IHIET 2026: Vol. XX, 2026
doi: 10.54941/ahfeXXXX

Teaching Multimodal Interaction in Cars to First-time Users

Thomas Marinissen[1], Jonas Glimmann[2] and Pavlo Bazilinskyy[3]

[1]Xenics NV, Leuven, Belgium
[2]Mercedes-Benz AG, Sindelfingen, Germany
[3]Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

This study examines three variations of a proactive method for teaching multimodal gaze and gesture interactions to first-time users in the context of an SAE Level 5 automated vehicle. The three variations differed in size, placement on the screen, and whether active user input was required to receive additional information. The results of a user study involving the gesture control prototype in a driving simulator (N=30) show that the greatest variation was more effective in teaching, caused by significant differences in visibility ratings (p<0.001), size (p<0.001) and duration (p=0.001) of the pop-ups. The results show no correlation between the measured effectiveness and the preference for a specific variation. Across all variations, participants are positive toward receiving proactive teaching from their car to learn new features. We conclude that proactively teaching users novel interaction methods has the potential to improve the user experience in future vehicles.

Keywords: User Interfaces, New Users, Multimodal Interaction, Gaze Control, Gesture Control

INTRODUCTION

Cars are transitioning into automated, connected environments where users can shift attention away from driving and engage in non-driving-related tasks (NDRTs) as automation increases (Naujoks et al., 2017; Pfleging et al., 2022). At SAE level 5 (SAE, 2021), vehicles operate without driver intervention, enabling flexible seating configurations such as reclining or rear-facing layouts (Mercedes-Benz AG, 2015; Volvo Cars, 2015). These changes challenge conventional interaction concepts that assume a fixed posture and easy physical reachability of controls. Previous human factors research highlights the role of reach, posture, and display accessibility in safe interaction (Wickens et al., 2022), making alternative modalities increasingly relevant.

Automotive user interfaces (UIs) now extend beyond physical controls to include speech (Cerence, 2021), gaze (MotorTrend, 2023), and gesture input (BimmerTech, 2020), following broader advances in multimodal Human-Computer Interaction (Bourguet, 2003). While multimodal interaction can offer more natural communication, increasing modality complexity risks overwhelming users or decreasing discoverability (Lentz et al., 2018). Raising awareness and providing appropriate teaching is therefore important for improving the user experience when new modalities are introduced.

Previous work on in-vehicle interaction has explored combinations of speech and gesture (Pfleging et al., 2012), gaze and haptics (Kern et al., 2010), and multimodal interaction (Aftab, 2019; Aftab and Von der Beeck, 2022). However, many of such approaches depend on physical controls or assume fixed seating positions, which limits their suitability for future level 5 interiors. As a result, natural modalities, especially gaze and mid-air gestures, are increasingly considered promising alternatives. Research also shows that multimodal systems require dedicated onboarding to help users understand available interaction methods and avoid feature non-use (Macek et al., 2014; Novick and Ward, 2006).

Commercial developments mirror these academic trends. Gesture-controlled interfaces appear in BMW’s 7-series (BMW Group, 2019), while Li Auto’s L9 implements rear-seat gesture input (screens, 2023). Audi’s Urbansphere concept demonstrates combined gaze–gesture interaction (Audi, 2022). Similar interaction philosophies are emerging in AR ecosystems such as Apple Vision Pro and Meta Quest, potentially increasing user familiarity over time.

Teaching and Learning of New Functions in Cars

Traditionally, learning about new functions in cars was done using a conventional printed user manual and/or by an explanation given by the salesperson. Today, printed manuals are perceived as old-fashioned, difficult to navigate, frustrating, inefficient, and not sufficiently detailed (Alvarez et al., 2010; Novick and Ward, 2006). The explanation in the dealership must remain concise due to time constraints, which means that not all features of the cars are covered (Bellini et al., 2017). Users prefer to find the answers to their questions about their car online, asking people they know, in the digital version of the manual, or may even choose to keep problems unsolved (Alvarez et al., 2010; Novick and Ward, 2006).

A common approach to modernising the conventional user manual involves integrating it into the UI (Mercedes-Benz AG, 2024; Skoda Auto AS, 2023; Wee, 2021). Similar indexing and explanations can be provided as in a regular user manual, with the added benefit of using interactive video and audio components (Mercedes-Benz AG, 2023a, 2023b). Another solution, as proposed by Alvarez et al., is voice-interfaced user help (Alvarez et al., 2010). Virtual assistant technology, such as Apple Siri or Amazon Alexa, is already integrated in many modern cars from manufacturers such as Audi, BMW, Ford, General Motors, and Mercedes-Benz (Amazon, 2024; Mercedes-Benz AG, 2023c, 2018). The benefit of this approach is that voice commands do not require the driver to take their eyes off the road and are therefore safe to use for users while driving the car. Studies show that users prefer interactive teaching methods to traditional user manuals (Macek et al., 2014; Novick and Ward, 2006). Both approaches share a limitation: they require the user to actively explore features, leaving some undiscovered.

After interactive teaching, the next step to helping users find and learn new features is proactive teaching. Proactive teaching means that the car informs and instructs the user about new features by tracking which features have not yet been found and have not yet been used. This could be through audio messages or visual information pop-ups. The risk of this approach is that the user is interrupted in their activity, which can make teaching suggestions a nuisance rather than an aid. This means that it is important to consider the user's willingness to learn. This can be achieved by gently providing the right type and amount of information at an appropriate time. This is called “*nudging*” the user (Caraban et al., 2019; Stryja and Satzger, 2018; Thaler and Sunstein, 2009; Wessel et al., 2019). This technique can be used to guide people towards a desired behaviour, such as promoting healthy food (Lee et al., 2011). Few automakers already use some form of nudging to attract user attention (BMW Group, 2015; Mercedes-Benz AG, 2024).

In Mercedes-Benz C-class and S-class, nudging is employed through on-screen information pop-ups, accessible as app icons, which explain new functions (Mercedes-Benz AG, 2024, 2023a, 2023b). Users can create a personal account on various car brands to track undiscovered or unused functions. However, a drawback of nudging is that if the pop-ups are inconspicuous, users may not perceive or click them, leading to missed teaching information in the menu.

Aim of Study

This study examines the effectiveness and satisfaction of proactively teaching multimodal interaction to first-time users in cars. In the context of SAE level 5 automated vehicles (AVs), with the possibility of varied seating configurations and with many car manufacturers equipping their vehicles with large or multiple screens, the reachability of controls could prove to be an issue. Examples of cars with multiple display areas are the Porsche Taycan, Honda E, and Mercedes-Benz EQS (Caricos, 2022; Porsche AG, 2024; Top Gear Magazine, 2020). The assessed multimodal interaction, consisting of gaze and gesture input, addresses the challenges when touchscreens or physical controls are not within reach of the user. Users must be aware of and learn this novel interaction method. A user study was conducted to compare three variations of a proactive visual teaching method presented to users "*on the fly*" during regular interactions with the UI. We evaluated participants' awareness, interaction, and adoption of presented information, and their preferences among the methods.

METHOD

Participants and Apparatus

Thirty adults in Germany took part in the study during 8–17 August 2023. Participants (14 female, 16 male) were over 18, held a driver's licence, and had a mean age of 45.7 years (SD=14.7). They were recruited through an external agency and scheduled in two-hour time slots. The study received ethics approval from Eindhoven University of Technology, and all participants gave informed consent.

The experiment took place in a full-scale seating buck simulating the interior of a luxury sedan. The setup included a 3840x900 px widescreen with a custom automotive GUI, an Intel RealSense camera for gaze and gesture detection, a centre console with an Apple Magic Trackpad as an alternative input method, and an electrically adjustable driver's seat with upright and reclined presets (approximately 18 cm backward shift and ~45° backrest angle). To simulate an SAE level 5 AV, the steering wheel and pedals were removed. The RealSense camera provided eye, head, and hand data to a machine-learning model trained to recognise gaze and gestures.

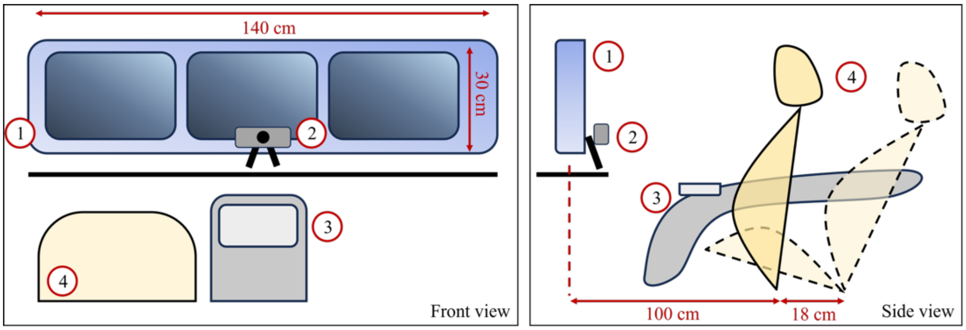


Figure 1: Front and side view of the user study setup.

Concept

The concept combined a gaze-and-gesture interaction for operating a custom automotive GUI with proactive pop-ups that taught this interaction to first-time users. The GUI, built in Protopie for a widescreen layout, contained three typical car display regions–instrument cluster (IC), head unit (HU), and passenger display (PD). See Figure 3 for the initial version (without the information pop-up). A music player and a video player were interactive and could be moved across these regions. Interaction followed an AR-glasses-inspired model: gaze selected the window, and gestures controlled its functions. These specific hand gestures were chosen because the system recognised them reliably. When users looked at a window, it was outlined for feedback; gesture feedback appeared as pictograms, time-stamp indicators (for forward/reverse), or a three-step animation for relocating windows (grab→drag→release).

The proactive teaching of how to operate one of the functions through multimodal interaction consisted of (1) animations that showed a person performing the gestures correctly, (2) textual explanations, and (3) pictograms. Figure 2 shows the static frames of the animations. An animation cycle lasted 3 s (for volume and mute) or 6 s (for play, pause, fast forward/reverse, and window relocation) and was looped until the entire pop-up disappeared from the screen. The three conditions varied in the following aspects: (1) how the teaching information was accessed, (2) the size of the pop-ups, (3) where the information pop-ups appeared on screen, and (4) how many animations could appear simultaneously on one slide of a pop-up. For all conditions, the animations were grouped according to the categories found in Figure 2, which means that each condition had one (set of) pop-up(s) for each category of functions.

The study compared three teaching conditions that varied in pop-up size, placement, and interactivity. *Condition 1* (C1; tiny) used the smallest pop-up (250×80 px), positioned on the side of the head unit (Figure 3). It indicated the availability of gaze and gesture controls, and users could open additional text and animation content via the touchpad (3 in Figure 1). If untouched, it disappeared after 30 s, and the instructional material remained inaccessible. This baseline resembled current Mercedes-Benz implementations (Mercedes-Benz AG, 2023a, 2023b). *Condition 2* (C2; medium) functioned identically but used a larger, centrally placed pop-up (1200×200 px) to increase visibility (Figure 3). Both C1 and C2 provided access to the same browseable instructional slides (Figure 4). *Condition 3* (C3; big) presented the largest instructional content (600×500 to

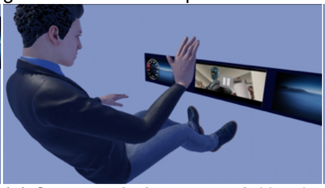
1500×500 px) and differed by removing user input entirely. Pop-ups appeared automatically on the instrument cluster. Animations started immediately after seat recline, and content disappeared after 30 s (Figure 5). Depending on the function category, one to three animations were played simultaneously. C3 was passive, prioritising visibility and immediacy over user-controlled exploration.

(a) **Audio control**: Thumbs down gesture for *volume down*.

(b) **Audio control**: Thumbs up gesture for volume up.

(c) **Audio control**: Index finger on lips gesture for mute.

(d) **Content timing control**: Pinch and drop gesture for *fast forward or reverse*.



(e) **Content timing control**: Hand up gesture for *pause*.

(f) **Content timing control**: Wave twice gesture for *start of play*.

(g) **Content timing control**: Wave twice gesture for *end of play*.

(h) **Screen layout**: Grab gesture for *start of relocating window*.

(i) **Screen layout**: Drop gesture for *end of relocating window*.

Figure 2: Frames of the animations for gesture categories: audio control (a–c), content timing control (d–g), and screen layout (h–i).

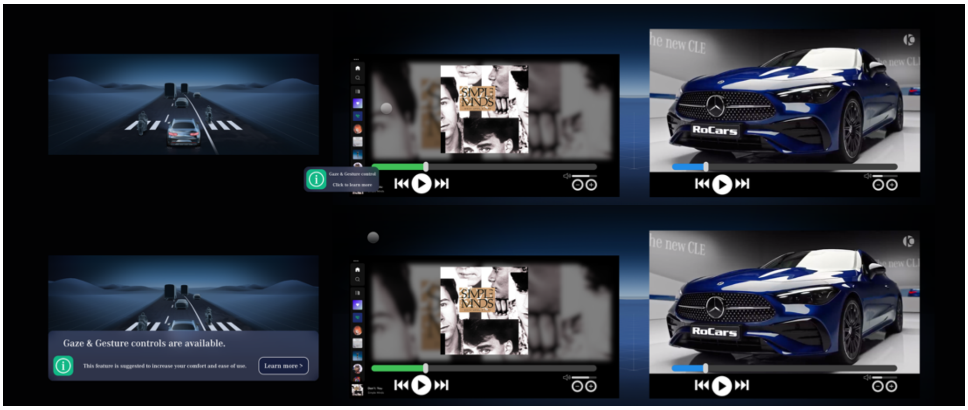


Figure 3: Information pop-ups. C1 (top) shows a small pop-up located on the side of the HU, while C2 (bottom) uses a larger, centrally placed pop-up for improved visibility.

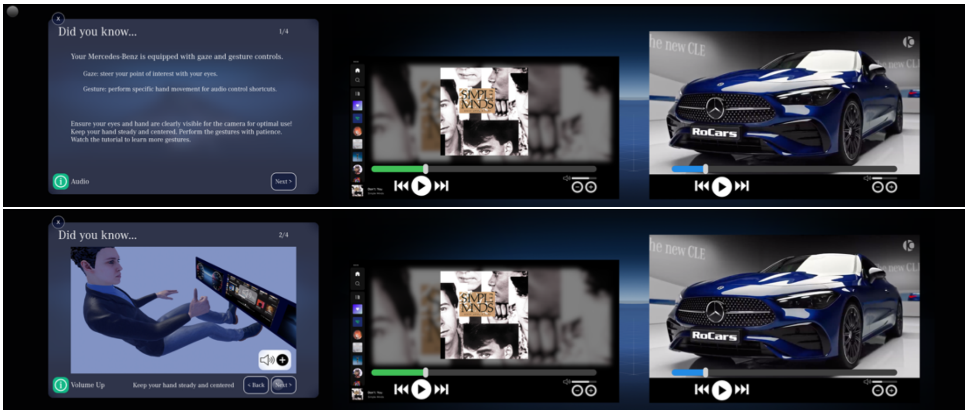


Figure 4: Slides from the information menu used in C1 and C2: textual explanation (top) and animated feedback (bottom).

Figure 5: Information pop-ups in C3 explaining gesture controls for different function categories: audio (top), content timing (middle), and screen layout (bottom).

Procedure and Data Analysis

The study comprised two parts, evaluating how participants experience and learn multimodal interaction in an AV. Age and gender were recorded at the start.

In Part 1, participants were gradually introduced to multimodal interaction without being told the study's purpose. Seated in a reclined position to simulate a naturalistic non-driving-related task (NDRT), they completed a structured secondary task: watching a widescreen video of the Mercedes-Benz CLE UI (https://youtu.be/ag9NQMrXM00; 1 in Figure 1), reading a one-page description of current Mercedes-Benz interfaces, and then re-watching the video (see supplementary material). This order was fixed. The examiner explained that the task served to assess comfort and realism of the reclined posture and to identify an incorrect statement in the text, framed as a challenge to maintain engagement. While participants were occupied, they experienced three remotely triggered seat recline movements. Each movement activated a 30 s pop-up tied to one function category (Figure 2), all presented under one randomly assigned teaching condition. Participants then used the functions in the shown category without guidance on which modality to use, after which the seat returned upright. This cycle occurred for all three categories. Ten participants experienced each condition, resulting in 30 exposures per function category. Behavioural observations captured whether participants noticed the pop-up, interacted with it, and perceived gaze feedback; failure to notice the content implied little chance of learning or gesture adoption.

In Part 2, participants were informed about the study's objective and the concept of multimodal interaction. They reviewed the pop-ups from Part 1, freely practised the techniques, and completed questionnaires evaluating the teaching method they had just experienced. They then viewed the two conditions they had not yet encountered and provided comparative feedback before answering open questions. After each condition, participants evaluated the teaching method on three scales: the 20-point NASA-TLX for workload (Hart and Staveland, 1988), the KANO scale for reactions to new features (Kano et al., 1984), and an Acceptance scale measuring usefulness and satisfaction on a −2 to +2 scale (Van Der Laan et al., 1997). A separate 12-item questionnaire assessed the visual properties and content of the pop-ups on a 7-point Likert scale. Items Q1–Q3 covered animations, clarity, and visibility; Q4–Q5 size and duration; Q6–Q7 informativeness and interference; Q8–Q11 timing, preference for animation vs. text, and openness to proactive learning; and Q12 general system guidance. Q1–Q7 were answered after each condition, while Q8–Q12 appeared only after the first. At the end, participants ranked the three teaching conditions and selected their most and least preferred gesture. The forms are available in the supplementary material.

NASA-TLX, Acceptance, and the main questionnaire data were analysed using ANOVA. Rankings were tested using a chi-square test. A significance level of α=0.05 was applied.

RESULTS

Notices and Interaction Rates

The data of all participants were retained. Table 1 displays the notices and interactions for each condition. C3 was noticed the most often of the three conditions (N=28), and C1 the least frequently (N=10). A Chi-Pearson square test was performed under the three conditions, which did not produce significant differences in the number of notices. For all conditions, the second appearance notices, during which participants performed the reading task, resulted in lower notices (N=1, 3, 8 for C1, C2 and C3, respectively) than for the first (N=5, 8, 10 for C1, C2 and C3, respectively) and third (N=4, 8, 10 for C1, C2, and C3, respectively). C1 did not produce interactions. For this condition, the user had to click on the information notification to reach the full explanation of gaze and gesture control. The complete lack of interactions for this condition implies that none of the participants in this group learned about the gaze and gesture control that was available. C2 produced only one interaction of the 30 possible interactions between the participants. C3 had 12 interactions out of 30 possible interactions, resulting in the highest interaction rate of the three conditions.

Table 1. Results (n → i) of participants' notices count (n) and interaction count (i) with each appearance of an information pop-up for all three conditions. p_n=0.72 (p_n is the p-value result of the Chi-Square test for n.

	C1	C2	C3
First appearance	5 → 0	8 → 0	10 → 4
Second appearance	1 → 0	3 → 1	8 → 4
Third appearance	4 → 0	8 → 0	10 → 4
Total	10 → 0	19 → 1	28 → 12

Table 2 shows the results of the ANOVA test for the questionnaire. For dimensions Q3, Q4 and Q5 (visibility, size, and duration), the results show significant differences between the three conditions and are marked in bold. The other dimensions did not show any significant differences. Table 3 displays the results for the NASA TLX, Acceptance scale, and Kano scale. The results show that there was a significant difference, marked in bold, only for the dimension of temporal demand. In this dimension, C2 scored the lowest demand. No significant differences were found in any of the other five dimensions of the NASA TLX. There were no significant differences in usability or satisfaction dimensions for the Acceptance scale.

Table 2. Results of the ANOVA test on all three conditions across the twelve dimensions of the questionnaire (on a 7-point Likert scale of 1–7).

Question	Mean value			*p*	Question	Mean value			*p*
	C1	C2	C3			C1	C2	C3	
Q1 animations	5.63	5.87	5.57	0.6	Q7 interference	3.77	4.43	4.67	0.11
Q2 understand.	5.37	6.07	5.97	0.13	Q8 moment	3.4	4.8	4.2	0.19
Q3 visibility	4.47	5.97	6.13	**<0.001**	Q9 windows	6.4	5.5	6.5	0.22
Q4 size	2.53	4.6	4.17	**<0.001**	Q10 description	2.7	3.2	1.9	0.19
Q5 duration	3.6	4.33	4.17	**0.001**	Q11 proactive	5.8	6.3	6.2	0.46
Q6 enough info	5.73	6.03	6.03	0.55	Q12 feedback	5.6	5.2	5.9	0.55

Table 3. Left: results of the ANOVA test on the conditions across the six dimensions of the NASA TLX scale (on a 20-point scale 1–20) (Hart and Staveland, 1988) and usefulness and satisfaction of the Acceptance scale (on a 5-point scale -2–+2)) (Van Der Laan et al., 1997). Right: results of the Kano scale (Kano et al., 1984).

	Mean value					Mean value		
NASA TLX	**C1**	**C2**	**C3**	***p***	**Kano scale**	**C1**	**C2**	**C3**
Mental demand	8.03	6.97	7.93	0.42	Performance	8	11	10
Physical demand	3.7	4.3	3.87	0.32	Must-have	13	13	11
Temporal demand	7.73	5.37	7.23	0.02	Attractive	1	0	2
Performance	6.27	5.5	6.33	0.35	Indifferent	7	4	5
Effort	7.1	6.7	7.63	0.52	Reverse	1	1	1
Frustration	4.8	5.0	5.53	0.63	Questionable	0	1	1
Acceptance scale	**C1**	**C2**	**C3**	***p***				
Usefulness	1.2	1.29	1.28	0.48				
Satisfying	1.16	1.26	1.26	0.62				

Ranking of Conditions and Gestures

Participants clearly preferred C3 and C2 over C1. C3 received the highest number of first-place rankings (13), followed closely by C2 (12), while C1 was most often ranked last (16). A statistical test confirmed that the distribution of rankings differed significantly from what would be expected by chance (p=0.048). For gesture preferences, the mute gesture emerged as the clear favourite (14 votes), followed by volume (7) and grab (6). Pause received only 2 votes, and both pinch and play received none. Pinch and play were instead rated as the least liked gestures (11 votes each), indicating a strong negative preference for these two interactions.

DISCUSSION

This study examined a proactive method for teaching first-time users gaze- and gesture-based interaction in an SAE level 5 AV. Thirty participants completed a secondary task while receiving one of three teaching variations, differing in size, placement, and interaction demands. The results show that participants were positive about receiving proactive instruction from their car, and most viewed such guidance as a must-have or performance feature. Among the three variations, C3 proved the most effective for teaching, producing the highest number of noticed messages and learned interactions, while C2 elicited only a single interaction and C1 none. These differences are largely driven by visibility and clarity: the larger size, central placement, and contrasting colours of C3 enhanced prominence in line with prior findings on visual salience and attention (Wickens et al., 2022). Its automatic presentation also reduced effort. Although participants rated C3 and C2 as favourites, acceptance scores did not differ significantly across conditions, suggesting that participants evaluated the *concept* of proactive teaching rather than specific implementations, consistent with broader patterns in automation trust and acceptance (Hoff and Bashir, 2015; Lee and See, 2004).

Participants' ranking of the conditions reinforces this interpretation: C3 and C2 were consistently preferred over C1 (p=0.048), primarily because they attracted more attention and provided clearer cues for interaction. Participants valued animations in C3 and the menu button in C2, while C1's small size limited noticeability. Despite large pop-ups covering much of the interface in C3 and C2, participants did not report stronger interference with their ongoing task, indicating that prominence did not inherently reduce comfort or usability. Perceived duration ratings (Q5) and NASA-TLX temporal demand differed significantly across conditions, despite identical 30-second durations, likely reflecting participants' perceived pressure to absorb the information quickly. These findings collectively point to a strong user preference for clear, visible, and minimally effortful proactive teaching when learning new multimodal interactions in AVs.

Gesture preferences followed similar patterns. The mute gesture emerged as the most popular due to its simplicity, reliability, and intuitive mapping between action and function. In contrast, the play and pinch gestures were the least preferred, largely due to recognition difficulties and the more complex animations used to illustrate them. Participants found the combined play-pause animation particularly confusing. The results highlight that simple gestures supported by concise, function-specific animations are most effective for first-time learning of multimodal interaction in automotive contexts.

This study has limitations. Participants interacted with the system briefly, whereas real-world learning unfolds over longer periods; future work should examine longitudinal learning and adaptive teaching strategies. The prototype supported only gaze and gesture input; integrating voice could enhance usability, particularly when touch is impractical. Gesture recognition errors, limited gaze-tracking resolution, and the absence of automated logging reduced analytical depth. More advanced tracking technologies, such as those found in the Apple Vision Pro, could improve accuracy. Further research should evaluate proactive teaching strategies over extended periods in realistic settings.

SUPPLEMENTARY MATERIAL

Supplementary material containing materials used in the user study is available at: https://doi.org/10.4121/9374acdd-e787-4544-a7e4-91201df979f5.

REFERENCES

Aftab, A.R., 2019. Multimodal driver interaction with gesture, gaze and speech, in: ICMI '19: 2019 International Conference on Multimodal Interaction. ACM, Suzhou, China, pp. 487–492. https://doi.org/10.1145/3340555.3356093

Aftab, A.R., Von der Beeck, M., 2022. Multimodal driver referencing: a comparison of pointing to object inside and outside the vehicle, in: IUI '22: 27th International Conference on Intelligent User Interfaces. ACM, Helsinki, Finland, pp. 483–495. https://doi.org/10.1145/3490099.3511142

Alvarez, I., Martin, A., Dunbar, J., Taiber, J., Wilson, D.-M., Gilbert, J.E., 2010. Voice interfaced user help, in: Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI). Pittsburgh, PA, USA, pp. 42–49.

Amazon, 2024. Vehicles with alexa [WWW Document]. URL https://www.amazon.com/alexa-auto (accessed 4.27.25).

Audi, 2022. Space travel through megacity [WWW Document]. URL https://press.audi.co.uk/releases/804 (accessed 6.19.25).

Bellini, E., Dell'Era, C., Frattini, F., Verganti, R., 2017. Design-driven innovation in retailing: An empirical examination of new services in car dealership. Creativity and Innovation Management 26, 91–107. https://doi.org/10.1111/caim.12184

BimmerTech, 2020. BMW gesture control–the next level of iDrive interaction [WWW Document]. URL https://www.bimmer-tech.net/blog/item/124-bmw-gesture-control (accessed 8.28.25).

BMW Group, 2019. Get the most out of gesture control [WWW Document]. URL https://www.youtube.com/watch?v=_mGwJh4da5w (accessed 4.12.25).

BMW Group, 2015. The new BMW 7 series [WWW Document]. URL https://www.press.bmwgroup.com/global/article/detail/T0221224EN/the-new-bmw-7-series (accessed 7.7.25).

Bourguet, M.-L., 2003. Designing and prototyping multimodal commands, in: Human-Computer Interaction–INTERACT'03. Zurich, Switzerland, pp. 717–720.

Caraban, A., Karapanos, E., Gonçalves, D., Campos, P., 2019. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction, in: CHI'19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Glasgow, UK, pp. 1–15. https://doi.org/10.1145/3290605.3300733

Caricos, 2022. 2022 Mercedes-Benz EQS–interior, cockpit [WWW Document]. URL https://www.caricos.com/cars/m/mercedes-benz/2022_mercedes-benz_eqs/images/58.html (accessed 9.1.25).

Cerence, 2021. How Mercedes is creating innovative multi-modal experiences with cerence look [WWW Document]. URL https://www.cerence.com/newsroom/blog/how-mercedes-is-creating-innovative-multi-modal-experiences-with-cerence-look (accessed 5.19.25).

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research, in: Human Mental Workload, Advances in Psychology. Elsevier, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hoff, K.A., Bashir, M., 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. Hum Factors 57, 407–434. https://doi.org/10.1177/0018720814547570

Kano, N., Seraku, N., Takahashi, F., Tsuji, S., 1984. Attractive quality and must-be quality. Journal of the Japanese Society for Quality Control 14, 39–48. https://doi.org/10.20684/quality.14.2_147

Kern, D., Mahr, A., Castronovo, S., Schmidt, A., Mueller, C., 2010. Making use of drivers' glances onto the screen for explicit gaze-based interaction, in: Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI). Pittsburgh, PA, USA, pp. 110–113. https://doi.org/10.1145/1969773.1969792

Lee, J.D., See, K.A., 2004. Trust in automation: Designing for appropriate reliance. Human factors 46, 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, M.K., Kiesler, S., Forlizzi, J., 2011. Mining behavioral economics to design persuasive technology for healthy choices, in: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, Canada, pp. 325–334. https://doi.org/10.1145/1978942.1978989

Lentz, A., Schlesinger, B., DiMartile III, J.T., Taubman, G., O'Dell, R., 2018. A logical layer to interpret user interactions.

Macek, T., Labsky, M., Vystrcil, J., Luksch, D., Kasparova, T., Kunc, L., Kleindienst, J., 2014. Interactive car owner's manual user study, in: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI). Seattle, WA, USA, pp. 1–4. https://doi.org/10.1145/2667317.2667412

Mercedes-Benz AG, 2024. Discover your owner's manual [WWW Document]. URL https://www.mercedes-benz.co.uk/passengercars/services/manuals.html (accessed 5.24.25).

Mercedes-Benz AG, 2023a. C-class saloon manual [WWW Document]. URL https://www.mercedes-benz.co.uk/passengercars/services/manuals.html/c-class-saloon-2023-09-w206-mbux/digital-owners-manual/calling-up-the-digital-owners-manual (accessed 3.5.25).

Mercedes-Benz AG, 2023b. S-class saloon manual [WWW Document]. URL https://www.mercedes-benz.co.uk/passengercars/services/manuals.html/s-class-saloon-2023-09-w223-mbux/digital-owners-manual/calling-up-the-digital-owners-manual (accessed 8.29.25).

Mercedes-Benz AG, 2023c. Mercedes-Benz takes in-car voice control to a new level with ChatGPT [WWW Document]. URL https://group.mercedes-benz.com/innovation/digitalisation/connectivity/car-voice-control-with-chatgpt.html (accessed 7.3.25).

Mercedes-Benz AG, 2018. It understands you perfectly. [WWW Document]. URL https://group.mercedes-benz.com/company/magazine/technology-innovation/mbux-voice-assistant-hey-mercedes.html (accessed 8.1.25).

Mercedes-Benz AG, 2015. The Mercedes-Benz F 015 Luxury in Motion [WWW Document]. URL https://media.mbusa.com/releases/the-mercedes-benz-f-015-luxury-in-motion (accessed 3.1.25).

MotorTrend, 2023. The 2024 BMW 5 series lets you steer with just your eyes [WWW Document]. URL https://www.motortrend.com/news/2024-bmw-5-series-eye-lane-change-tech (accessed 6.4.25).

Naujoks, F., Befelein, D., Wiedemann, K., Neukum, A., 2017. A review of non-driving-related tasks used in studies on automated driving, in: Advances in Human Aspects of Transportation. Springer, Los Angeles, CA, USA, pp. 525–537. https://doi.org/10.1007/978-3-319-60441-1_52

Novick, D.G., Ward, K., 2006. Why don't people read the manual, in: SIGDOC '06: Proceedings of the 24th Annual ACM International Conference on Design of Communication. ACM, Myrtle Beach, SC, USA, pp. 11–18. https://doi.org/10.1145/1166324.1166329

Pfleging, B., Kun, A.L., Shaer, O., 2022. Automated vehicles as a space for work & wellbeing, in: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems. New Orleans, LA, USA, pp. 1–7. https://doi.org/10.1145/3491101.3503766

Pfleging, B., Schneegass, S., Schmidt, A., 2012. Multimodal interaction in the car–combining speech and gestures on the steering wheel, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI). Portsmouth, NH, USA, pp. 155–162. https://doi.org/10.1145/2390256.2390282

Porsche AG, 2024. Taycan [WWW Document]. URL https://www.porsche.com/international/models/taycan/taycan-models/taycan

SAE, 2021. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles [WWW Document]. URL https://www.sae.org/standards/content/j3016_202104 (accessed 6.9.24).

screens, 2023. Li auto L9 | gesture control [WWW Document]. URL https://www.youtube.com/watch?v=q09VOemNji8 (accessed 11.22.25).

Skoda Auto AS, 2023. Digital manual [WWW Document]. URL https://www.skoda-auto.com/connectivity/infotainment-apps-digital-manual (accessed 7.5.25).

Stryja, C., Satzger, G., 2018. Digital nudging to overcome cognitive resistance in innovation adoption decisions. The Service Industries Journal 39, 1123–1139. https://doi.org/10.1080/02642069.2018.1534960

Thaler, R.H., Sunstein, C., 2009. Nudge: Improving decisions about health, wealth and happiness. Penguin.

Top Gear Magazine, 2020. Honda e review [WWW Document]. URL https://www.topgear.com/car-reviews/honda/e (accessed 11.25.25).

Van Der Laan, J.D., Heino, A., De Waard, D., 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies 5, 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Volvo Cars, 2015. Volvo Cars unveils Concept 26, delivering the luxury of time [WWW Document]. URL https://www.volvocars.com/intl/media/press-releases/4F93E8898A363A84 (accessed 7.24.25).

Wee, D., 2021. From print to voice and beyond: How Toyota is transforming the car manual [WWW Document]. URL https://cloud.google.com/blog/topics/manufacturing/toyota-modernizes-the-car-manual-with-google-cloud (accessed 11.17.25).

Wessel, G., Alterndorf, E., Schwalm, M., Canpolat, Y., Burghardt, C., Flemisch, F., 2019. Self-determined nudging: A system concept for human-machine interaction. Cognition, Technology and Work 21, 621–630. https://doi.org/10.1007/s10111-019-00556-5

Wickens, C.D., Gutzwiller, R.S., McCarley, J.S., 2022. Applied Attention Theory, 2nd ed. CRC Press, Boca Raton. https://doi.org/10.1201/9781003081579