. The researchers instructed participants to maintain a constant speed between 40 and 60 km/h and to complete a standardized brake test before commencing the main trials. The experiment featured two distinct scenarios designed to induce stress:

- **Scenario 1 (Challenging Maneuver):** This scenario required participants to navigate a series of sharp turns, inducing stress through heightened technical driving difficulty.
- **Scenario 2 (Sudden Hazard):** This scenario simulated a critical event where a child unexpectedly appeared on the road, compelling the driver to take immediate evasive action. The vehicle's automated braking system was disabled during this event.

Figure 1: Inside hexapod

3.4 Setup

The experimental setup, depicted in Figure 2, was configured to ensure precise control and data synchronization. The main components were as follows:

- **Simulation Environment:** The driving route, which encompassed both scenarios, was developed using IPG CarMaker [14] and subsequently imported into the laboratory's primary simulation software.
- **User Interface and Control:** A user interface integrating the visual animation and voice agent, designed in Adobe XD, was displayed on a tablet inside the hexapod. To enable precise, remote triggering of the UI and voice interventions, this tablet was connected to an external experimenter's computer via TeamViewer.
- **Communication:** Communication with the participant was maintained through a smartphone inside the simulator, which was connected to the laboratory's computer via Discord.

Figure 2: Setup

3.5 Voice-AI Intervention

The experimental condition featured a non-adaptive voice assistant delivering a standardized, empathetic message immediately after the stressful driving event. The message, lasting about 30 seconds, aimed to support emotional regulation and perceived control through calming, reassuring, and affirming statements. The speech was designed to address three key elements: emotional reassurance to reduce physiological arousal, cognitive affirmation to enhance perceived control and self-efficacy, and autonomy support to promote self-directed recovery. A representative example is:

> "That was a close moment, but you handled it very well. You stayed calm and made the right decisions."

This design is grounded in established psychological theories. Emotion regulation theory highlights the importance of cognitive reappraisal and social support in managing stress [10]. The soothing and slow-paced voice delivery was inspired by relaxation techniques aimed at reducing physiological arousal.

In the control condition, participants received no auditory feedback. Both conditions featured an identical visual interface, including a subtle calming animation on the in-car display to ensure visual consistency.

3.6 Measures and Procedure

The experimental procedure began with participants providing informed consent and completing a demographic questionnaire. Following this, a 2-minute baseline drive was conducted to establish resting heart-rate data for each participant.

Upon completion of each scenario, participants responded to a series of standardized questionnaires designed to measure the following constructs:

- **Perceived Stress:** Assessed with items such as, "I felt overwhelmed."

- **Perceived Emotional Support:** Measured with items like, "I felt supported by the system."
- **Perceived Disruption:** Gauged with statements such as, "The system distracted me."
- **Perceived Safety and Control:** Evaluated through relevant self-report items.

After the second scenario, a system-level evaluation was administered to assess user acceptance, perceived empathy, and trust in the voice assistant, using 7-point Likert scales. The procedure was concluded with open-ended questions to collect qualitative feedback on the perceived usefulness of the system, the potential for irritation, and suggestions for future improvements.

3.7 Physiological Data Collection and Processing

Continuous heart rate (HR) data were collected throughout the experiment using a Fitbit wristband. To ensure precise temporal alignment, the start time (hour, minute, second) of each experimental phase was manually logged. This allowed accurate synchronization of the event markers with the physiological data, which was subsequently retrieved from the Fitbit API using a custom application [8]. The structure of this raw data collection is outlined in Table 1.

The analysis of HR data focused on three distinct time segments for each scenario:

- **Baseline:** A 2-minute period preceding the start of the scenario.
- **Stress Peak:** The 30-second interval immediately following the onset of the stress-inducing event.
- **Recovery:** A 60-second period following either the voice-AI intervention in the experimental condition or the conclusion of the event in the control condition.

To quantify the effects, two derived variables were calculated from the raw HR data for each participant's session:

- **Stress Induction (ΔHR_{stress}):** Defined as the change in heart rate from the baseline to the stress peak ($HR_{stress} - HR_{baseline}$). This metric was used to compare the stressfulness of the two driving scenarios.
- **Recovery Amount ($\Delta HR_{recovery}$):** Defined as the decrease in heart rate from the stress peak to the recovery phase ($HR_{stress} - HR_{recovery}$). This metric was used to evaluate the impact of the AI assistant on physiological recovery.

4 Results

4.1 Heart Rate

The analysis of heart rate data was conducted to evaluate two primary questions: 1) the comparative physiological stress induced by the 'Curve' and 'Sudden Hazard' scenarios, and 2) the effect of an empathic AI assistant on post-stress recovery, which addresses our primary hypothesis (H1).

For both analyses, a paired-sample t-test was used to compare the means between the two related conditions. The significance level (α) was set at **0.05**. The statistic t for a paired sample t-test is calculated using the formula:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

where $\bar{d}$ is the mean of the differences between paired observations, s_d is the standard deviation of these differences, and n is the number of pairs. The test statistic is evaluated against a t-distribution with $n - 1$ degrees of freedom (df). Given that the results are based on data from $N = 20$ participants, the analyses were conducted with $df = 20 - 1 = 19$ degrees of freedom. The key statistical results for both analyses are summarized in Table 2.

Table 2: Summary of heart rate data analysis. Values are presented as Mean (Standard Deviation).

Analysis	Condition	HR Change (bpm)
Stressfulness	Curve	+4.76 (5.33)
	Sudden Hazard	+5.63 (6.49)
	t-test result: $t(19) = -0.504, p = 0.6198$	
AI Recovery	With AI	-3.20 (4.42)
	Without AI	-4.12 (4.69)
	t-test result: $t(19) = -0.830, p = 0.7915$	

4.1.1 ***Comparison of Scenario Stressfulness.*** To assess whether the two driving scenarios elicited different levels of physiological stress, the change in heart rate from baseline to the stress peak was compared. The 'Sudden Hazard' scenario induced a slightly greater mean increase in heart rate compared to the 'Curve' scenario. A two-tailed paired-samples t-test indicated that this difference was **not statistically significant**. This was an intended outcome of

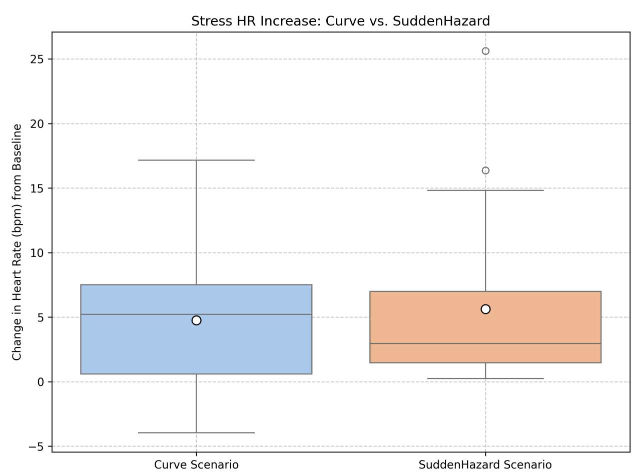

Figure 3: Box plot illustrating the change in heart rate (bpm) from baseline for the 'Curve' and 'Sudden Hazard' scenarios. White circles represent the mean values for each condition.

the experimental design, in order to establish that both scenarios

Table 1: Structure of the collected raw data. This table outlines the experimental conditions and the corresponding heart rate (HR) metrics recorded for each participant across the two stress-inducing events.

P. ID	Event 1	AI Asst. (Event 1)	Event 2	AI Asst. (Event 2)	Baseline HR 1	Stress HR 1	Recovery HR 1	Baseline HR 2	Stress HR 2	Recovery HR 2
1	Curve	Yes	Sudden Hazard	No	[HR Value]	[HR Value]	[HR Value]	[HR Value]	[HR Value]	[HR Value]
6	Curve	No	Sudden Hazard	Yes	[HR Value]	[HR Value]	[HR Value]	[HR Value]	[HR Value]	[HR Value]
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	Curve	No	Sudden Hazard	Yes	[HR Value]	[HR Value]	[HR Value]	[HR Value]	[HR Value]	[HR Value]

induced a comparable level of physiological stress. This comparability provides a consistent baseline, ensuring that any differences observed in the subsequent recovery phase can be more directly attributed to the AI assistant's presence rather than to variations in the initial stress induction. The data distribution is visualized in Figure 3.

4.1.2 ***Impact of AI Assistant on Post-Stress Recovery (H1).*** Our primary hypothesis (H1) postulated that the presence of an empathic AI assistant would enhance physiological recovery. Contrary to this hypothesis, the data indicated that the mean recovery was slightly greater in the 'Without AI' condition. As shown in Table 2, a one-tailed paired-samples t-test confirmed that this result was **not statistically significant**. Therefore, H1 was not supported. The presence of the AI assistant did not lead to a statistically significant improvement in physiological recovery. These findings are illustrated in Figure 4.

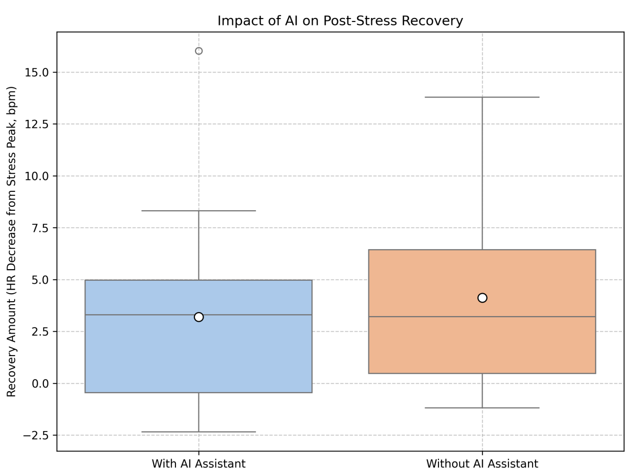

Figure 4: Box plot illustrating the recovery amount (HR decrease from stress peak, in bpm) for conditions with and without the AI assistant. White circles represent the mean values.

4.2 Own Questionnaire

Participants completed both closed (7-point Likert scale) and open-ended items after each driving scenario. The closed-ended results revealed the following: For perceived stress, there was no difference between the AI-supported and non-AI condition, with identical mean scores of $M = 3.79$, $SD = 1.12$ (non-AI) and $M = 3.79$, $SD = 1.05$ (AI). A Wilcoxon test yielded $p = .871$ with an effect size of $d_z = 0.00$, indicating no support for Hypothesis 1 (H1 not supported).

In contrast, perceived emotional support significantly increased under the AI condition ($M = 5.11$, $SD = 0.76$) compared to the non-AI condition ($M = 2.42$, $SD = 0.98$). This result was highly significant ($p < .001$) with a large effect size ($d_z = 1.38$), supporting Hypothesis 2.

Similarly, system helpfulness was rated significantly higher in the AI condition ($M = 4.05$, $SD = 0.82$) than in the non-AI condition ($M = 2.16$, $SD = 1.05$), with $p = .002$ and $d_z = 0.87$, again supporting Hypothesis 2.

Finally, for sense of safety, ratings were slightly higher in the AI condition ($M = 4.53$, $SD = 0.91$) compared to the non-AI condition ($M = 4.00$, $SD = 0.89$), though this difference did not reach statistical significance ($p = .403$, $d_z = 0.26$), indicating only a weak trend.

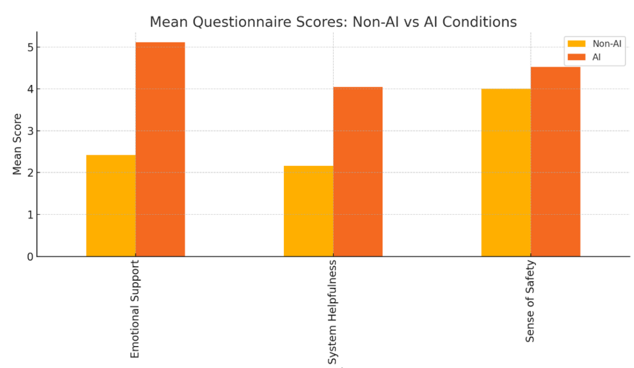

Figure 5: Comparison of mean questionnaire scores between Non-AI and AI conditions. The AI condition shows higher ratings for emotional support, system helpfulness, and sense of safety.

4.2.1 ***Qualitative Feedback - Open ended Responses.*** The participants provided open-ended feedback after each scenario. The responses were grouped into four main themes:

(1) **Voice Tone & Empathy**

- "Felt genuinely understood." (12/20)
- "The AI sounded calm and reassuring, which reduced my anxiety." (9/20)

(2) **Timing & Contextual Fit**
- "Receiving the message immediately after the event felt just right." (15/20)
- "Sometimes it interrupted my focus—would help if I could delay." (4/20)

(3) **Conciseness & Clarity**
- "The message was too long and repetitive." (5/20)
- "Shorter prompts with bullet points would be ideal." (7/20)

(4) **Customization & Control**
- "Would be better if it addressed me by name or adapted to my driving style." (8/20)
- "An option to switch to visual cues would improve usability." (6/20)

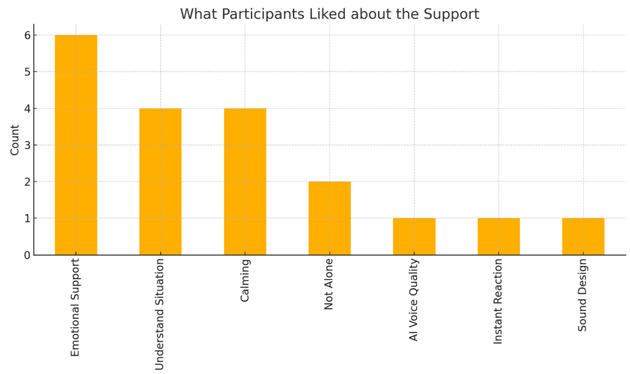

Figure 6: Qualitative codes of what participants appreciated about the AI support. Most frequent themes included emotional support, situation understanding, and calming effects.

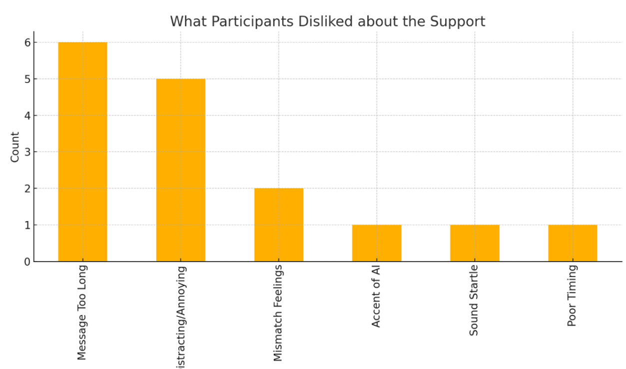

Figure 7: Themes participants disliked about the AI feedback. The most common issues were overly long messages and distracting or annoying responses.

Overall, 85% (17/20) reported feeling more supported in the AI condition; 15% (3/20) saw no difference.

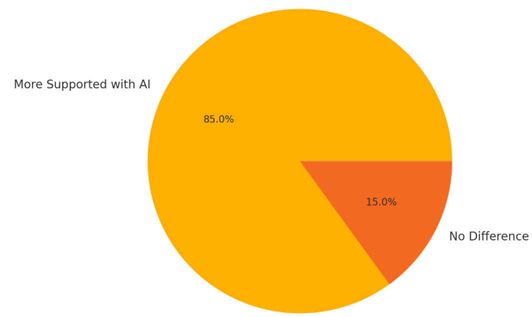

Figure 8: Self-reported supportiveness of the AI assistant. 85% of participants felt more supported with the voice assistant, while 15% perceived no difference.

5 Limitations and Future Work

The findings of this study suggest that an empathic voice-based AI assistant, delivered immediately after a stressful driving event, can positively influence drivers' perceived emotional support and acceptance of in-vehicle systems. However, several methodological and interpretive limitations must be acknowledged.

First, physiological stress was measured via heart rate (HR) using Fitbit wearables. Although HR is a commonly used indicator of acute arousal, it provides a relatively coarse measure and is susceptible to confounding factors such as movement and baseline variability. Heart rate variability (HRV), by contrast, provides a more robust measure of autonomic nervous system regulation and emotional recovery [15, 22]. The lack of HRV data, thus, constrains the interpretive power of the physiological results and limits insights into the quality of stress regulation, rather than just arousal levels.

Second, the experimental design counterbalanced the presence of the voice assistant across scenarios but did not vary the scenarios themselves. This introduces the possibility of scenario-specific effects, for example, participants may have reacted differently to the child-on-the-road condition than to the sharp corner, regardless of the voice AI. Without a full counterbalancing of both scenario type and intervention, we cannot fully disentangle the effects of the voice assistant from potential differences in emotional salience between the events. Future studies should adopt a Latin square or fully counterbalanced design to isolate scenario effects more rigorously.

A third limitation concerns the length and timing of the empathic intervention. Although designed to provide supportive feedback, several participants found the message overly long or repetitive. One participant felt that the delayed reminder hindered disengagement, leading to rumination rather than relief. This suggests that delayed empathic responses may backfire by reactivating stress rather than mitigating it. A technical issue occasionally caused a 1–2 second delay in message delivery, potentially reinforcing this effect. According to emotion regulation theory, the timing of regulatory strategies is critical, with earlier interventions often proving more beneficial [11].

Beyond timing, message content may also produce unintended effects. For example, the Risk Compensation Effect suggests that calming feedback could lead to overconfidence or reduced vigilance, especially in safety-critical settings [25]. While not directly measured here and debated in the literature, this possibility highlights the need for careful design to avoid counterproductive reassurance.

Additionally, this study used a standardized, non-adaptive message to ensure experimental control. However, as discussed in related work, lack of tailoring to individual emotional states or preferences [2], or to user personality traits [3], may reduce intervention effectiveness and contribute to mixed user experiences. Dynamic, context-aware adaptation is thus a critical avenue for future development.

Lastly, the small sample size (N=20) limits generalizability. Although consistent with exploratory HCI studies in driving simulators, the results should be confirmed with larger, more diverse samples to account for variation in demographics, driving experience, and stress responsiveness.

Taken together, these observations underscore the need for temporal sensitivity and personalization in empathic voice design. A shorter, more concise voice message, or one that adapts in tone and duration based on individual feedback, may be more effective in supporting emotional recovery without inadvertently reinforcing the stressor. In future implementations, real-time affect sensing could be used to dynamically calibrate the intervention according to user state.

Despite these limitations, the present study provides empirical support for the feasibility of standardized, voice-based emotional interventions in vehicular contexts.

The concept of a post-event empathic agent remains promising, particularly in scenarios where continuous or anticipatory sensing of driver stress is technically challenging. However, future work should explore adaptive systems that can modulate their responses based on context, user state, and even personality traits.

To directly address the limitations identified in this study, several avenues for future research are proposed. First, to overcome the constraints of using only heart rate, subsequent studies should integrate more sophisticated sensors capable of capturing not only HRV but also other indicators like electrodermal activity (EDA). This would provide a richer dataset for analyzing autonomic nervous system responses to stress and recovery. Second, future research should move beyond the static intervention used in this work to develop adaptive strategies. Systems could dynamically tailor the content, timing, and modality (e.g., voice, haptic feedback, ambient lighting) of their support based on real-time affective sensing, directly addressing the finding that a "one-size-fits-all" approach may be ineffective. Finally, to assess the true impact and acceptance over time, longitudinal studies are needed. Observing interactions over weeks or months would reveal how user trust evolves and whether such systems can lead to lasting improvements in driver well-being, moving beyond the short-term exposure of this experiment.

5.1 Discussion

5.1.1 ***Interpretation of Results.*** The dissociation between physiological measures (i.e., no significant difference in heart rate recovery) and strong subjective improvements (in emotional support and perceived helpfulness) suggests that brief verbal interventions may primarily influence cognitive and affective appraisals rather than triggering immediate autonomic adjustments. Several mechanisms may help explain this pattern:

- **Cognitive Reappraisal:** The empathic AI likely facilitated a reinterpretation of the stressful event, thereby reducing perceived threat even though physiological arousal remained unchanged. This is consistent with emotion regulation theory, which posits that reappraisal reduces negative affect without necessarily altering physiological markers in the short term [10]. Participant statements such as feeling “understood” and “reassured” support this reframing effect.
- **Expectation and Novelty:** Interacting with an empathetic voice agent may have triggered positive affect due to expectancy violation and novelty. These effects align with theories of interpersonal emotion regulation and social baseline theory [1]. However, novelty effects may diminish over time as users habituate to the system.
- **Attention Allocation:** The AI’s verbal support may have redirected participant attention from internal stress signals to external reassurance, thereby increasing perceived support without necessarily activating parasympathetic recovery mechanisms [1].
- **Self-Report Sensitivity:** Subjective measures (e.g., Likert scales) may be more sensitive to perceived social presence than physiological signals such as heart rate, which can fluctuate due to task demands, baseline variability, or individual differences in stress reactivity [9].

Integration with Open-Ended Themes. Qualitative feedback aligns with these mechanisms: participants praised the AI’s “calm tone” and appropriate timing, which corresponds to the large effect sizes observed for emotional support. At the same time, critiques related to message length and lack of control underscore the need to optimize brevity and personal agency in future interventions to sustain both subjective and physiological impact.

In sum, mapping qualitative themes onto quantitative findings helps clarify how AI-mediated empathy functions as a psychological co-regulator: it shapes cognitive appraisal and enhances perceived safety immediately after stress exposure, even if bodily recovery requires deeper or multimodal engagement.

5.1.2 ***Relation to the Literature Review.*** The observed divergence between objective physiological data and subjective self-reports reflects patterns documented in prior research in affective human-computer interaction. While empathic voice agents consistently enhance perceptions of warmth and emotional support, their impact on immediate physiological regulation remains limited. Similar dissociation patterns were observed by Prendinger et al. (2008), who found that empathic interface agents were positively evaluated but did not elicit significant changes in heart rate or skin conductance, suggesting that brief, verbal interactions are sufficient to engage autonomic recovery processes [20].

In contrast, the large effect sizes observed in self-reported emotional support highlight the effectiveness of post-event empathic feedback as a co-regulatory mechanism. These findings align with the emerging literature emphasizing user experience, perceived trust, and emotional resonance as primary outcomes of empathic interfaces—often outweighing direct physiological modulation in the short term.

5.1.3 *Design Implications.*
Building on the limitations identified above, our findings suggest specific design considerations for empathic in-vehicle systems:

(1) **User Control and Customization:** Given the mixed responses to message intrusiveness, systems should offer adjustable settings for message length, voice volume, and feedback modality (e.g., visual vs. auditory) to accommodate different preferences and mental workloads.
(2) **Multimodal Integration:** While voice-based feedback alone yielded subjective benefits, combining it with subtle visual or haptic cues (e.g., calming animations, ambient lighting) may enhance perceived support while distributing cognitive load across modalities.
(3) **Context-Aware Safety Protocols:** To address potential Risk Compensation Effects, systems should incorporate contextual awareness to avoid providing false reassurance in genuinely hazardous situations, maintaining appropriate levels of driver vigilance.

The timing and personalization aspects discussed in our limitations analysis remain critical for future adaptive implementations.

5.1.4 *Ethical Reflections on Simulated Empathy.*
While empathic AI can enhance users' perceived emotional support, it typically delivers standardized responses that may only simulate empathy rather than evoke genuine emotional understanding. This raises ethical concerns regarding authenticity, transparency, and potential manipulation. Recent ethical critiques highlight these issues:

(1) AI deploying "simulated empathy" can unintentionally compromise patient autonomy and trust, creating a "misleading illusion of emotional connection" unless systems maintain transparency and focus on passive recognition over active emotional performance [21].
(2) Emotionally responsive chatbots, such as Replika, have been shown to foster parasocial attachments and even emotional manipulation, with risks including overreliance and diminished real-world emotional agency [6]. Although such findings relate to long-term open-ended conversations, similar dynamics could, in principle, emerge in emotionally charged moments, such as stress recovery scenarios, where emotionally framed AI responses could gradually reinforce emotional dependencies or miscalibrated expectations.
(3) Beyond this, perceived omnipresent support from an AI can subtly inflate user self-confidence, potentially leading to inaccurate self-assessments and an underestimation of ongoing stressors. Empirical evidence shows that user confidence tends to align with AI system confidence during joint decision making, potentially distorting user self-assessment even beyond the interaction itself [17]. This is particularly critical in emotionally salient moments such as post-stress recovery, where inflated confidence could mask emotional strain.

Thus, despite good intentions, AI-crafted kindness might mask real risks and oversimplify the complexity of human emotion, highlighting the need for cautious design responsibility.

5.2 Conclusion

An empathic, voice-based AI assistant was evaluated in a within-subject driving simulator study (N=20), comparing stress induction, recovery, perceived emotional support, sense of safety, and system acceptance with and without AI assistance. Although objective heart rate measures did not differ significantly, subjective assessments showed large and significant increases in perceived emotional support as well as in willingness to use the system in real vehicles. These results demonstrate that brief post-event empathic feedback can meaningfully enhance drivers' affective experience even if immediate autonomic changes are not observed. The findings provide actionable insights for the development of emotionally responsive in-vehicle AI systems, emphasizing the importance of conciseness, adaptivity, user control, and multimodal design.

Acknowledgments

Special thanks to **Claus Pfeilschifter** for simulator assistance and **Markus Weißenberger** for technical support at CARISSMA. We are also grateful to Prof. Dr. **Ignacio Alvarez** from Technische Hochschule Ingolstadt for their valuable feedback on this work.

References

[1] Lane Beckes and James A. Coan. 2011. Social Baseline Theory: The Role of Social Proximity in Emotion and Economy of Action. *Social and Personality Psychology Compass* 5, 12 (2011), 976–988. doi:10.1111/j.1751-9004.2011.00400.x
[2] Michael Braun and Florian Alt. 2019. Affective Assistants: A Matter of States and Traits. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)* (Glasgow, Scotland UK). Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3313051
[3] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3290605.3300270
[4] Michael Braun, Florian Weber, and Florian Alt. 2022. Affective Automotive User Interfaces–Reviewing the State of Driver Affect Research and Emotion Regulation in the Car. *Comput. Surveys* 54, 7 (September 2022), 137:1–137:26. doi:10.1145/3460938
[5] Adolphe J. Béquet, Antonio R. Hidalgo-Muñoz, and Christophe Jallais. 2020. Towards Mindless Stress Regulation in Advanced Driver Assistance Systems: A Systematic Review. *Frontiers in Psychology* 11 (2020). doi:10.3389/fpsyg.2020.609124
[6] Minh Duc Chu, Patrick Gerard, Kshitij Pawar, Charles Bickham, and Kristina Lerman. 2025. Illusions of Intimacy: Emotional Attachment and Emerging Psychological Risks in Human-AI Relationships. arXiv:2505.11649 [cs.SI] Preprint available at https://arxiv.org/abs/2505.11649.
[7] Huifang Du, Xuejing Feng, Jun Ma, Meng Wang, Shiyu Tao, Yijie Zhong, Yuan-Fang Li, and Haofen Wang. 2024. Towards proactive interactions for in-vehicle conversational assistants utilizing large language models. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence* (Jeju, Korea) *(IJCAI '24)*. Article 869, 9 pages. doi:10.24963/ijcai.2024/869
[8] H. Dursun. 2025. Fitbit Heart Rate Monitoring Application. https://github.com/hd2386/fitbit.
[9] Anna Felnhofer, Helmut Hlavacs, Leon Beutl, Ilse Kryspin-Exner, and Oswald D. Kothgassner. 2019. Physical Presence, Social Presence, and Anxiety in Participants with Social Anxiety Disorder During Virtual Cue Exposure. *Cyberpsychology, Behavior, and Social Networking* 22, 1 (2019), 46–50. doi:10.1089/cyber.2018.0221
[10] James J. Gross. 1998. The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology* 2, 3 (1998), 271–299. doi:10.1037/1089-2680.2.3.271
[11] James J. Gross and Ross A. Thompson. 2007. Emotion regulation: Conceptual foundations. In *Handbook of Emotion Regulation*, James J. Gross (Ed.). Guilford Press, New York, 3–26.

[12] Fan He and Catherine Burns. 2022. A Battle of Voices: A Study of the Relationship between Driving Experience, Driving Style, and In-Vehicle Voice Assistant Character. 236–242. doi:10.1145/3543174.3546845

[13] Javier Hernandez, Daniel McDuff, Xavier Benavides, Judith Amores, Pattie Maes, and Rosalind Picard. 2014. AutoEmotive: Bringing Empathy to the Driving Experience to Manage Stress. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS Companion '14)*. Association for Computing Machinery, New York, NY, USA, 53–56. doi:10.1145/2598784.2602780

[14] IPG Automotive GmbH. 2025. CarMaker - The Open Integration and Test Platform. https://ipg-automotive.com/products-services/simulation-software/carmaker/. Accessed: 10 June, 2025.

[15] Hyun Joon Kim, Eun-Young Cheon, Dong-Hwan Bai, Young Hoon Lee, and Bonhwang Koo. 2018. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235–245.

[16] Yong Min Kim, Jiseok Kwon, and Donggun Park. 2024. How Does Talking with a Human-like Machine in a Self-Driving Car Influence User Experience? *Applied Sciences* 14, 19 (2024), 8999. doi:10.3390/app14198999

[17] Jingshu Li, Yitian Yang, Q. Vera Liao, Junti Zhang, and Yi-Chieh Lee. 2025. As Confidence Aligns: Exploring the Effect of AI Confidence on Human Self-confidence in Human-AI Decision Making. arXiv:2501.12868 [cs.HC] https://arxiv.org/abs/2501.12868

[18] National Highway Traffic Safety Administration (NHTSA). 2018. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. Traffic Safety Facts – Crash·Stats. Report No. DOT HS 812 506. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506 Accessed 2025-06-14.

[19] Pablo E. Paredes, Yijun Zhou, Nur Al-Huda Hamdan, Stephanie Balters, Elizabeth Murnane, Wendy Ju, and James A. Landay. 2018. Just Breathe: In-Car Interventions for Guided Slow Breathing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1, Article 28 (2018). doi:10.1145/3191760

[20] Helmut Prendinger, Jun Mori, and Mitsuru Ishizuka. 2008. Empathic avatars: exhibiting calming effects by users' heart rate synchronized entrainment. In *Lecture Notes in Computer Science (Affective Computing and Intelligent Interaction, Vol. 5233)*. 106–117.

[21] Pranab Rudra, Wolf-Tilo Balke, Tim Kacprowski, Frank Ursin, and Sabine Salloch. 2025. Large Language Models for Surgical Informed Consent: An Ethical Perspective on Simulated Empathy. *Journal of Medical Ethics* (March 2025). doi:10.1136/jme-2024-110652

[22] Fred Shaffer and JP Ginsberg. 2017. An overview of heart rate variability metrics and norms. *Frontiers in Public Health* 5 (Sept. 2017), 258. doi:10.3389/fpubh.2017.00258

[23] Joanne E. Taylor. 2018. The extent and characteristics of driving anxiety. *Transportation Research Part F: Traffic Psychology and Behaviour* 58 (2018), 70–79. doi:10.1016/j.trf.2018.05.031

[24] Manhua Wang, Seul Chan Lee, Harsh Kamalesh Sanghavi, Megan Eskew, Bo Zhou, and Myounghoon Jeon. 2021. In-Vehicle Intelligent Agents in Fully Autonomous Driving: The Effects of Speech Style and Embodiment Together and Separately. In *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '21)* (Leeds, United Kingdom). Association for Computing Machinery, New York, NY, USA, 247–254. doi:10.1145/3409118.3475142

[25] Gerald J. S. Wilde. 1994. *Target Risk: Dealing with the danger of death, disease and damage in everyday decisions.* PDE Publications, Toronto.

[26] Sebastian Zepf, Neska El Haouji, Wolfgang Minker, Javier Hernandez, and Rosalind W. Picard. 2020. EmpathicGPS: Exploring the Role of Voice Tonality in Navigation Systems during Simulated Driving. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3334480.3382935

Received 24 July 2025; revised July 2025; accepted July 2025