Exploring Emotion–Motion Coupling in Level-2 Automated Urban Turning using In-Cabin Facial Expressions and Vehicle Dynamics

Siyi Dai* und Jakob Huemer† und Klaus Bengler‡

Zusammenfassung: Automated-driving user experience depends not only on whether a maneuver succeeds, but also on how its motion unfolds. This exploratory follow-up study tests whether in-cabin facial emotion-channel probabilities co-vary with vehicle dynamics during Level-2 automated urban turns, which motion cues dominate, and which associations remain after participant-baseline control. Across 44 participants, 497 clips, and 2626 two-second windows, weak but reproducible coupling emerged. Lateral acceleration and speed were most consistent, with yaw/curvature and longitudinal jerk secondary. Facial channels are therefore treated as contextual behavioral markers, not standalone emotion detectors.

Schlüsselwörter: affective computing, automated driving, facial expression, vehicle dynamics

1 Introduction

Advanced driver assistance systems increasingly combine longitudinal and lateral control. They therefore shape not only the objective trajectory of a vehicle, but also the moment-to-moment experience of the person supervising the automation. In SAE Level 2 automation, the driver remains responsible for supervision and intervention, while the system controls parts of the driving task and can expose the user to motions that feel smooth, assertive, cautious, or surprising [1, 2]. Comfort, enjoyment, perceived safety, and acceptance depend on driving style and familiarity, and acceleration and jerk are central physical determinants of automated-vehicle comfort [3, 4]. This is particularly relevant in urban left turns, where curvature change, speed reduction, yaw motion, lateral acceleration, and jerk occur within a few seconds. A maneuver can be technically successful and still be experienced as uncomfortable, unsafe, or poorly matched to user expectations.

The simulator study from which the present analysis is derived examined Level-2 automated urban turning styles and their relation to subjective user experience. The earlier analysis focused on scenario-level evaluations: after each scenario, participants rated perceived safety, comfort and relaxation, and stimulation and fun. Those results showed that

*BMW Group, Knorrstraße 147, 80807 Munich, Germany. (e-mail: siyi.dai@bmw.de)

†BMW Group, Knorrstraße 147, 80807 Munich, Germany. (e-mail: jakob.huemer@bmw.de)

‡Technical University of Munich (TUM), Boltzmannstraße 15, 85748 Garching near Munich, Germany. (e-mail: bengler@tum.de).

automated turning style and user expertise influence subjective ratings, with acceleration/deceleration style being particularly diagnostic for perceived safety and comfort [5]. Post-scenario ratings remain essential because they capture what users can report directly. However, they compress a multi-second maneuver into one retrospective judgment and therefore cannot show whether expressive signals fluctuate with the motion profile inside the maneuver.

In-cabin cameras offer a time-resolved complement. In many fields, camera data is used to analyze facial behavior and expressions and draw conclusions about possible human emotions. For example, approaches for media and entertainment analysis [6], innovative training methods in sports [7] or the optimization of human-robot interaction [8], camera data is used with computer vision to analyse facial behavior. In the automotive sector, facial behavior has been studied as an indicator of discomfort or affective response in automated driving [9], and recent face-analysis toolkits can produce continuous action-unit, gaze, and emotion-channel outputs from ordinary video streams [10, 11]. These outputs can be synchronized with vehicle dynamics, but they require caution. Facial emotion recognition is not a direct measure of a user's ground-truth affective state; broader affective-science work warns against inferring discrete emotions directly from facial movements, and recent automated-driving work has shown that facial emotion recognition can fail to capture feeling unsafe [12, 13]. The present paper therefore treats the emotion outputs as classifier channels that may contain useful behavioral information, not as validated labels of what the user truly felt.

The research gap is the temporal link between vehicle motion and in-cabin facial behavior during short automated urban maneuvers. Prior work often relates motion descriptors to post-scenario ratings or studies facial behavior around predefined events. Less is known about whether continuous facial emotion-channel probabilities co-vary with continuous vehicle dynamics inside a maneuver, which motion variables show the strongest associations, and whether such associations remain after controlling for stable individual differences in facial expressiveness. This paper addresses that gap through an exploratory secondary analysis of aligned simulator and in-cabin video data. It does not attempt to build an affect detector. Instead, it asks whether the facial channels contain repeatable information about vehicle motion during short automated urban maneuvers:

- **RQ1:** Is there measurable time-resolved coupling between facial emotion-channel probabilities and vehicle dynamics during Level-2 automated urban scenarios?
- **RQ2:** Which motion cues carry the strongest and most interpretable coupling?
- **RQ3:** Which associations remain credible after participant-baseline correction, clip-level aggregation, and window-size sensitivity checks?

The contribution is a reproducible coupling map that identifies where facial behavior and motion co-vary, separates robust associations from analysis-dependent ones, and defines concrete hypotheses for future multimodal user-experience modelling.

2 Study Logic and Method

2.1 Analysis Logic

In automated driving, motion is not only a consequence of path planning and control; it is also an input to subjective evaluation. A user feels acceleration, deceleration, yaw rotation, and jerk directly through the body, and these signals can be interpreted as smoothness, hesitation, assertiveness, or lack of margin. The earlier analysis of this simulator study followed that logic at the scenario level by relating whole-turn ratings to manipulated driving styles. The present paper moves one level lower. Instead of asking whether a whole turn is rated positively, it asks whether local changes in the motion time series are mirrored in local changes in facial behavior.

The interpretation of facial emotion probabilities is deliberately conservative. A classifier output named "fear" or "disgust" does not mean that the participant consciously felt fear or disgust. The output can be affected by head pose, gaze, lighting, facial morphology, expression intensity, and the classifier's training data. It can also reflect concentration, reduced expressiveness, bracing, or small action-unit combinations that are not emotional in the everyday sense. The term emotion–motion coupling is therefore used statistically: it denotes covariation between probability channels and vehicle-dynamic variables.

The analysis was exploratory but task-guided. Urban left turns contain approach and speed reduction, entry into curvature, yaw rotation through the intersection, and exit. We therefore expected speed, lateral acceleration, yaw rate, curvature, and jerk to be more informative than variables with little variation during the short clips. We also expected participant-specific facial baselines to matter, because some participants are generally more expressive than others. The most credible results were defined as those visible in both participant-demeaned pooling and clip-level aggregation.

2.2 Simulator Study and Analytic Sample

The data come from a motion-base driving simulator study of Level-2 automated urban driving. Participants first completed a short familiarization drive and then experienced 13 automated scenario clips along a predefined urban route. The route included nine experimentally manipulated unprotected left-turn scenarios and four contextual scenarios. The manipulated left turns crossed curvature style with acceleration/deceleration style, each at assertive, moderate, and defensive levels. Curvature profiles were derived from representative naturalistic intersection trajectories, and force-level behavior was varied through the acceleration/deceleration profile and corresponding lateral-acceleration targets [14, 5]. The present analysis uses the available aligned facial and simulator streams rather than the post-scenario questionnaire data.

The present analysis includes 44 participants for whom usable aligned motion and in-cabin facial streams were available. Across these participants, 497 scenario clips were analyzed. The median clip duration was 9.2 s, with an interquartile range of 1.2 s. At the primary two-second window length, the dataset contained 2626 windows. For clip-level correlations, 397 clips had at least five windows and were therefore eligible for meta-analysis, contributing 2226 windows in total. The analysis was organized at three levels: 10 Hz samples for signal preparation, averaged two-second windows for pooled and participant-demeaned correlations, and scenario clips for Fisher-z meta-analysis. This se-

Tabelle 1: Analytic dataset for the emotion–motion coupling analysis.

Quantity	Value
Participants with usable aligned data	44
Scenario clips	497
Scenario indices	13
Total 10 Hz samples	48,822
Median clip duration	9.2 s
Two-second windows	2626
Meta-analysis eligible clips (≥ 5 windows)	397
Windows in meta-analysis eligible clips	2226

paration prevents sample count, window count, and clip count from being treated as the same evidence unit.

2.3 Signals, Windowing, and Statistics

In-cabin video was processed with OpenFace 3.0, a lightweight multi-task facial-behavior toolkit that estimates landmarks, action units, gaze, and emotion probabilities [10]. We used the eight basic emotion-channel probabilities provided in the aligned files: neutral, happy, sad, surprise, fear, disgust, anger, and contempt. Simulator signals were available on the aligned 10 Hz grid and included speed, longitudinal acceleration, lateral acceleration, yaw rate, steering-related variables, and lane offsets. The analysis script preserved canonical motion predictors and derived yaw acceleration, curvature, and dynamic jerk terms. Curvature was computed as yaw rate divided by speed when speed was nonzero. Longitudinal and lateral jerk were computed in the vehicle frame as

$$j_x = \dot{a}_x - a_y r + v r^2, \tag{1}$$

$$j_y = a_x r + v \dot{r}, \tag{2}$$

where v is speed, a_x and a_y are longitudinal and lateral acceleration, and r is yaw rate. A Savitzky–Golay filter was applied only in the derivative chain to reduce numerical noise before jerk and yaw-acceleration estimation. The main figures focus on eight motion variables: speed, longitudinal acceleration, lateral acceleration, yaw rate, yaw acceleration, curvature, longitudinal jerk, and lateral jerk. Lane offsets were retained in the full output tables but omitted from the main figures because left and right offsets were largely redundant for the present coupling question.

The facial and motion streams were averaged into non-overlapping windows. The primary window length was 2 s, chosen as a compromise between reducing sample-level autocorrelation and retaining within-maneuver temporal structure. Spearman correlations were used because both facial probabilities and vehicle-dynamic variables can be skewed, bounded, or affected by local outliers. Three correlation views were computed: pooled window-level correlations, participant-demeaned window-level correlations, and clip-level correlations aggregated by fixed-effect Fisher-z meta-analysis with weights $n-3$ [15]. False-discovery-rate control used the Benjamini–Hochberg procedure across the 80

Tabelle 2: Strongest clip-level meta-analytic associations and corresponding within-participant correlations.

Pair	r_{meta}	95% CI	q_{meta}	r_{within}
Sad–speed	0.263	[0.205, 0.319]	$< .001$	0.150
Neutral–lateral accel.	0.231	[0.173, 0.288]	$< .001$	0.256
Contempt–lateral accel.	0.213	[0.154, 0.271]	$< .001$	0.173
Happy–speed	-0.213	[-0.270, -0.154]	$< .001$	-0.128
Disgust–yaw rate	0.207	[0.148, 0.265]	$< .001$	0.166
Sad–lateral accel.	-0.206	[-0.264, -0.147]	$< .001$	-0.047
Contempt–speed	-0.196	[-0.254, -0.137]	$< .001$	-0.168
Disgust–curvature	0.191	[0.132, 0.249]	$< .001$	0.162

emotion–motion tests in each analysis view [16]. A sensitivity check repeated the within-participant analysis with 1 s, 2 s, and 4 s windows for the strongest meta-analytic pairs.

3 Results

3.1 Participant-Demeaned Window-Level Associations

Figure 1 shows the participant-demeaned Spearman correlations for the eight paper-level motion predictors. The effect sizes were small in absolute terms, but several associations were consistent after false-discovery-rate correction. The largest within-participant association was between neutral probability and lateral acceleration ($r = 0.256$). Neutral probability was also lower at higher speed ($r = -0.211$), higher with longitudinal jerk ($r = 0.182$), lower with curvature ($r = -0.132$), and lower with yaw rate ($r = -0.119$). Fear also decreased with speed ($r = -0.187$) and increased with lateral acceleration ($r = 0.146$).

Disgust showed a different pattern: it decreased with lateral acceleration ($r = -0.173$) and longitudinal acceleration ($r = -0.129$), but increased with yaw rate ($r = 0.166$) and curvature ($r = 0.162$). Contempt increased with lateral acceleration ($r = 0.173$) and longitudinal jerk ($r = 0.144$), but decreased with speed ($r = -0.168$). Happy probability decreased with speed ($r = -0.128$) and increased modestly with lateral acceleration ($r = 0.110$). Sad probability increased with speed ($r = 0.150$), while its within-participant association with lateral acceleration was weak ($r = -0.047$). The main result is therefore not a single emotion label, but a structured coupling pattern concentrated in speed, lateral acceleration, yaw/curvature cues, and longitudinal jerk.

3.2 Clip-Level Meta-Analysis

The clip-level analysis asked whether the same patterns appeared across individual scenario clips rather than only in pooled windows. Of the 80 tested emotion–motion pairs, 34 were significant after FDR correction. Table 2 lists the strongest meta-analytic associations and their corresponding within-participant correlations.

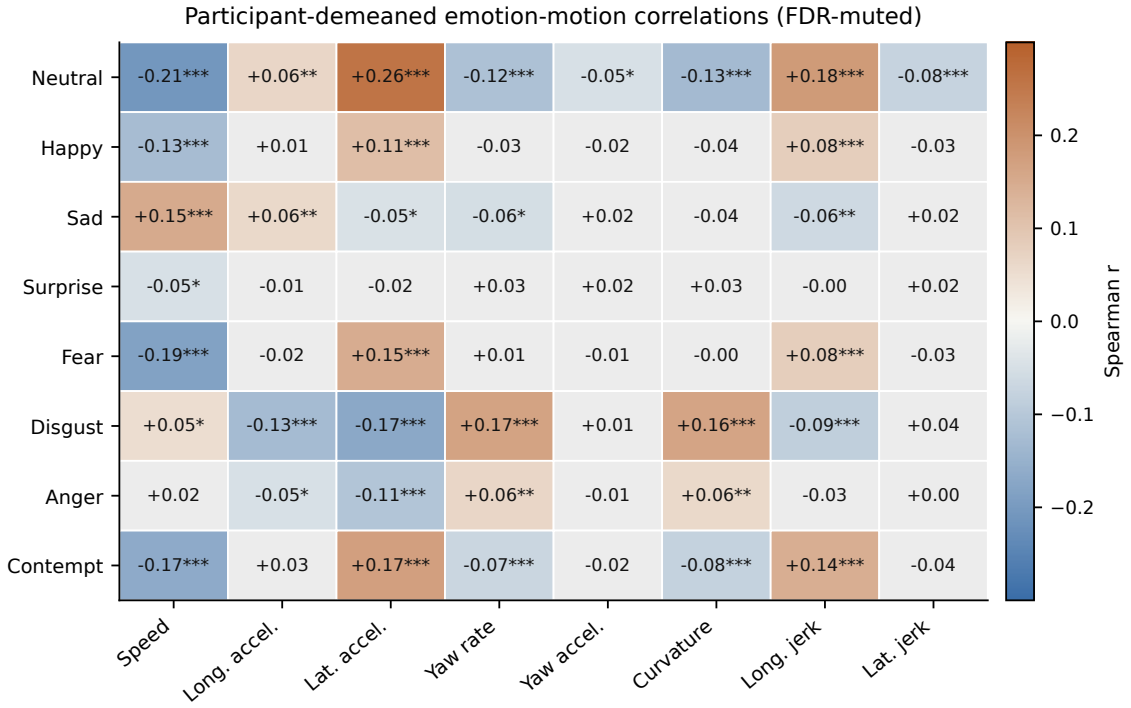

Abbildung 1: Participant-demeaned Spearman correlations between facial emotion-channel probabilities and vehicle-dynamic variables. Values are correlations on two-second window means after removing each participant's mean for every variable. Blue cells indicate negative coupling, orange cells indicate positive coupling, and gray cells did not survive FDR correction. Stars mark FDR-adjusted significance levels: $^{*}q < 0.05$, $^{**}q < 0.01$, $^{***}q < 0.001$.

The strongest meta-analytic effect was sad probability versus speed ($r_{\text{meta}} = 0.263$), followed by neutral probability versus lateral acceleration ($r_{\text{meta}} = 0.231$). The latter also had the largest participant-demeaned correlation, making it the most consistent single association across the two main views. Several other pairs were directionally similar across meta-analysis and within-participant pooling, including contempt–lateral acceleration, happy–speed, disgust–yaw rate, contempt–speed, and disgust–curvature. The pair sad–lateral acceleration was strong in the clip-level meta-analysis but weak after participant demeaning; this suggests that part of that association may depend on clip composition, participant baselines, or maneuver-phase structure rather than robust within-person coupling.

Figure 2 shows the full clip-level meta-analytic matrix. Compared with the participant-demeaned heatmap, the meta-analysis gives more weight to repeatability across scenario clips. Speed is associated with several emotion channels, lateral acceleration is prominent for neutral and contempt channels, and yaw/curvature variables are especially visible for disgust. Longitudinal jerk also appears in several positive associations. Yaw acceleration and lateral jerk show weaker and less systematic coupling in this dataset.

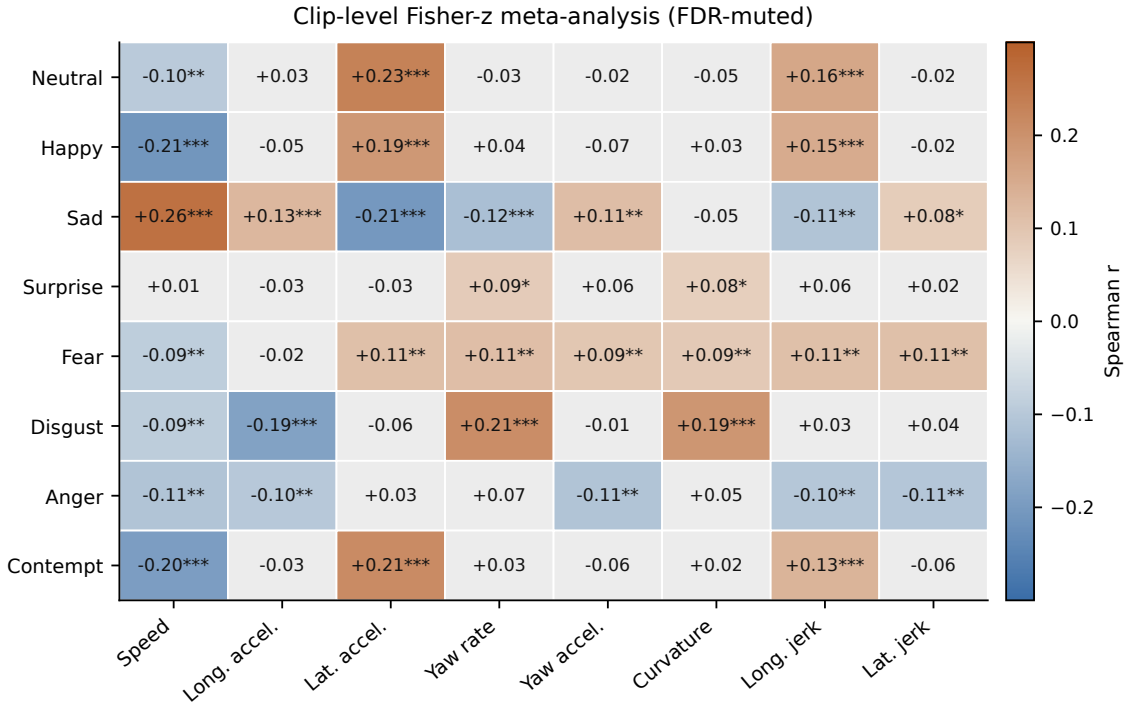

Abbildung 2: Clip-level Fisher-z meta-analysis of per-clip Spearman correlations. Values are meta-analytic correlations across clips with at least five windows. Blue cells indicate negative coupling, orange cells indicate positive coupling, and gray cells did not survive FDR correction. Stars mark FDR-adjusted significance levels.

3.3 Cross-View Agreement and Window-Size Sensitivity

Cross-view comparison showed a cluster of pairs with similar direction in both participant-demeaned and clip-level analyses, especially neutral–lateral acceleration, contempt–lateral acceleration, disgust–yaw rate, and sad–speed. Other pairs depended more strongly on the aggregation method. This distinction is important for interpretation: pairs visible in both views are better candidates for follow-up hypotheses than pairs that appear only in one analysis view.

The strongest pairs were generally stable when the window size was changed from 1 s to 4 s. Neutral–lateral acceleration remained the strongest participant-demeaned association, increasing slightly from $r = 0.227$ at 1 s to $r = 0.262$ at 4 s. Contempt–lateral acceleration and disgust–yaw rate also strengthened slightly with wider windows. Sad–speed was nearly unchanged across the three window sizes ($r \approx 0.15$), and happy–speed remained negative with similar magnitude. The weak sad–lateral-acceleration association stayed near zero and lost significance at 4 s. The core associations were therefore not artifacts of the selected 2 s window, but the sensitivity check reinforces that the effects are small.

4 Discussion

This exploratory follow-up analysis answers the research questions in a constrained way. First, measurable time-resolved emotion–motion coupling was present, but the effects we-

re small. Second, the clearest motion cues were lateral acceleration and speed, with yaw rate, curvature, and longitudinal jerk providing secondary coupling. Third, the most credible effects were those visible in both participant-demeaned and clip-level views. These findings should not be read as evidence that facial emotion recognition detects user experience. A more defensible interpretation is that facial behavior channels carry time-resolved information that is partly structured by vehicle motion.

The most robust single finding was the positive relation between neutral probability and lateral acceleration. In everyday language this may appear counterintuitive: higher lateral acceleration does not simply make users more "neutral." A more cautious interpretation is that the neutral channel may capture a face configuration associated with attention, gaze stabilization, reduced overt expressiveness, or maneuver phase. The speed-related effects point in the same direction. Sad probability increased with speed, whereas happy, fear, contempt, and neutral probabilities tended to decrease with speed. These patterns may partly reflect where in the maneuver the vehicle is: approach, rotation through the turn, and exit differ in both motion state and facial behavior.

Disgust and contempt channels were most clearly connected to yaw, curvature, lateral acceleration, and jerk. One possible interpretation is that these channels respond to face configurations produced by concentration, bracing, or small facial muscle changes during sharper turning phases. Another is that the classifier maps some non-emotional facial movements into those categories. The present data cannot distinguish these explanations. What it can show is that the channels are not random with respect to the vehicle trajectory. Some facial outputs systematically track dynamic features of the maneuver.

The analysis complements the earlier scenario-level user-experience study. Post-scenario ratings remain the more direct measure of perceived safety, comfort, and stimulation. The present results add temporal resolution by identifying parts of the vehicle-dynamics signal that are most likely to co-vary with facial behavior. For future modelling, motion context should be included when facial behavior is interpreted during automated driving; participant normalization should be used to account for stable facial baselines; and discrete emotion labels should not be treated as the only affective target. Action units, gaze, head pose, valence–arousal representations, physiological signals, and continuous subjective comfort ratings may provide a richer description of user state.

4.1 Limitations

Several limitations are central. First, this is a secondary, exploratory analysis of a simulator dataset. The driving environment was controlled, and the clips were short. Generalization to real urban traffic, longer automation exposures, and more varied passengers requires replication. Second, the facial variables are OpenFace emotion probabilities. They are useful machine-readable channels, but they are not validated labels of participant affect in this study. The probability channels are also compositional: changes in one channel necessarily interact with changes in others. Third, the analysis is correlational. Vehicle motion may influence facial behavior, but facial behavior can also co-vary with time within the scenario, gaze direction, head movement, anticipation, or unmeasured situational factors.

Fourth, windowed correlations reduce but do not remove temporal autocorrelation. The clip-level meta-analysis partly addresses this by aggregating per-clip associations, yet it remains a fixed-effect summary of an exploratory signal search. Fifth, the analytic

subset contains only participants and clips with usable aligned video and motion streams. This is appropriate for the present question but should be distinguished from the full questionnaire sample of the original simulator study. Finally, no continuous self-report, skin conductance, heart-rate, or eye-tracking target was used here to validate the facial-channel changes. Future work should combine these modalities and use time-lagged or event-triggered models to test causal timing more directly.

5 Conclusion

The study provides a reproducible exploratory analysis of emotion–motion coupling in Level-2 automated urban scenarios. Across 44 participants and 2626 two-second windows, facial emotion-channel probabilities showed small but consistent associations with vehicle dynamics. The most stable patterns involved lateral acceleration, speed, yaw rate, curvature, and longitudinal jerk. The findings are hypothesis-generating rather than diagnostic. They suggest that continuous facial analysis can complement post-scenario user-experience ratings, provided that emotion probabilities are interpreted as behavioral model outputs and validated against richer affective and subjective measures in future work.

Literatur

[1] SAE International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE Standard J3016_202104, 2021. doi: https://doi.org/10.4271/J3016_202104.

[2] K. Bengler, K. Dietmayer, B. Färber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014. doi: https://doi.org/10.1109/MITS.2014.2336271.

[3] F. Hartwich, M. Beggiato, and J. F. Krems, "Driving comfort, enjoyment and acceptance of automated driving: Effects of drivers' age and driving style familiarity," *Ergonomics*, vol. 61, no. 8, pp. 1017–1032, 2018. doi: https://doi.org/10.1080/00140139.2018.1441448.

[4] K. N. de Winkel, T. Irmak, R. Happee, and B. Shyrokau, "Standards for passenger comfort in automated vehicles: Acceleration and jerk," *Applied Ergonomics*, vol. 106, p. 103881, 2023. doi: https://doi.org/10.1016/j.apergo.2022.103881.

[5] S. Dai, J. Huemer, and K. Bengler, "Linking vehicle dynamics to user experience in L2 automated urban turning: A driving simulator study," in *Proc. IEEE Int. Conf. Human-Machine Systems (ICHMS)*, 2026, to appear.

[6] M. Sajjad, S. Zahir, A. Ullah *et al.*, "Human Behavior Understanding in Big Multimedia Data Using CNN based Facial Expression Recognition," *Mobile Networks and Applications*, vol. 25, pp. 1611–1621, 2020. doi: https://doi.org/10.1007/s11036-019-01366-9.

[7] O. J. Santana, D. Freire-Obregón, D. Hernández-Sosa *et al.*, "Facial expression analysis in a wild sporting environment," *Multimedia Tools and Applications*, vol. 82, pp. 11395–11415, 2023. doi: https://doi.org/10.1007/s11042-022-13654-w.

[8] Z. Liu *et al.*, "A facial expression emotion recognition based human-robot interaction system," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 668–676, 2017. doi: https://doi.org/10.1109/JAS.2017.7510622.

[9] M. Beggiato, N. Rauh, and J. F. Krems, "Facial expressions as indicator for discomfort in automated driving," in *Intelligent Human Systems Integration 2020*, Springer, 2020, pp. 932–937. doi: https://doi.org/10.1007/978-3-030-39512-4_142.

[10] J. Hu, L. Mathur, P. P. Liang, and L.-P. Morency, "OpenFace 3.0: A lightweight multitask system for comprehensive facial behavior analysis," arXiv preprint arXiv:2506.02891, 2025. doi: https://doi.org/10.48550/arXiv.2506.02891.

[11] J. A. P. Lim, G. P. Kusuma, and A. Zahra, "Facial Emotion Recognition Using Computer Vision," in *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, Jakarta, Indonesia, pp. 46–50, 2018. doi: https://doi.org/10.1109/INAPR.2018.8626999.

[12] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019. doi: https://doi.org/10.1177/1529100619832930.

[13] A. van Elburg, K. Gkentsidis, M. Sarrazin, S. Barendswaard, V. Kotian, and R. Happee, "Facial emotion recognition does not detect feeling unsafe in automated driving," arXiv preprint arXiv:2509.04490, 2025. doi: https://doi.org/10.48550/arXiv.2509.04490.

[14] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The inD dataset: A drone dataset of naturalistic road user trajectories at German intersections," in *Proc. IEEE Intelligent Vehicles Symposium*, 2020, pp. 1929–1934. doi: https://doi.org/10.1109/IV47402.2020.9304839.

[15] R. A. Fisher, "On the probable error of a coefficient of correlation deduced from a small sample," *Metron*, vol. 1, no. 4, pp. 1–32, 1921.

[16] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, pp. 289–300, 1995. doi: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[17] R. Coutinho and M. Antonio Placucci, "Recognition of Facial Emotions through the Use of Computer Vision and Intelligence: A Context of an Intelligent Classroom," in *ICERI2024 Proceedings*, pp. 1351–1360, 2024. doi: https://doi.org/10.21125/iceri.2024.0408.