Research on the Emotional Recognition and Interactive Influence Mechanism of the Main and Co-Pilots

Qiuyue Wang[a]
Sydney Smart Technology College, Northeastern University, Qinhuangdao, Hebei, China

Keywords: Emotion Recognition, Emotion Linkage, Human-Vehicle Interaction, Intelligent Cockpit.

Abstract: With the evolution of intelligent cockpit technology and human-vehicle interaction systems, the ability to recognize and regulate emotions in vehicles has increasingly become a key research direction in intelligent driving. Existing studies mostly focus on the perception and intervention of the emotional state of the main driver. However, in actual driving scenarios, the co-driver, as an important interactive subject, has a significant interrelationship with the emotional state of the main driver, which may have a profound impact on driving behavior and driving safety. In response to this relatively weak research area, this paper reviews the latest progress in the recognition of the emotional states of the main and co-drivers, and then focuses on the transmission mechanism and behavioral impact path of the emotional linkage between them. Finally, based on the analysis of the challenges and gaps in current research, the research trends in the future in the directions of linkage modeling, multimodal fusion, and human factor-adaptive interaction are discussed, aiming to provide a theoretical basis and practical reference for building a more intelligent and collaborative emotional perception human-vehicle interaction system.

1 INTRODUCTION

In the current era, the intelligence of smart cabins has been significantly enhanced. The human-machine interaction scenarios inside the vehicle are becoming increasingly complex. Driver emotion perception and regulation are vital for maintaining road safety and enhancing human-vehicle interaction (Li et al., 2023; Guo et al., 2023). Against this backdrop, emotion recognition technology has gradually expanded from a single visual or auditory modality to a multi-modal fusion direction, including expressions, postures, speech content, and some internal physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), respiration (RSP), and temperature (T), continuously enhancing the emotion perception capability of the cabin system (Wu, 2023). However, current research mainly focuses on the emotion recognition and intervention mechanisms of the main driver. In actual driving environments, the co-driver is also a high-frequency interaction object, and there is a potential interrelationship between their emotional states and those of the main driver, which has a significant impact on the driving behavior of the main driver and the system response.

Unlike ordinary passengers, in real driving scenarios, the co-driver often participates in interactions such as navigation, reminders, and communication, and is an important factor influencing the emotional state of the main driver. The emotional linkage between the main and co-drivers may affect each other through various means such as language, facial expressions, tone changes, and body postures, and thereby indirectly influence driving behavior and the intelligent response of the vehicle system. However, current research on such linkage phenomena is still in its infancy. Most existing studies lack a systematic framework, and the modeling of emotional contagion, emotional synchronization, and their transmission paths remains fragmented. Moreover, constructing a multi-agent emotional linkage model in intelligent cabins still faces many challenges, such as the alignment of multimodal information during data collection, the trade-off between real-time performance and accuracy, and individual differences in emotional

[a] https://orcid.org/0009-0003-3274-3087

Wang, Q.
Research on the Emotional Recognition and Interactive Influence Mechanism of the Main and Co-Pilots.
DOI: 10.5220/0014350100004718

In *Proceedings of the 2nd International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2025)*, pages 239-245
ISBN: 978-989-758-792-4

responses, all of which limit the in-depth advancement of related research.

Therefore, from the perspective of intelligent cabin interaction research, this paper first reviews the key technologies and research progress in the emotion recognition of main and co-drivers, clarifying the application basis in multimodal perception and individual identification. On this basis, it further focuses on the formation mechanism, transmission path, and impact on driving behavior of the emotional linkage between the main and co-drivers, revealing the dynamic interaction characteristics of emotional states among multiple subjects in the vehicle. In response to the challenges and deficiencies in multi-agent modeling, linkage perception, and response strategy design in existing research, it deeply analyzes the modeling methods and recognition frameworks of emotional linkage, and proposes design ideas for an interaction system oriented towards multi-agent joint perception and collaborative regulation. The related research is conducive to improving the existing vehicle emotion recognition system, providing theoretical support and practical paths for enhancing the emotional interaction capabilities of intelligent cabins and building a safer and more stable in-vehicle emotional ecosystem.

2 DRIVER AND CO-DRIVER EMOTION RECOGNITION

2.1 Emotion Recognition of the Main Driver

Within intelligent cockpit systems, accurately identifying the driver's emotional state serves as a critical component for enhancing both road safety and customized user experiences. The current mainstream research focuses on facial emotion recognition. Xiao et al. proposed a road driver emotion recognition method based on facial expressions called FERDERnet. The method divides the recognition task into three modules: first, the driver's face is located through the face detection module; second, the data is expanded and balanced using the enhancement-based resampling module; in the concluding stage, a deep convolutional neural network, pre-trained on FER and CK+ databases then optimized through fine-tuning, performs the driver's emotional state classification. This method integrates five different backbone networks and optimizes them with an integration strategy. To verify the effectiveness of the method, the authors constructed a driver facial expression dataset containing a variety of real road scenes. Experimental results show that FERDERnet, which uses Xception as the backbone network, outperforms the baseline network and some advanced methods in terms of recognition accuracy and processing efficiency, and shows excellent performance in real road environments (Xiao et al., 2022).

In addition, some studies have also used changes in the driver's tone, speaking speed, volume, etc. to judge his emotional state. Meng et al. proposed a new deep learning architecture ADRNN (composed of dilated convolution, residual block, BiLSTM and attention mechanism) for speech emotion recognition (Meng et al., 2019). This method first converts the original speech signal into a three-dimensional Log-Mel spectrogram as input features, and uses a dilated convolutional network to expand the receptive field, skip connections to retain shallow historical information, BiLSTM to learn long-term dependencies, and attention mechanism to further enhance key feature extraction. In addition, the authors introduced a combination of softmax and center loss in the loss function to improve classification performance. The experiment was evaluated on two commonly used emotional speech databases, IEMOCAP and Berlin EMODB. Experimental findings demonstrated that the proposed approach achieved a speaker-dependent accuracy of 74.96% and a speaker-independent accuracy of 69.32% in the IEMOCAP database, which were better than the 64.74% of previous methods. Evaluation of the EMODB dataset showed significant improvements, yielding 90.78% accuracy for speaker-dependent scenarios and 85.39% for speaker-independent cases, outperforming prior results of 88.30% and 82.82%. In addition, the method also showed good robustness and generalization ability in cross-corpus experiments, achieving a recognition accuracy of 63.84% (Meng et al., 2019). Tang et al. developed a novel end-to-end architecture for speech emotion recognition that combines dilated causal convolutions with context stacking. Their design incorporates parallel processing blocks that expand the model's receptive field to encompass complete input sequences while maintaining computational efficiency. Additionally, the incorporation of context stacking enhances the model's capacity to capture long-range dependencies. Experiments in the regression and classification tasks of emotion recognition show that the model achieves better recognition performance with only about one-third of the parameters of the current mainstream end-to-end model. In addition, the authors compared the impact of different input representations (raw audio vs Log-Mel spectrogram), verified the advantages of the end-to-end learning method over hand-crafted features, and demonstrated that the model can effectively extract embedded features that retain emotional information in the intermediate layers (Tang et al., 2021).

At present, the application of driver emotion recognition technology has gradually achieved a leap from single modality to multimodal fusion, combining vision, voice and even some physiological signals to achieve a more comprehensive perception of the driver's emotional state. Li et al. used CogemoNet to enhance driver emotion recognition with cognitive features. This method is different from the traditional research method of single modality information. The driver's facial expression is used together with cognitive features for research. The research team constructed a multimodal dataset containing facial videos, cognitive feature data and self-emotional assessment of 40 drivers. The experimental results show that CogemoNet shows good cross-database recognition performance on both discrete emotion models and dimensional emotion models, proving its effectiveness and superiority in the driver emotion recognition task (Li et al., 2021). Mou et al. proposed a new multimodal fusion framework based on a convolutional long short-term memory network (ConvLSTM) to recognize driver emotions. This method is the first to integrate non-invasive eye movement features, vehicle dynamics data and environmental information with driving context features to comprehensively model the emotional state in driving situations. The experimental data was collected on a highly simulated driving simulator platform, which simulated real road scenes through hydraulic motion systems, sound systems and visual simulation systems. The simulation environment supports a variety of weather conditions (such as rain and fog), time (day and night) and road curvature changes, inducing drivers to have diverse emotional states. The proposed model was verified in multiple scenarios and multiple subjects. Following the "one scenario left out per subject" evaluation protocol, the system accomplished mean accuracy scores of 97.64% in valence prediction, 97.27% in arousal detection, and 96.47% in dominance estimation; in the "leave one subject" experiment, the accuracy rates of the three dimensions were 88.16%, 81.65% and 85.34% respectively, and the recall rates reached 80.97%, 72.66% and 83.62%. In addition, the ablation experiment further revealed the different effects of different modal features on the recognition performance of each emotion dimension, providing a reference for multi-task modeling (Mou et al., 2023).

2.2 Emotion Recognition of the Co-pilot

At present, the emotion recognition research of smart cockpit is still mainly focused on the main driver. As the primary operator of vehicles, drivers' emotional states have been extensively researched for their impact on driving performance. However, as the frequent interaction object in the car, the co-pilot is often simplified to the identity of "passenger", focusing on building and improving the entertainment and comfort of the smart cockpit to meet its various needs (Liu, Shi, & Jiang, 2021). The system assumes that its emotions have little impact on driving, and has not yet built an independent emotion modeling framework for the co-pilot. As an active interactive participant in the "third life scene", its emotional changes may also affect the cognitive load and psychological state of the main driver through language communication, facial expression feedback or even silent attitude, but related research has not been systematically carried out.

During the driving task, the co-pilot is often in a non-dominant task state and lacks a clear interaction goal. Its emotional change mechanism is more hidden and unstable. At the same time, it also lacks a standardized annotation system and behavior label, which makes it difficult to directly apply the existing emotion recognition model. On the other hand, there are more visual occlusions and perspective offsets in the co-pilot area, which further increases the difficulty of facial image acquisition and emotion analysis (Yu, 2022). In voice interaction, the co-pilot talks to the main driver more as a "cooperative participant". There are also high individual differences in the frequency and content structure of their speeches, which makes it difficult for the system to uniformly model them.

In response to the above challenges, future research can introduce emotion recognition methods based on body movements. Literature shows that body movements such as head, arms, and body postures can express emotions to a certain extent (Tracy & Robins, 2004; Dael, Goudbeek, & Scherer, 2013), which can be applied to the emotion recognition of co-pilots, which are more severely restricted by light, viewing angle, etc.

3 EMOTIONAL LINKAGE AND INFLUENCE MECHANISM BETWEEN THE MAIN DRIVER AND CO-DRIVER

In recent years, in the research of group intelligent interaction, social robots and multi-user human-computer interaction, multi-agent emotion recognition has gradually developed into one of the core directions of affective computing. Multi-agent scenarios usually involve more than two interacting subjects, whose emotional states not only change independently, but also have complex linkage relationships, such as emotional synchronization, infection, and confrontation. In this closed, high-frequency communication environment, the emotional dynamics between the driver and the co-pilot are more linked, and the relevant experience provides a modeling perspective that can be used for reference in similar scenarios.

3.1 Emotional Contagion and Transmission Mechanism

Emotional contagion refers to the flow of emotions between people (Van Haeringen, Gerritsen, & Hindriks, 2023). In the closed and high-frequency interactive space of the smart cockpit, there is often a significant emotional resonance and emotional contagion effect between the driver and the co-pilot. Based on the "emotional contagion" theory of social psychology, when one driver has obvious emotions (such as anxiety, anger, and tension), the other driver is easily affected unconsciously and shows a synchronized emotional response (Pinus et al., 2025). Emerging research indicates that the transmission of emotions in multi-agent interactions often presents asymmetry. In the in-car interactive scene, the main driver's emotional state is often more likely to affect the co-pilot due to his dominant position in vehicle control. That is, the main driver's emotions often have a stronger guiding effect on the co-pilot. There may also be nonlinear dynamics, such as emotion amplification or delayed response. In the in-car scene, this emotional linkage may be achieved through multimodal channels such as language, expression, and body movements. Its transmission mechanism has cross-modal, multi-stage, and multi-path characteristics, which still need to be systematically modeled.

3.2 The Mediating Effect of Emotional Influence on Driving Behavior

The emotional linkage between the driver and the co-pilot not only changes each other's psychological state, but also may indirectly regulate the driving behavior of the driver. Studies have shown that emotions may lead to aggressive driving operations, distracted attention or slow response, thus affecting driving behavior (Ma, Xing, Wu, & Chen, 2024). Driven by anger, speeding behavior will increase, thereby increasing the possibility of traffic accidents (Habibifar & Salmanzadeh, 2022). Fear emotions will lead to increased heart rate, mental tension, and reduced concentration time, resulting in a slow response of the driver, thereby increasing the probability of operational errors (Samuel et al., 2019). The cognitive burden induced by stress negatively impacts drivers' ability to maintain optimal driving performance. (Halim & Rehan, 2020). When the co-pilot shows anxiety or excessive intervention, the emotional stress level of the driver will increase significantly, which may trigger defensive or confrontational behavior patterns. Furthermore, emotional changes are often reflected in driving behavior as quantifiable indicators such as steering angle, braking frequency, and acceleration fluctuations. Therefore, emotional linkage is not only the object of emotion recognition, but also an important mediating variable for understanding driving risk status.

3.3 Linkage Recognition Framework and Interactive System

For the emotional linkage characteristics of the driver and the co-driver, the emotional states of multiple people can be constructed into an emotional propagation map through a graph neural network (GNN) (Gao & Wang, 2024). On this basis, in the future, we can try to build a linkage recognition framework that integrates multimodal input. This type of framework generally includes three key modules: an emotion perception module that uses signals such as voice, expression, posture, and eye movement to extract individual emotional characteristics; a linkage modeling module that uses a graph neural network, a temporal neural network, or a causal reasoning mechanism to construct the emotional propagation path between drivers; and an interactive feedback module that dynamically adjusts the cabin lighting, voice assistant, or driving assistance prompts based on the recognition results to achieve an emotional response closed loop. For example, some intelligent cockpit systems can warn the driver of potential tension by detecting the emotional fluctuations of the co-driver's voice, and reduce the volume of the audio system in the cockpit in a timely manner to help maintain driving concentration. The key to this type of system lies in the dynamic understanding of the emotional relationship between multiple subjects and the construction of a real-time adaptive feedback mechanism, marking an important transition from perception to intervention in vehicle-mounted emotional intelligence.

4 CURRENT CHALLENGES AND RESEARCH GAPS

4.1 Difficulties in Collecting Multi-Subject and Multi-modal Data

In actual driving environments, the synchronous collection of multi-subject and multi-modal data in the car faces great technical challenges. First, due to the physical space layout of the smart cockpit, there may be problems such as occlusion, posture deflection and uneven light between the driver and the co-pilot, especially the co-pilot's side face and eye movement features are more likely to be occluded, affecting data integrity. Secondly, there are natural differences in sampling frequency, timing granularity and alignment mechanism between visual, voice, physiological signals and other modalities, and traditional synchronous fusion methods are difficult to adapt to this asynchronous characteristic. In addition, for privacy and security reasons, it is difficult to obtain high-quality, long-term, and multi-channel data in real environments. At present, most studies still rely on simulation scenarios, lacking a large-scale, natural interaction dataset of the driver and co-pilot emotional linkage, which restricts the model's generalizability and hinders the advancement of research findings.

4.2 Difficulty in Dynamic Modeling of the Emotional Linkage Mechanism

The emotional linkage between the driver and the co-pilot has strong interpersonal interaction properties, and its propagation process is affected by multi-factor coupling, individual differences and emotional asynchrony. Different from the single-person recognition task, emotion linkage modeling needs to deal with complex dynamic features such as emotion source recognition, propagation direction and intensity estimation. At present, most models focus on individual modeling, lack relationship expression and cross-modal and cross-time modeling capabilities, and have not yet formed a unified linkage graph construction method. At the same time, the relationship between individual emotion changes and driving behavior feedback is complex, and the causal chain is difficult to explicitly construct, which restricts the in-depth understanding and mechanism mining of the linkage mechanism.

4.3 System Performance Trade-off Between Real-Time and Accuracy

Smart cockpits place strict requirements on the real-time performance of emotion recognition systems, and perception and response must be completed within sub-seconds. However, although existing deep learning models have good expressiveness, they consume large computing resources and are difficult to run efficiently on vehicle-mounted edge devices, often limited by latency and power consumption. There is currently a lack of a unified optimization framework for fusion model compression, inference acceleration and modality selection mechanisms. At the same time, multimodal data has noise and redundancy problems, and blind fusion may reduce recognition accuracy instead, and dynamic scheduling mechanisms need to be developed urgently.

5 FUTURE RESEARCH DIRECTIONS AND TRENDS

5.1 Construction of Multimodal Graph Model for Linkage Identification

In the future, a graph model of driver-copilot emotional interaction can be constructed based on a graph neural network (GNN), with the driver set as a graph node and interaction events of different modes modeled as edges to express "who influences whom", "with what emotion" and "through what mode" for transmission. Combined with time series modeling mechanisms such as Transformer, the direction and strength of the transmission chain can be captured, "emotion source nodes" and "highly sensitive nodes" can be identified, and the modeling ability of linkage emotions in complex interaction situations can be improved.

5.2 Design of Personalized and Adaptive Intelligent Cockpit Emotion Engine

For the problem of significant individual differences, a collaborative modeling framework that integrates individual characteristics and linkage modes should be developed. Through methods such as federated learning and transfer learning, personalized modeling can be achieved under the premise of protecting privacy; at the same time, historical linkage trajectory analysis is introduced to achieve emotional evolution prediction and early warning of the driver-copilot combination. The system can actively adjust the interaction atmosphere based on the current state, improve positive feedback and emotional synchronization, and enhance collaborative stability.

5.3 Development of Multi-Scenario Highly Robust Emotion Recognition System

Smart cockpits need to adapt to a variety of scenarios including commuting, family travel, and long-distance driving. The emotional interaction mode between the driver and the co-pilot may also migrate with the situation. Therefore, it is urgent to build a linkage recognition system with scene adaptability. On the one hand, a cross-scenario linkage database can be established to train a recognition model with universal adaptability to typical linkage modes; on the other hand, a context-aware mechanism (such as task intensity, in-car noise, time period, etc.) is introduced to dynamically adjust the weights and recognition thresholds of each modality.

5.4 Fusion of Risk Intervention Strategies Driven by Emotion Perception

Current intervention systems focus more on the driver and ignore the regulatory role of the co-pilot. In the future, a dual-subject collaborative intervention framework based on emotional linkage relationships can be constructed: for example, in the case of anxiety-indifference combination, the co-pilot is guided to intervene in communication; in the case of double negative emotions, music, lighting and other soothing means are used to intervene. Further combining the linkage path with the feedback of the intervention effect, a closed-loop system of "recognition-response-regulation" is constructed to improve the in-cabin emotional management and safety assurance capabilities.

6 CONCLUSIONS

This paper focuses on the research issues of the emotional linkage and influence mechanism between the main and co-pilot in the intelligent cockpit environment. With the accelerated advancement of autonomous vehicle technologies, emotion recognition, as a key technology to improve the in-vehicle human-computer interaction experience and driving safety, has received widespread attention. However, existing research mostly focuses on the main driver, ignoring the emotional state of the co-pilot and its potential impact on the driving process, and the emotional linkage relationship between the main and co-pilot is still lacking in systematic discussion.

To fill this gap, this paper sorts out the current emotion recognition methods and research basis around the emotional linkage problem between the main and co-pilot, focusing on the emotional propagation mechanism and behavioral mediation effect, and proposes a corresponding recognition framework and interactive system ideas. On this basis, the main challenges in this field in terms of data collection, dynamic modeling and system performance are summarized, and it is pointed out that current research still has problems such as insufficient multi-agent collaborative modeling and scene adaptability. Looking to the future, this paper proposes key directions such as building a linkage graph model, developing a personalized emotional engine, improving system robustness and designing a closed-loop intervention mechanism, which provides theoretical support and research reference for achieving more efficient and stable emotional collaboration in intelligent cockpits.

REFERENCES

Dael, N., Goudbeek, M., & Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. Perception, 42(6), 642‑657.

Gao, S., & Wang, Y. (2024). A review of group emotion recognition based on images. Computers and Modernization, (08), 98‑107.

Guo, Y., Zou, Y., Xu, C., & Cao, D. (2023). A review of emotion recognition research in smart cockpit scenarios. In Proceedings of the 27th Annual

Conference on New Network Technologies and Applications (pp. 11-15).

Habibifar, N., & Salmanzadeh, H. (2022). Relationship between driving styles and biological behavior of drivers in negative emotional state. Transportation Research Part F: Traffic Psychology and Behaviour, 85, 245-258.

Halim, Z., & Rehan, M. (2020). On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. Information Fusion, 53, 66-79.

Li, W., Cao, D., Tan, R., Shi, T., Gao, Z., Ma, J., ... & Wang, L. (2023). Intelligent cockpit for intelligent connected vehicles: Definition, taxonomy, technology and evaluation. IEEE Transactions on Intelligent Vehicles, 9(2), 3140-3153.

Li, W., Zeng, G., Zhang, J., Xu, Y., Xing, Y., Zhou, R., ... & Wang, F. Y. (2021). Cogemonet: A cognitive-feature-augmented driver emotion recognition model for smart cockpit. IEEE Transactions on Computational Social Systems, 9(3), 667-678.

Liu, H., Shi, R., & Jiang, J. (2021). Discussion on the development trend of automotive intelligent cockpit in the 5G communication era. Journal of Guangdong Communications Vocational and Technical College, 20(01), 33-37.

Ma, Y., Xing, Y., Wu, Y., & Chen, S. (2024). Influence of emotions on the aggressive driving behavior of online car-hailing drivers based on association rule mining. Ergonomics, 67(10), 1391-1404.

Meng, H., Yan, T., Yuan, F., & Wei, H. (2019). Speech emotion recognition from 3D log-mel spectrograms with deep learning network. IEEE Access, 7, 125868-125881.

Mou, L., Zhao, Y., Zhou, C., Nakisa, B., Rastgoo, M. N., Ma, L., ... & Gao, W. (2023). Driver emotion recognition with a hybrid attentional multimodal fusion framework. IEEE Transactions on Affective Computing, 14(4), 2970-2981.

Pinus, M., Cao, Y., Halperin, E., Coman, A., Gross, J. J., & Goldenberg, A. (2025). Emotion regulation contagion drives reduction in negative intergroup emotions. Nature Communications, 16(1), 1387.

Samuel, O., Walker, G., Salmon, P., Filtness, A., Stevens, N., Mulvihill, C., ... & Stanton, N. (2019). Riding the emotional roller-coaster: Using the circumplex model of affect to model motorcycle riders' emotional state-changes at intersections. Transportation Research Part F: Traffic Psychology and Behaviour, 66, 139-150.

Tang, D., Kuppens, P., Geurts, L., & van Waterschoot, T. (2021). End-to-end speech emotion recognition using a novel context-stacking dilated convolution neural network. EURASIP Journal on Audio, Speech, and Music Processing, 2021(1), 18.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. Psychological Science, 15(3), 194-197.

Van Haeringen, E. S., Gerritsen, C., & Hindriks, K. V. (2023). Emotion contagion in agent-based simulations of crowds: A systematic review. Autonomous Agents and Multi-Agent Systems, 37(1), 6.

Wu, L. (2023). Research on driver anger emotion recognition and regulation methods (Master's thesis, Chongqing University).

Xiao, H., Li, W., Zeng, G., Wu, Y., Xue, J., Zhang, J., ... & Guo, G. (2022). On-road driver emotion recognition using facial expression. Applied Sciences, 12(2), 807.

Yu, M. (2022). Research on the current status and development trend of automobile intelligent cockpit design. Engineering Management, 3(5), 187-189.