# Frustration-Aware Assistance Systems

## Assessment and Causes of In-Vehicle Frustration

vorgelegt von
M. Sc.
Esther Bosch

an der Fakultät V - Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:
Vorsitzende: Prof. Eva Wiese
Gutachterin: Prof. Meike Jipp
Gutachterin: Prof. Christine Ahrend
Gutachter: Prof. Martin Baumann

Tag der wissenschaftlichen Aussprache: 29. Juni 2023

Berlin 2023

Digital Art as produced by DALL · E with the keyword 'frustrated driver with in-vehicle frustration-aware assistance system.' Source: https://openai.com/dall-e-2/.

# Abstract

Frustration is an emotion that occurs when goal-directed behavior is hindered. This frequently happens in transportation, for example in driving. Problematically, frustration can lead to risky driving behavior when a human driver is in charge of the vehicle. Furthermore, frustration may hinder the acceptance of innovative automated vehicles. The option to automatically recognize negative emotions, such as frustration, of vehicle users with so-called affect-aware systems has gained increasing attention within the last few years. These systems enable to adapt vehicle functions, such as the human-machine interface, in real-time depending on the current traveler state and corresponding needs. However, the automated recognition of emotion requires high-quality data sets to train algorithms on. These are insufficiently present in the affective computing community so far. Furthermore, a wide variety of measures for affect recognition exist, but methods to compare different modalities for measures of frustration are lacking. Previous research found that emotional expressions in the face and body form promising indicators for user frustration. Previous studies have investigated expressions of frustration in the context of driving and mobility but have neglected interindividual differences. Furthermore, knowledge of possible causes of frustration is needed to successfully mitigate frustration, which are yet unknown. To properly design frustration-aware systems and to develop methods to capture frustration, it is, therefore, necessary to 1) provide a training dataset, 2) find a method to compare different modalities of frustration recognition, 3) improve recognition of frustration by facial expressions and 4) investigate causes of frustration in driving. This dissertation did so and thereby enabled a more reliable recognition of frustration in driving. This could make manual driving safer and contribute to the development of affect-aware systems, which also have the potential to facilitate the acceptance of automated driving systems. In total, this dissertation presents three studies published in four papers. Paper 1 of this dissertation presents high-quality and continuously frustration-labeled expression data that we provide for the research community. It contains a thorough description of the data and a benchmark algorithm that automatically recognizes frustration in video data. Paper 2 found that in addition to previously described frustration-typical expressions, individual-typical expressions of frustration exist. Paper 3 presents a latent variable model that can evaluate which measurements for frustration are necessary. Finally, paper 4 investigated causes and coping strategies for frustration in driving through a diary study and a focus group study. The overall goal of this dissertation is to contribute to the underlying research necessary to develop frustration-aware assistance systems. Based on the findings of the three studies, this dissertation helps to expand our knowledge of

how to measure in-vehicle frustration. The discussion highlights this dissertation's contributions for such a development, but also points out limitations of the current studies. Ethical aspects of automated emotion recognition are discussed.

# Zusammenfassung

Frustration ist eine Emotion, die auftritt, wenn ein zielgerichtetes Verhalten behindert wird. Dies geschieht häufig im Verkehr, zum Beispiel beim Autofahren. Problematisch ist, dass Frustration zu riskantem Fahrverhalten führen kann, wenn ein menschlicher Fahrer das Fahrzeug lenkt. Außerdem kann Frustration die Akzeptanz innovativer automatisierter Fahrzeuge behindern. Die Möglichkeit, negative Emotionen, wie z.B. Frustration, von Fahrzeugnutzern mit sogenannten affektbewussten Systemen automatisch zu erkennen, hat in den letzten Jahren zunehmend an Aufmerksamkeit gewonnen. Diese Systeme ermöglichen es, Fahrzeugfunktionen, wie z.B. die Mensch-Maschine-Schnittstelle, in Echtzeit an den aktuellen Zustand des Fahrers und die entsprechenden Bedürfnisse anzupassen. Die automatische Erkennung von Emotionen erfordert jedoch qualitativ hochwertige Datensätze, auf denen Algorithmen trainiert werden können. Diese sind in der Affective-Computing-Community bisher nur unzureichend vorhanden. Darüber hinaus gibt es eine Vielzahl von Messungen für die Erkennung von Emotionen, aber es fehlt an Methoden zum Vergleich verschiedener Modalitäten für die Messung von Frustration. Vorherige Forschungen haben ergeben, dass emotionale Ausdrücke im Gesicht und im Körper vielversprechende Indikatoren für die Frustration von Benutzern sind. Frühere Studien haben Frustrationsausdrücke im Zusammenhang mit dem Autofahren und der Mobilität untersucht, dabei aber interindividuelle Unterschiede vernachlässigt. Darüber hinaus sind Kenntnisse über mögliche Frustrationsursachen erforderlich, um Frustration erfolgreich zu mindern, die bisher noch unbekannt sind. Um frustrationsbewusste Systeme zu entwerfen und Methoden zur Erfassung von Frustration zu entwickeln, ist es daher notwendig, 1) einen Trainingsdatensatz bereitzustellen, 2) eine Methode zu finden, um verschiedene Modalitäten der Frustrationserkennung zu vergleichen, 3) die Erkennung von Frustration durch Gesichtsausdrücke zu verbessern und 4) die Ursachen von Frustration beim Fahren zu untersuchen. Diese Dissertation hat dies getan und damit eine zuverlässigere Erkennung von Frustration beim Autofahren ermöglicht. Dies könnte das manuelle Fahren sicherer machen und zur Entwicklung von affektbewussten Systemen beitragen, die auch das Potenzial haben, die Akzeptanz von automatisierten Fahrsystemen zu erleichtern.

Insgesamt werden in dieser Dissertation drei Studien vorgestellt, die in vier Beiträgen veröffentlicht wurden. Paper 1 dieser Dissertation präsentiert qualitativ hochwertige und kontinuierlich mit Frustrationsmarkern versehene Expressionsdaten, die wir der Forschungsgemeinschaft zur Verfügung stellen. Er enthält eine ausführliche Beschreibung der Daten und einen Benchmark-Algorithmus, der automatisch Frustration in Videodaten erkennt. In Beitrag 2 wurde festgestellt, dass es neben den zuvor beschriebenen frustrationstypischen Ausdrücken auch individualtypische Ausdrücke von Frustration gibt. In Beitrag 3 wird ein latentes Variablenmodell vorgestellt, mit dem bewertet werden kann, welche Messungen für Frustration notwendig

sind. In Beitrag 4 schließlich wurden Ursachen und Bewältigungsstrategien für Frustration beim Autofahren anhand einer Tagebuchstudie und einer Fokusgruppenstudie untersucht. Das übergeordnete Ziel dieser Dissertation ist es, einen Beitrag zur Grundlagenforschung zu leisten, die für die Entwicklung frustrationsbewusster Assistenzsysteme notwendig ist. Auf der Grundlage der Ergebnisse der drei Studien trägt diese Dissertation dazu bei, unser Wissen über die Messung von Frustration im Fahrzeug zu erweitern. In der Diskussion werden die Beiträge dieser Dissertation für eine solche Entwicklung hervorgehoben, aber auch die Grenzen der aktuellen Studien aufgezeigt. Ethische Aspekte der automatischen Emotionserkennung werden diskutiert.

**List of Publications**

Bosch, E., Corbí, R. L. H., Ihme, K., Hörmann, S., Jipp, M., & Käthner, D. (2022). Frustration Recognition Using Spatio Temporal Data: A Novel Dataset and GCN Model to Recognize In-Vehicle Frustration. IEEE Transactions on Affective Computing. DOI: https://doi.org/10.1109/TAFFC.2022.3229263. Postprint.

Bosch, E., Käthner, D., Jipp, M., Drewitz, U., & Ihme, K. (2023). Fifty shades of frustration: Intra-and interindividual variances in expressing frustration. Transportation research part F: traffic psychology and behaviour, 94, 436-452. `https://doi.org/10.1016/j.trf.2023.03.004`. Published Version.

Bosch, E., Klosterkamp, M., Guevara, A., Kaethner, D., Bendixen, A., & Ihme, K. (2022, September). Multimodal Estimation of Frustrative Driving Situations Using a Latent Variable Model. In 2022 13th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000011-000016). IEEE. `https://doi.org/10.1109/CogInfoCom55841.2022.10081636`. Postprint.

Bosch, E., Ihme, K., Drewitz, U., Jipp, M., & Oehl, M. (2020). Why drivers are frustrated: results from a diary study and focus groups. European transport research review, 12(1), 1-13. `https://doi.org/10.1186/s12544-020-00441-7`. Published Version.

# Acknowledgements

First of all, I would like to thank Meike J., for your guidance that was always combined with a sincere, encouraging and eye-to-eye atmosphere. Also, for a shared humor regarding christmas surprises and Stsetstiekmaskremfua. Furthermore, I would like to thank the NZM - research group: Klas I., for always having my back, for very successful emotion recognition regarding me, for being considerate, patient, empathic, respectful, and eye-to-eye, but also funny in just the right moments. Also, for sharing a 'müsste gehen' spirit. Thank you also for believing and trusting in me, and keeping project work away in moments I intensively worked on this thesis. I will take you as role model for every student I will supervise in the future. I would like to thank David K., for giving me a glimpse of how little I understand about statistics, which motivated me to discover a topic I now love. Also, for not giving up on me and our analysis in moments that anyone else certainly would have. Also, for asking questions that I hate, but that always change my work for the better. Further, I would like to thank Uwe D., for your contagious enthusiasm, your creativity, and your empathy in the moments that matter. Phuong N., for saving me in endless situations that concerned many cables, cryptic C++ code, or finding creative solutions. Of course, also for Figure 1.1 of this thesis, and a trip to Drewitz. I also thank the people that made the complex data acquisition possible in the technical sense, especially to Jan W. and Johannes R. in the driving simulator and Reza D., Jonas R and Claus K. in the automated car. I also thank Michael O. for helpful conversations in our office. Thank you also to my new group, Merle L., Duc Hai L. and Fabian W. I also thank Meng Z. for your advice on R, and for being a great co-thesising-buddy. Also, thank you to many of my colleagues at DLR for loving boardgames as much as I do.

From all the friends that accompanied me during the time, I would especially like to thank Theresa H., Sarah F., and Joachim F. for hours of conversation that kept me motivated in moments I could not see the end of this work, and for your advice in all other regards, too. You mean the world to me. Theresa H., also for Bluetooth-showers, lists in top drawers and epic fights with Greek grasshoppers. I would also like to thank Alexandra K. and Laura Q. for your support, your friendship and advice. My thanks also go to Nicolas B., for adventures from Qpark to Woog, your trust, loyalty and empathy, to Larissa G., for being my soulmate, to Sarah K., for being my travel companion, to Alexandre K., for your ideas, creativity and helpfulness, and for a shared love for zebra finches and Bostonian supervisors. To Felizitas E., for being adventurous and spontaneous with me, to Eva C. and Nina F., for advice on life and lectures, to all three of you for a shared enthusiasm for p53-pathways. To Tobias R., for your friendship and understanding advice, to Monica R., for an insight into a different way of seeing the world, to Sophia H., for sneeuwpoppen en chocoladeontbijt, to Vivian H., for

Charles River Standup-Paddling, meetings in all corners of the world, and showing me how to stay true to oneself.

Furthermore, I thank the students that accompanied me during the time. Especially, I would like to thank Marie K. and Raquel L. H. C. for your enthusiasm, and for your help that lead to the papers included in this thesis. I would also like to thank Anne R. for Gemüsedip and a big list of failures that lead to a summer that changed my life to the better, and for your friendship. I also thank the students that enabled me to spend significantly more time on my thesis by greatly helping in my other projects, especially Stefan B., Peru D., Helena W., David S., Charlotte B. and Ricarda L.

I would like to thank my family. To my parents, for giving me an endless amount of love, the courage to always walk one step ahead of my comfort zone, and for giving me the drive to go discover the world (including, but not limited to, the scientific world). Gregor, for being the best brother in the world, for eine Dingsbums-Krimskrams- UND Gerätehöhle!, for many of the best hours of my life, and for being a great advice-giver on top of that. We could conquer the world together.

Lastly, I would like to thank Thomas K. Regarding this work, you significantly (yes, significantly) helped in making the last, but necessary few steps of this thesis. Regarding everything else, for being exactly the human being that you are, and making us an amazing team.

# Table of Contents

# Acronyms

**AAI** frontal Alpha Asymmetry Index

**CANBUS** Controller Area Network

**ECG** Electrocardiogram

**EEG** Electroencephalogram

**FAAS** Frustration-Aware Assistance System

**fNIRS** Functional near-infrared spectroscopy

**GDPR** General Data Protection Regulation

**GPS** Global Positioning System

**HMI** Human-Machine-Interface

**HRV** Heart Rate Variability

**IARPA** Intelligent Advanced Research Project Agency

# 1

# Introduction

## 1.1 Frustration in Transportation

Frustration occurs when goal-directed behavior is blocked (Lazarus, 1991). Mehrabian and Russell (1974) proposed to describe emotions in a space of the dimensions of valence, arousal and dominance. The third dimension, dominance, helps to distinguish between emotions like anger and fear, which are both associated with high arousal and a negative valence. Fear, on the contrary, is accompanied by a feeling of withdrawal (low dominance) and anger is accompanied by a feeling of approach (high dominance) (Demaree et al., 2005). Frustration is defined by negative valence, high arousal, and low dominance (Reuderink, Mühl, and Poel, 2013). Persistent frustration leads to the experience of anger and aggressive behavior (Berkowitz, 1989; Myounghoon Jeon, 2015; Shinar, 1998), causes experienced stress (Myounghoon Jeon and Zhang, 2013), and negatively influences the acceptance of human-machine systems (Picard and Klein, 2002).

In driving, negative emotions, such as frustration, negatively affect cognitive skills required for the driving task (Myounghoon Jeon, 2015; Lee, 2010). They have been shown to impact driver performance as measured by lane excursions and lateral control of the car (G. M. Hancock, P. A. Hancock, and Janelle, 2012). Mesken et al. (2007) have found that participants who report being angry are more likely to exceed the speed limit than participants who did not. Myounghoon Jeon (2016) finds that participants who experience negatively valenced emotions like anger or sadness make more driving errors than in a neutral valenced drive. Generally, anger has been shown to be related to aggressive driving (Nesbit, J. C. Conger, and A. J. Conger, 2007; Precht, Keinath, and Krems, 2017)). Frustration is thought to be a precursor-emotion for anger (Berkowitz, 1989). It is, therefore, a suitable emotion to mitigate in order to prevent anger in-time. Many potential sources of frustration exist in modern transportation, where most participants need to reach their destinations as quickly as possible. These frustrators include, for example, traffic jams on the highway caused by congestion, accidents, or construction sites; red traffic lights during urban rush hour; or slow tractors driving ahead on rural roads. These numerous triggers of frustration and the link between frustration and aggressive behavior (Berkowitz, 1989), such as speeding or risky overtaking maneuvers, suggest that frustration contributes to the number of aggression-related accidents, some of which are fatal. Furthermore, it is speculated that people often compensate poorly for the momentary impairments of cognitive functions caused by negative emotions, as they are unaware of the corresponding effects, which is in contrast to impairments caused by distractions (e.g., mobile phone) (Myounghoon Jeon, 2012). In conclusion, frustration is a negative emotion that has been linked to various negative effects in driving.

These harmful impacts of frustration on driving and traffic are expected to be remedied by the development of automated cars, which is designed to optimize traffic flow in general. SAE defines six levels of automated driving, in which the human's role as traveler changes from driver to passenger. In levels zero to two, the human driver is fully responsible for the car's behavior. This changes from level three onward, where the automated driving system is responsible for the driving task while the 'automated driving features are engaged'[1]. The fact that humans no longer have to steer themselves in automated traffic means that, firstly, they can no longer exhibit aggressive driving patterns. This is because the automated vehicle will relieve the driver from unpleasant tasks and, moreover, will only give back control if the driver is ready to drive (Braunagel, Rosenstiel, and Kasneci, 2017). Secondly, they can use their time in the car sensibly so that the loss of time due to blocking traffic is no longer so significant. Steck et al. (2018) showed that the value of travel time savings can be reduced by 31% by automated driving vehicles. They also showed that automated driving is perceived 10% less negative than manual driving (Steck et al., 2018). However, the road to fully autonomous driving still seems a long way off, so the promise of perfectly functioning, "frustration-free" traffic will still be several years to decades away. Thus, humans will still have to take the wheel and be needed as (part-time) drivers in certain situations. Additionally, because autonomous vehicles are highly technical systems that are difficult for the common user to understand, it is difficult for engineers and designers to create them in an understandable way. As a result, when interacting with autonomous vehicles, especially the first generation, users will probably become frustrated often. In these cases, frustration can translate into a negative user experience (cf. (Picard and Klein, 2002), which can negatively affect the evaluation and acceptance of automated vehicles. Traveler's acceptance is highly relevant in transportation, especially for a shift towards innovative and sustainable mobility solutions.

In summary, frustration experienced during driving negatively impacts the overall safety of driving, as well as the user experience and thus acceptance of automated vehicles. Reducing frustration is, therefore, highly desirable. Frustration-inducing factors come in many forms, differ from person to person (Ceaparu, 2004), are not always predictable and therefore cannot be avoided by design. Because of this, an increasing amount of literature has been written in recent years about how to recognize frustration. For example, Zepf, Dittrich, et al. (2019) presented a support vector machine that recognizes moments of frustration in driving. In a next step, they used this classification to initiate frustration mitigation by either ambient light or a driving assistant (Zepf, Dittrich, et al., 2019). They found a trend for a more positive user experience when the frustration detection was active and triggered the voice assistant, than when it was triggered at random. In conclusion, frustration is highly relevant in manual

[1] https://www.sae.org/news/2021/06/sae-revises-levels-of-driving-automation

and automated driving research. Its reduction is of interest to make manual driving safer and future mobility concepts like automated driving attractive to use. One option to achieve this could be the development of frustration-aware assistance systems, which can recognize and mitigate frustration in real-time.

### 1.1.1 Frustration-Aware Assistance Systems

One idea to reduce frustration in cars is to design emotion- or, more specifically, frustration-aware assistance systems (Bruce, 1993; Harris and Nass, 2011; Krüger et al., 2021; McDuff and Czerwinski, 2018; Oehl et al., 2019; Picard and Klein, 2002; Stephan, 2015) that can detect a traveler's current level of frustration, derive the traveler's current needs and offer specific assistance. Driver assistance systems in general are defined as technological systems that assist the human driver (Kukkala et al., 2018). Examples already implemented nowadays are lane keep assistant or the parking assistant. Already Bruce (1993) suggested to develop systems that could help robots to understand human faces and their expressions to improve communication. Klein, Moon, and Picard (1999) were the first to propose a computer that reacts to a user's frustration through a text display. Their results showed that frustration levels as measured by subsequent game interaction time are significantly lower in the condition with the affective text display compared to the control conditions.

After detection of frustration, the frustration-aware assistance system aims at either reducing the traveler's level of frustration and bringing him or her to a different target state (e.g. relaxation, pleasure, or high attention) or mitigating the negative consequences of frustration by providing support. Braun, Pfleging, and Alt (2018) have shown that drivers would like to use a system to help them mitigate negative emotions. The following user story gives an exemplary interaction between a human and a frustration-aware assistance system (see Figure 1.1).

*Peter is on his way to a work appointment with external project partners. He just picked up the rental car and starts driving towards the highway. He starts the car's navigation system to put in the address. Instead, the radio starts playing very loudly. He quickly turns down the volume and tries to find the button to change from radio to navigation system. After a few tries of different buttons, he finally sees a map, but it shows an entirely wrong destination. He also cannot find the usual input bar for the address. Outside, the entry to the highway's different directions is getting closer and closer, and Peter starts to become very frustrated and hectic... Suddenly, a friendly voice asks Peter: 'hey, I can see you are trying to enter an address. You can either just tell me where you want to go or use the drop-sign in the lower right corner to type it in.' Peter is surprised but answers with the address he'd like to go to. The navigation system shows on which highway to turn just in time, and Peter can relaxedly continue his journey.*

An interaction between a traveler in a context and a frustration-aware assistance system is necessary for such a system. Figure 1.2 shows an overview of how such a system would work. The traveler has long-term attributes such as age and gender and short-term attributes like changes in physiology or expressions. The traveler exists in a time- and location-dependent context but is also influenced by upcoming appointments in the calendar, the use of technical artifacts like a cell phone or navigation system, or surrounding traffic. The traveler's attributes can be

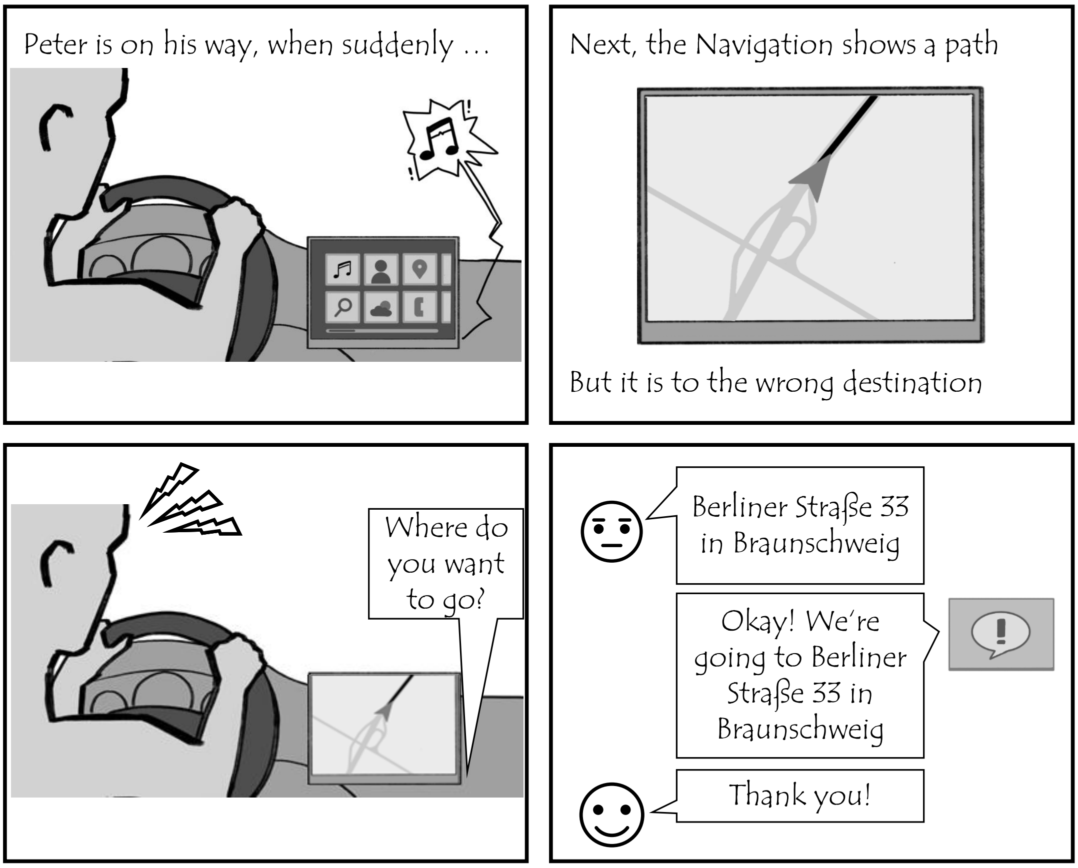


**Figure 1.1:** User story of Peter, trying to interact with his rental car.

captured by a camera, Electrocardiogram (ECG), skin conductance, or Electroencephalogram (EEG). Smartphone data (in the case of the calendar), Global Positioning System (GPS) data (e.g. traffic and weather), and camera data (for the use of technical artifacts) can capture the context's attributes. The system can use captured traveler attributes to recognize changes typical for frustration. Concomitantly, the captured context attributes can be utilized to detect situations that typically cause frustration. By this, a decision rule that combines the input from the traveler and context frustration detection can recognize moments of traveler frustration. A mitigation strategy suitable for the situation can then be initialized and executed. This process, in turn, leads to a change in context and traveler state.

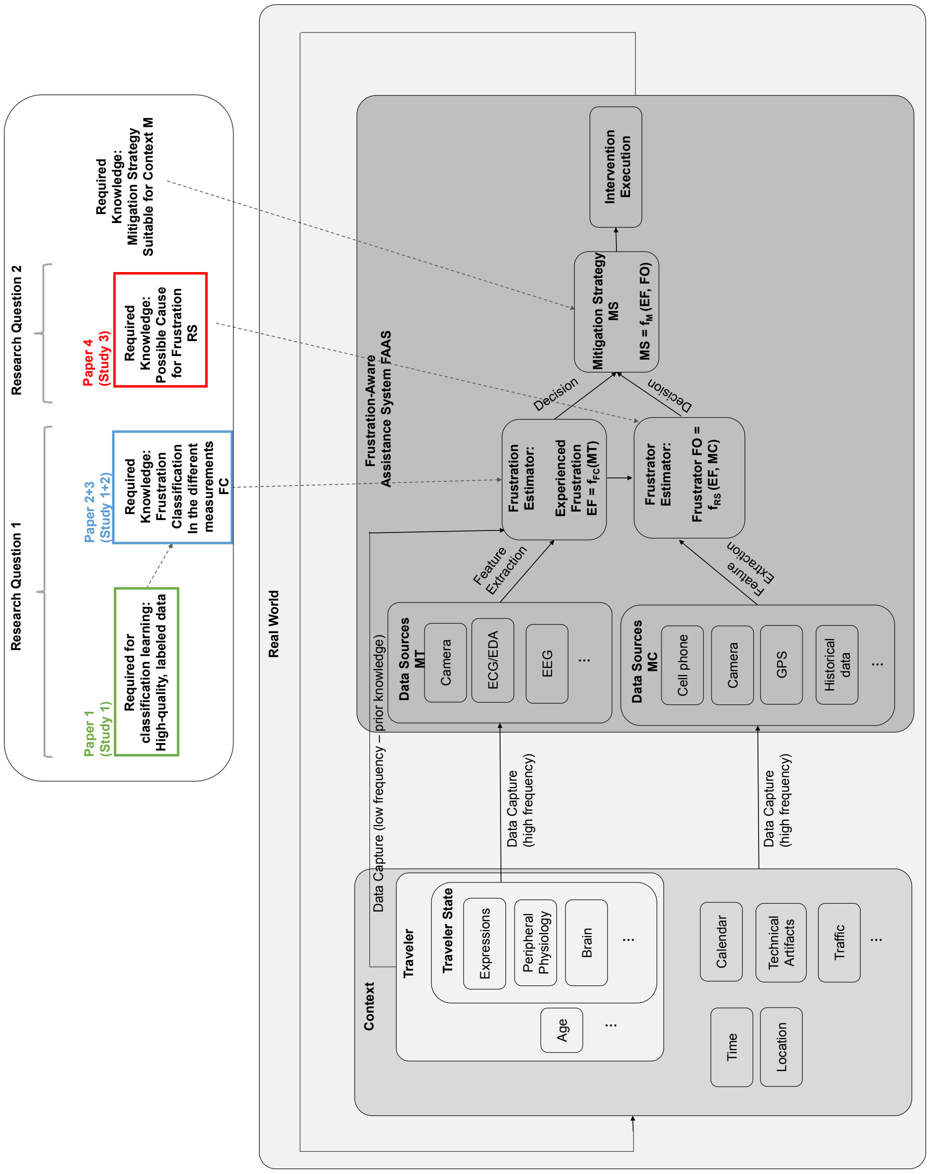


**Figure 1.2:** Overview of how a Frustration-Aware Assistance System would work.

#### 1.1.1.1 Context and Traveler

The traveler exists in a context that is, among others, influenced by factors like upcoming plans, an interaction of the traveler with different technical agents like the car-human interface, and factors outside the vehicle, like traffic situation or weather. Braun, Pfleging, and Alt (2018) find that negative emotions are mostly triggered by traffic-related incidents. Positive emotions mainly occur due to pleasant surroundings, music, ride enjoyment, and personal interaction (Braun, Pfleging, and Alt, 2018). Generally, the traveler has long-term attributes like age and gender and short-term attributes connected to the current traveler state. Regarding a Frustration-Aware Assistance System (FAAS), the traveler state of interest is the emotional traveler state. According to Scherer and Moors (2019), facial, vocal, and gestural expressions and physiological processes accompany the emotion process. These user states are important for the frustration-aware assistance system because sensors placed in the vehicle can measure them.

### 1.1.2 Frustration-Aware Assistance System

#### 1.1.2.1 Data Sources and Estimation - Traveler

Data that is captured from the traveler or the context, respectively, are the basis of the FAAS. Regarding the traveler, sensors assess video data (by a camera), peripheral physiology data (by ECG and skin conductance measurement), and brain activity data (by EEG or Functional near-infrared spectroscopy (fNIRS)), for example. From this data, relevant features are extracted. These can be, for example, action units following Ekman and Friesen (1978) for facial expressions, heart rate, and skin conductance peaks for peripheral physiology and alpha-band activity for an EEG measurement. Wearable devices can provide these measurements (Schmidt et al., 2019). Next, a classifier can categorize these extracted features into 'no frustration' and 'frustration.'

Based on labeled data, a frustration estimator can classify whether frustration is present. Previous work aiming to classify frustration used video (Grafsgaard et al., 2013; Hoque and Picard, 2011; K. Ihme, A. Unni, et al., 2018; K. Ihme, Dömeland, et al., 2018; Malta et al., 2010; McCuaig, Pearlstein, and Judd, 2010; Sidney et al., 2005), physiological (Belle et al., 2010), speech (Song, Mallol-Ragolta, et al., 2021) and neurophysiological (Fan et al., 2018; K. Ihme, A. Unni, et al., 2018) data, also in multimodal settings (Zepf, Hernandez, et al., 2020). Many of the studies that tried to find relevant indicators for frustration focused on frustration-typical facial expressions (Grafsgaard et al., 2013; Hoque and Picard, 2011; Kapoor, Burleson, and Picard, 2007), also in in-car settings (K. Ihme, A. Unni, et al., 2018; K. Ihme, Dömeland, et al., 2018). These expressions are most often described by the Facial Action Coding System (Ekman and Friesen, 1978), which describes facial muscle movements based on the activation of 27 different facial muscle groups. For example, Grafsgaard et al. (2013) found that Brow Lowerer, Brow Raiser and Dimpler correlate positively with frustration experienced during learning. Hoque and Picard (2011) found that participants often smiled when they were frustrated. K. Ihme, A. Unni, et al. (2018) frustrated participants in a driving simulator setting and describe that participants moved muscles in the mouth region (Chin Raiser, Lip Pucker, Lip Pressor) significantly more often in the frustrating than non-frustrating drives. In summary, muscle activations of Brow Lowerer, Dimpler, Brow Raiser, Smile, Chin Raiser, Lip Pucker and Lip Pressor are often shown in frustration. While not studied for frustration yet, bodily expressions can also be relevant indicators of emotion (Kleinsmith and Bianchi-Berthouze, 2012; Noroozi et al., 2018; Wallbott, 1998).

The basis for further improving such a frustration expression classifier is data that includes objective measurements obtained by sensors, and a subjective rating, which is used as 'ground truth' for subjectively experienced frustration. Song, Z. Yang, et al. (2019) published a video and audio dataset that recorded students while playing a frustrating game. They performed a binary classification into frustration and no frustration and achieved a classification accuracy of 60.3%. Similar to most previous work, their work is based on video data, as its advantage is its feasibility and non-invasiveness. Li et al. (2021) presented a dataset of drivers' anger, happiness, and neutral facial expressions with video stimuli. Ong et al. (2021) published a multimodal dataset that contained subjectively annotated unscripted life stories. To the best of our knowledge, no video-based dataset exists so far that

1. contains naturally occurring frustration of
2. a driver
3. that has continuous subjective frustration labels.

Therefore, the first contribution of this dissertation is such a dataset (see Figure 1.2, green box) with a thorough description of the data. It will be published so that developers of affect-aware systems, i.e. the affective computing community, can profit from it.

For example, facial expressions represent a low-cost and non-invasive method to assess driver frustration. To further improve in-vehicle frustration recognition, it is helpful to understand the occurrence of non-coincidental variance in frustration expression. Facial expressions are considered an individual trait that differs across the user's culture (Ekman, Friesen, et al., 1987) or gender (Chaplin, 2015). Wilms, Lanwehr, and Kastenmüller (2021) describe that an individual effect and a situation effect determine a construct's realization. In their research, they find that the individual effect and the situation effect account almost equally for variance found in emotion regulation. Cohn et al. (2002) conducted two studies that showed that it is possible to recognize individuals solely based on their facial expressions in response to emotional stimuli. This recognition worked in two different contexts and over long-time intervals (12 and 4 months, respectively). One context was that participants watched a film alone, and the other was a clinical interview. Gross (2008) describes that individuals differ in levels of emotion experience, behavioral responses, physiological responses, and subsequent emotion regulation. Differences on all these levels can lead to differences between individuals in expressing emotion. Barr, Kahn, and Schneider (2008) developed a taxonomy of individual differences in the expression of emotions and found two higher-order factors, emotional constraint, and emotional expression. Accordingly, . Sangineto et al. (2014) built a personalized classifier to account for individual differences in the expression of emotion. However, they take a black box approach, not considering the nature of differences between individuals. Also, the data contains videos in which the difference between individuals is not only the person but also the situation in which the person is, as situations differ per individuum. It is, therefore, impossible to differentiate between variance caused by individual or by situational differences. Kosch et al. (2020) investigated if detecting facial expressions via computer vision is feasible for mobile in-the-wild studies. They find that a re-calibration of the individual facial expressions on a per-user basis increases the correctness of emotions detected through facial expressions by 33%.

It is established that relevant individual differences in the expression of emotion exist, but the precise nature of differences has not yet been researched for frustration or in-vehicle. This information could relevantly improve in-vehicle recognition of frustration. Therefore, I argue that a descriptive approach that aims to study individual explanation of variance in the expression of frustration is essential for building generalizable FAAS. Here, I see the second contribution of this dissertation (see Figure 1.2, blue box).

Another possible source for assessing frustration is neurophysiological data (Fan et al., 2018; K. Ihme, A. Unni, et al., 2018; Reuderink, Mühl, and Poel, 2013). This data may be used in the future by unintrusive EEG systems that can, for example, be built-in into glasses. A frequently used method for inferring frustration from EEG is the calculation of the frontal Alpha Asymmetry Index (AAI) (Smith et al., 2017). The AAI is based on the difference in the activation measured at frontal electrodes, most commonly F3 and F4, thereby comparing activity over the left and right hemispheres of the brain (Smith et al., 2017). Numerous studies have used the AAI as an indicator for emotion-related state and trait measures, analyzing mood inductions, alterations, and dispositional mood (e.g. (Palmiero and Piccardi, 2017; Smith et al., 2017)). Regarding the recognition of emotion by EEG data, Reuderink, Mühl, and Poel (2013) and Schuster (2014) describe that relative left frontal alpha band activation (negative AAI values when AAI = F4 - F3/F3 + F4) characterizes low dominance. Huang et al. (2012) also showed that negative valence is characterized by relative left frontal alpha band activation (negative AAI values). According to these results, negative AAI values would characterize frustration as an emotion with low valence and low dominance. The AAI is, therefore, an additional modality that can give information on experienced frustration as measured by the body's physiological changes.

The component process model of emotion (Scherer, 2009) states that emotion consists of an event's appraisal and a subsequent appropriate expressive and physiological response. Accordingly, previous studies have measured subjective as well as expressive and physiological data to learn what patterns look like for various subjectively rated emotions. However, as both the subjective appraisal and the expressive and physiological response are part of the emotion, it is more accurate to estimate the actually experienced emotion in a model that uses both subjective, expressive, and physiological data as estimation variables. This estimation is possible using a latent variable model described in Ben-Akiva et al. (2002). Such latent variable models have previously been used to combine driving simulator and physiological data to model how stress changes car-following behavior (Paschalidis, Choudhury, and Hess, 2019) and how a driver's cognitive effort impacts route choice decisions (Agrawal and Peeta, 2021). A methodological upside that this latent variable model brings is that this 'complete' model of emotion (that includes subjective and objective data) enables a comparison between the complete full model and models that use fewer variables. Therefore, this method can evaluate which measured variables are needed to get an estimation close to the full model. This knowledge is crucial for a FAAS, as only a system that achieves a good classification without the continuous measurement of subjective ground truth data is feasible in an applied setting.

Such latent variable models have not previously been used to assess a situation's frustrativeness based on multimodal data. They offer a great way to estimate experienced frustration by subjective and objective measures. Therefore, I propose this as a suitable method for evaluating future measurements of frustration. The development of this method is the third contribution of this dissertation (see Figure 1.2, blue box).

#### 1.1.2.2 Data Sources and Estimation - Context

Next, the FAAS needs to recognize the cause of the traveler's frustration. For this, one needs background knowledge of the range of possible frustrators. The following step needs to recognize which of the possible frustrators is present. Regarding the recognition of emotions based on context data, Bethge, Kosch, et al. (2021) used contextual cell phone data like vehicle speed, weather, road types, and traffic flow to differentiate between 'anger','disgust','happiness', 'neutral,' and 'surprise' and can improve classification performance up to 38%. Also Malta et al. (2010) report that their frustration recognition algorithm improves when considering contextual data. Zepf, Dittrich, et al. (2019) present emotional triggers while driving and find traffic and driving task, vehicle and equipment, human-computer interface and navigation, and environment as the main types of emotional triggers. M. Jeon and Walker (2011) find 33 emotion-inducing contexts that can be categorized into driving-irrelevant in-vehicle contexts, driving relevant in-vehicle contexts, and driving irrelevant out-of-vehicle contexts. Liu et al. (2021) present a model that uses the vehicle's front-view camera and Controller Area Network (CANBUS) data. They can differentiate between six emotions with an accuracy of 71%. Bořil, Boyraz, and Hansen (2012) recognize stress by speech and CANBUS data, and Karaduman et al. (2013) differentiate aggression and calmness purely by CANBUS data. In summary, the first attempts to categorize several context data sources into emotion source categories have been made.

No studies have researched what common frustrators during driving are. This knowledge could form an essential indicator on which a FAAS could base successful frustration mitigation. A first step for recognizing causes of in-vehicle frustration is knowing which possible causes exist on the road. The answer to this is the fourth contribution of this dissertation (see Figure 1.2, red box).

#### 1.1.2.3 Mitigation Strategy and Intervention Execution

When the FAAS determines that 1) frustration is present, 2) what the causes of frustration is, and 3) what the suitable mitigation strategy is, the FAAS can prompt a mitigation strategy that can, for example, by voice, interact with the traveler. This interaction leads to the mitigation of frustration and therefore changes the traveler's state. This mitigation is realized by initiating intervention methods, like the presentation of information, via visual or auditory human-machine interfaces. Previous studies have found these intervention methods to be effective in increasing user acceptance (Grippenkoven et al., 2018; Herrenkind et al., 2019; Keller et al., 2019; Millonig and Fröhlich, 2018). For example, Löcken, Ihme, and Unni (2017) collected several ideas for mitigation strategies for frustration and proposed an ambient light that changes from white to blue depending on the frustration level. Nass et al. (2005) frustrated participants and subsequently helped reappraise the situation with a voice assistant. Positive reappraisal was successful and mitigated frustration and improved driving performance. A follow-up study found that matching the voice assistant's tone to the driver's state increased driving performance (Harris and Nass, 2011). Braun, Schubert, et al. (2019) compared 'Ambient Light,' 'Visual Notification,' 'Voice Assistant,' and 'Empathic Assistant' in a simulator study with 60 participants, inducing anger and sadness in two different groups. The intervention method 'ambient light' consisted of a LED strip that changed its color to purple-blue light in the anger group and green-yellow in the sadness group. 'Visual notification' showed an angry face when anger was detected and a sad face when sadness was detected. 'Voice Assistant' was a voice sample that said, 'I detect that you are distracted. Would you like to listen to some radio to concentrate better?'. The Empathic Assistant was similar to the Voice Assistant but empathized with the participant by saying, 'Hey, are you alright? I can understand that you are a bit angry, sometimes I feel the same way. How about some music to take your mind off things?'. They found that the empathic voice assistant was most effective in reducing negative emotions. Zepf, Hernandez, et al. (2020) detected drivers' frustration through facial expressions and heart rate and mitigated frustration with ambient light and a voice assistant. Research on mitigation methods is still ongoing, and the path to a natural and universally working frustration mitigation is still a long way off. However, mitigation is only possible if frustration was accurately recognized in a first step. Therefore, this dissertation focuses on contributing to the successful recognition of frustration and finding common causes of in-vehicle frustration, and does not contribute to investigating different mitigation methods.

### 1.1.3 Structure of this dissertation

This dissertation presents four original research papers that answer two main research questions. These research questions are:

1. How can we improve the classification of frustration?
2. What contexts lead to in-vehicle frustration?

Accordingly, the dissertation is structured as follows: After this introduction, Chapter 2 entails the work done on the research question (1) as published by papers 1-3. Paper 1 presents a dataset of 43 participants who experienced frustration in driving-related situations in a simulator (Bosch, Corbí, et al., 2022). The data set contains a continuous subjective ground truth label, hand-annotated face and body expressions, facial landmark coordinates, and demographic data. In addition, a descriptive analysis and description of the data's characteristics are provided with a Graph Convolution Network-based model to recognize frustration. This work is valuable for researchers of the affective computing community because it provides realistic and natural data with an in-depth description of its characteristics and a benchmark model for automated frustration recognition. This is part of the data used for the analyses conducted in papers 2 and 3. Paper 2 examined the possibility of improving the recognition of frustration by considering individual differences (Bosch, Kaethner, et al., n.d.). For this, I used the driving simulator study data from study 1, and, additionally, the real car study data from study 2. An analysis of participants' facial expressions during frustrating driving situations confirms previously reported expressions of frustration (Brow Lowerer, Dimpler, Brow Raiser, Smile, and Lip Press). In addition, the results also hint toward high variance between and low variance within participants for all other expressions, suggesting the existence of individual-typical expressions of frustration. Hence, future frustration-aware systems could benefit from considering these individual differences by using a universally trained algorithm that is customized to each individual. Paper 3 presents a latent variable model that estimates frustration in two different contexts by continuous subjective frustration rating, facial expressions, and frontal alpha asymmetry in the EEG (Bosch, Klosterkamp, et al., 2022). We then compare this full model to models with fewer measurement variables to evaluate which ones can be left out. Our results show that expression frequency and subjective frustration contribute to the experienced frustration model. This paper presents a proof of concept for using a latent variable model to evaluate collected measures to estimate an experienced emotion. This method can inform researchers which measurements are most informative in different circumstances. Additionally, the method shows how well purely objective measurements (the only feasible measurements in most applied settings) perform compared to a model including subjective ratings. The third chapter presents the paper concerning research question 2. Paper 4 employs a combination of diary study and user focus groups to shed light on the causes of why humans become frustrated during driving (Bosch, K. Ihme, et al., 2020). In addition, we asked the participants of the focus groups for their usual coping methods with frustrating situations. We revealed that the main causes of driving frustration are traffic, in-car reasons, self-inflicted causes, and weather. Coping strategies drivers use in everyday life include cursing, distraction by media, and thinking about something else. This knowledge will help design a frustration-aware system that monitors the driver's environment according to the spectrum of frustration causes in the research presented here. Finally, I will discuss the meaning these four contributions have for developing FAAS in specific and transportation research in general.

# 2

# Research Question 1: How can we improve the classification of frustration?

Frustration Recognition Using Spatio Temporal Data: A Novel Dataset and GCN Model to Recognize In-Vehicle Frustration.

This is a postprint version of the manuscript published as:

Bosch, E., Corbí, R. L. H., Ihme, K., Hörmann, S., Jipp, M., & Käthner, D. (2022). Frustration Recognition Using Spatio Temporal Data: A Novel Dataset and GCN Model to Recognize In-Vehicle Frustration. IEEE Transactions on Affective Computing. DOI: https://doi.org/10.1109/TAFFC.2022.3229263.

# **F**rustration **R**ecognition **U**sing **S**patio **T**emporal Data: A Novel Dataset and GCN Model to Recognize In-Vehicle Frustration

Esther Bosch*, Raquel Le Houcq Corbí*, Klas Ihme, Stefan Hörmann, Meike Jipp, and David Käthner

**Abstract**—Frustration is an unpleasant emotion that is highly prevalent in several target applications of affective computing, such as human-machine interaction, learning, (online) customer interaction and gaming. To redeem this issue, one idea is to recognize frustration in order to offer help or mitigation in real-time, e.g. by a personal assistant. The recognition of frustration is not limited to these applied contexts, but can also inform emotion research in general. In this paper, we present a dataset of 43 participants who experienced frustration in driving-related situations in a simulator. The data set contains a continuous subjective ground truth label, hand annotated face and body expressions, facial landmark coordinates of both the frontal camera and the camera that was placed above the tablet, and the participants' age and sex information. A descriptive analysis and description of the data's characteristics are provided together with a Graph Convolution Network based model to recognize frustration. Allowing for a tolerance of 10%, the model could correctly identify frustration with a similarity of 79.4 % and a variance of 7.7 %. This work is valuable for researchers of the affective computing community because it provides realistic and natural data with an in-depth description of its characteristics as well as a benchmark model for automated frustration recognition. Our FRUST-dataset is publicly available under: tbd.

**Index Terms**—Frustration Recognition, Naturalistic Dataset, Graph Convolution Network, Affect-Aware Systems.

## 1 Introduction

Frustration is an emotion that occurs when goal achievement is blocked [1] in a variety of contexts. The circumplex model of affect characterizes frustration to be of negative valence, high arousal and low dominance [2] by Russell [3]. Frustration can lead to diminished motivation to learn among students [4], dissatisfied customers in interaction with human or artificial customer service personnel [5], diminished player engagement in gaming [6], decreased performance and mental health of employees [7], and a decreased acceptance of (new) technical systems due to failed user interaction [8], [9]. To counteract negative effects of frustration, affect-aware systems that recognize frustration in real-time and react accordingly have been proposed. A growing body of literature explores how to recognize and mitigate frustration in various contexts [10].

Traffic is one context in which frustration occurs. Drivers experience frustration frequently [11] and supposedly even in automated driving modes [12]. In manual driving, frustration plays a critical role regarding road safety. Aggressive driving maneuvers [13], [14] and distracted driving [15] often result from the experience of frustration. Moreover, unpleasant emotions like frustration diminish cognitive skills important for safe driving [15]. As the driving task shifts from human to machine, the effects of frustration change from safety- to acceptance-critical. Both during development and deployment of automated driving systems, uncertainty and subsequent frustration of users caused by system glitches and sub-optimal technical functionality are to be expected. This could impede the acceptance of automated vehicles, which are one building block for the change towards modern, innovative and sustainable mobility solutions. Here, an affect-aware system that offers solutions, gives information or initiates dialogues could alleviate or even prevent the occurrence of frustration [16], [17] and thus aid the formation of acceptance towards automated systems.

The first required step for building frustration-aware systems is a robust frustration recognition. Previous work that aimed to recognize frustration has used video [18], [19], physiological [20], speech [21] and neurophysiological [2], [10] data, also in combination [17]. For successful recognition, it is crucial to train algorithms on high-quality, realistic data which are barely available as of now. Song et al. [22] published a video- and audio-dataset that recorded students while playing a frustrating game. They performed a binary classification into frustration and no frustration and achieved a classification accuracy of 60.3%. Like most previous studies, their work is based on video data because of its advantage regarding feasibility and non-invasiveness. Macary et al. [23] published a dataset of continuously frustration-annotated call-center conversations. Li et al. [24] present a dataset of driver's facial expressions that induced anger and happiness and is annotated after every drive (see their introduction for an overview of more emotion expression datasets). The work of Li et al. [25] is similar,

- *E. Bosch, R. Le Houcq Corbí, K. Ihme and D. Käthner are with the Institute of Transportation Systems, German Aerospace Centre, Braunschweig, Germany.*
  *E-mail: esther.bosch@dlr.de*
- *R. Le Houcq Corbí and S. Hörmann are with the Chair of Human-Machine-Communication, Department of Electrical and Computer Engineering, Technical University of Munich, Munich, Germany.*
- *M. Jipp is with the Institute of Transport Research, German Aerospace Centre, Berlin, Germany.*

**These two authors contribute equally to the work.*

TABLE 1
Comparison to other publicly available datasets

| Dataset | Emotions | Stimulus | Input Data | Labels |
|---|---|---|---|---|
| MGFD [22] | Frustration | Video game-play | Audio& Video | Static & discrete frustration levels |
| AlloSat [23] | Frustration | Call center recordings | Audio | Static valence score |
| DEFE [24] | Neutral, happiness, anger, valence, arousal, dominance | Driving Scenarios | Video | Static & discrete emotion model Dimensional valence arousal model |
| PPB-Emo [25] | Anger, sadness, fear, disgust, neutral, surprise, happiness | Driving scenarios | Video | Static & discrete emotion model Dimensional valence arousal model |
| FRUST (Ours) | Frustration | Driving scenarios | Video | Continuous frustration rating |

but induced all six basic emotions. However, to the best of our knowledge, no public dataset exists so far that fulfill all of the following criteria: it contains 1) manually annotated expression and automatically annotated landmark data of 2) naturally occurring frustration and 3) has continuous subjective frustration labels 4) recorded in a driving setting (see Table 1). In summary, the main contributions of the following work are:

- The introduction of a novel and continuously labeled dataset containing frustrating driving scenarios, called **F**rustration **R**ecognition **U**sing **S**patio **T**emporal Data (FRUST).
- A thorough description of our FRUST dataset.
- A discussion of challenges that need to be considered when working with the specific dataset and subjectively labelled emotion data in general.
- A state-of-the-art Graph Convolutional Network (GCN) to automatically recognize frustration levels as benchmark.

Emotion recognition is a complex problem. Therefore, it is necessary to address it from different perspectives. Previous work on emotion recognition has attempted to solve this challenge on the basis of hand-annotated expressions and classical statistical analysis in the earlier days and evolved towards automated recognition of expressions and classification by machine learning in recent years. Both approaches have respective advantages by providing different insights into the underlying data; while classical statistical analysis gives insight into which expressions are important, machine learning approaches are more feasible in an automated setting. For both approaches, it is crucial to work with data that was acquired in a credible setting and in which a continuous subjective rating is acquired. We provide a high-quality dataset that is feasible for both approaches and thereby opens up a possibility to investigate their relationship. We include a descriptive statistical analysis of hand-annotated expressions and a machine learning model based on automated recognized facial landmarks in this work. Our dataset includes the following variables: 1) a continuous subjective frustration rating; 2) a questionnaire rating of 24 emotions after each drive; 3) hand-annotated expressions; and 4) automatically recognized facial landmarks.

# 2 FRUST DATASET

The collection of this dataset is also described in [26].

Fig. 1. Experimental Setup-Up.

## 2.1 Data Collection

### 2.1.1 Participants

Fifty participants were recruited through the institute's participant pool. Of these, 7 participants were excluded from data analyses due to motion sickness (2 cases), faulty logging (1 case), a condition of facial myoclonus (1 case) and partly missing frustration rating data (3 cases). Of the N = 43 participants included in the analyses 13 were female and 30 male. Participants age ranged from 20 to 59 years (M = 31.8, SD = 12.2). Participants were informed about all data recordings, potential risks of driving in simulators (e.g., the experience of simulator sickness) and the duration of the experiment. Participants could take a break or abort their participation at any time. All participants gave written informed consent to take part in the study. As reimbursement for their time, the participants received 5 € per commenced half hour. After the study, the true goal of the experiment (evoking frustration) was revealed and the necessity to conceal this goal with a cover story was explained. The collected data were processed according to European General Data Protection Regulations.

### 2.1.2 Experimental Set-Up

The data set was recorded in a 360-degree full-view driving simulator [27]. The participants sat in a vehicle mock-up and could use a conventional interface with throttle, brake pedal, steering wheel, and indicators to drive the mock-up car in the driving simulation (Virtual Test Drive, Vires Simulationstechnologie, Bad Aibling, Germany). On a tablet (Microsoft Surface Pro 7, 12.3') mounted to the car's center console, a user interface (UI) was shown (required for the frustration induction, see subsubsection 2.1.3). The setup is shown in Figure 1. During all drives, one frontal camera and one mounted above the tablet filmed the participant's face and body. In the automated condition, the latter was employed to record the face of a participant who was oriented towards the UI. Another camera facing forward, mounted between the driver and co-driver seats, captured the entire scene. All three cameras were Axis M1065-L network cameras that recorded at a resolution of 1280 x 800 pixels.

### 2.1.3 Stimuli

We collected data in two different driving modes (manual vs. automated). In the manual driving mode, participants were told to assume they were supposed to meet friends at a movie theater. The driving track consisted of a rural road in the first half and an urban road in the second half. The participants were informed that the average travel time to the movie theater was less than ten minutes. They were told that if they arrived at their location on time, they would receive a 2 € prize. The time remaining for punctual arrival at the movie theater was displayed on a clearly visible clock. In the baseline condition ('Baseline Manual'), participants drove five minutes and the setting featured low traffic density ensuring that participants reached the movie theater in the allotted time. In the frustration conditions ('Manual Frustration-inducing condition 1' and 'Manual Frustration-inducing condition 2'), a phone call simulated a group of friends already waiting at the movie theater, reminding the participant that they had the entry tickets for everyone. Slow lead vehicles and red lights obstructed the way during the following drive. Throughout the journey, the clearly visible clock displayed a time that was 30 seconds shorter than the time it took to drive the route. Between the two frustration conditions, we varied the car types and the track that was driven to disguise the fact that the same driving scenario was driven twice, with one track taking seven minutes and the other ten. Frustrating incidents were the same in both driving scenarios and took the same amount of time. In earlier investigations, tasks comparable to the ones used in this study have been demonstrated to successfully elicit frustration [10], [18].

In the automated driving mode, participants completed a task (joining an online conference or changing the destination) on the in-car UI shown on the tablet. Meanwhile, the car drove fully automated on a highway. The participants were told that if they completed their work successfully, they would get a reward of 2 €. Before the start of the experiment, all participants read the same story in all three automated drives. They were asked to imagine to drive to a business meeting in an autonomous car. In each of the modes, participants experienced one baseline drive and two frustration-induction drives. In the baseline condition ('Baseline Automated') the participants were requested to browse a web page in the baseline condition, which was a simple task. They were then instructed to push a single button that appeared in various locations of the UI. They were assured not to be under any time constraints and asked to interact with the UI as relaxed as possible. The drive took three minutes. In the frustration condition ('Automated Frustration-inducing condition 1' and 'Automated Frustration-inducing condition 2'), the participants received a scripted call from their 'boss' in the frustration condition, telling them that they were urgently required for another, more important meeting and that they needed to turn around quickly to be on time. Following that, the participants had to adjust the navigation system's destination. This was difficult to do in seven minutes due to vague button names, imprecise iconography, and confusing click-paths. In the second automation scenario, a 'boss' called and urged the participant to join an online meeting with clients as soon as possible. Again, the UI was so difficult to comprehend that it was challenging to achieve the purpose of participating in the online conference. The drive took seven minutes.

#### 2.1.4 Measures - Subjective Frustration Rating as Manipulation Check

In order to acquire a time-resolved assessment of frustration, a continuous subjective assessment was collected when a participant had completed all drives. The participants assessed their frustration using a joystick on a scale of 0 to 100 percent while watching the videos that were recorded during all drives of the whole scene (the participant's face was not visible). When not touched, the joystick could only go in one direction and immediately returned to zero. The participants received a visual feedback of their current rating next to the video. They were instructed to move the joystick according to their level of frustration in the circumstances depicted in the video. This allowed for the collection of a continuous frustration rating for each drive and each participant.

After each drive, an emotion questionnaire was completed. First, participants completed a questionnaire that included four distraction questions on gaze behavior that were in accordance with the cover story. Afterwards, the participants rated to what extend certain emotional words described their current subjective emotional state on a 5-point scale from 'not at all' to 'extremely'. The questionnaire was based on the Positive and Negative Affect Scale (PANAS [28]). From the original PANAS items, the following emotion words were used: active, afraid, alert, ashamed, attentive, determined, distressed, enthusiastic, excited, inspired, interested, jittery, nervous, proud, scared and upset. Additionally, the following emotion words were used: angry, frustrated, insecure, relaxed, sad and surprised. The emotions besides frustration were queried to record emotions that were possibly co-triggered, unintentionally, by the experiment. The dimensional scales of valence, arousal and dominance were also assessed on a 5-point scale from from negative to positive, excited to calm and influenced to independent, respectively.

#### 2.1.5 Measures - Data Annotation

We include an annotation of human observers as well as automated recognition of participants' facial and bodily expressions in our FRUST dataset. On the one hand, it has been found that humans outperform automated detection of affective expressions [29]. On the other hand, only an automated recognition of expressions is feasible in applied settings. We therefore used the commercially available software Affectiva [30] for detection of facial landmarks and action units. Figure 2 shows the landmarks that are labeled by Affectiva. Kulke et al. [31] found that Affectiva's results are comparable to Electromyographic (EMG) results, a common method to recognize facial muscle activity. In comparison to EMG, recognition by Affectiva is less intrusive and therefore more feasible in applied contexts like a vehicle. All data was sampled with a frequency of 20 Hz. Not all frames were successfully detected by Affectiva. The detection rate varies between videos. However, we included all videos in our dataset for reasons of completeness.

The term 'expression' is used throughout the text to refer to facial motions that involve numerous face components (e.g., an expression of joy). The action units specified in

Fig. 2. Landmarks as annotated by Affectiva [30].

TABLE 2
Expressive units that were annotated by hand.

| | |
|---|---|
| **Lower Face** | Blow Cheeks, Cheek Puffer, Chin Raiser, Dimpler, assymetric Dimpler, Jaw Drop, Laugh, Lip Corner Depressor, Lip Funneler, Lip Press, Lip Puckerer, Lip Stretcher, Pout, assymetric Pout, Smile, Smirk, Tongue Out, Upper Lip Raiser, Lip Corner Puller, assymetric LipCornerPuller, LipTightener, assymetric LipTightener, Lower Lip Depressor, Mouth Stretch, Nose Wrinkler |
| **Upper face** | Brow Lowerer, Brow Raiser, Roll Eyes, Squint, Upper Lid Raiser, Eyes Closed |
| **Head** | Head Back, Head Shake, Head Tilt, Head Wiggle, Chin Back, Head Forward, Swallow Hard, Move Jaw |
| **Body** | Deep Breath, Hands To Air, Hands To Face, Shrug, Straighten Up |

the facial action coding system [32] were expanded to incorporate head, torso, and hand motions, as they have been proven to be important emotional expressors as well [33], [34]. We employ the term *expressive unit* for these annotated facial and bodily expressions. Expressive units were manually annotated over the whole length of all six drives for all drives and participants by two independent raters. A third rater then decided on all occasions when the rater 1 and rater 2 disagreed. Table 2 lists the categories that were annotated. All annotated terms (see Table 2) refer to expressive units. The annotation was carried out with the help of the program ELAN [35]. Analysis of the hand-annotated expressive units was done by means of expression frequency as measured by how often an expressive unit was shown within one minute. The automated annotation of 34 facial landmarks was done with the Affectiva module from iMotions [30]. The automated annotation of 25 body landmarks was done with OpenPose [36].

#### 2.1.6 *Procedure*

On arrival, participants filled out an informed consent form as well as a data privacy declaration. The researcher told the cover story that the study was analyzing changes in gaze behavior between manual and automated driving modes. This was done to hide the true purpose of frustration induction and allow spontaneous emotion emergence. To lessen the impact of unfamiliarity, all participants practiced manual and automated driving modes before the start of the experiment until they were comfortable with the simulator and the driving conditions. Following the six drives, the participants were told of the study's true purpose. Then, they provided a continuous post-hoc frustration rating for

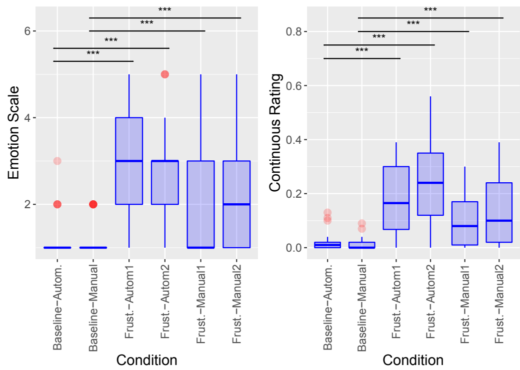


Fig. 3. Subjective frustration ratings. Left: emotion scale rating; right: continuous post-hoc rating.

each of the six drives. The entire process took 2 hours on average.

#### 2.1.7 *Experimental Design*

Each participant experienced six drives in a 2 (driving mode: 3 automated vs. 3 manual drives) x 2 (frustration induction: 2 frustration vs. 1 baseline drive) within-subject design in a driving simulator. Three of the drives were driven in manual driving mode and three in automated driving mode. Both driving modes had one baseline drive and two frustration-inducing drives each. The drives' order for each participant was determined by a balanced Latin square design.

### 2.2 Descriptive Statistics

Frustration induction was successful as indicated by both subjective ratings. Figure 3 shows the continuous frustration rating (left) and the emotion-scale rating of frustration after each drive (right), both per condition. The Spearman correlation between the continuous frustration rating and the emotion-scale rating is 0.59. This is a high correlation according to Cohen [37]. A Friedman's ANOVA revealed that both frustration ratings – emotion scale rating and continuous rating – varied significantly between conditions, $\chi^2(5) = 122.07, p < .001$ (emotion scale rating) and $\chi^2(5) = 116.63, p < .001$ (continuous rating, see Figure 4). Holm-corrected post hoc tests revealed that per driving mode, both subjective ratings were higher in the frustrating drives than in the baseline drives, but not different within baseline and frustration drives. In summary, we assume that frustration was induced successfully in both, the automated and manual driving modes.

The two raters agreed on annotations in 55% of all cases and gave a similar rating in 10% of the cases (e.g. Lip suck and Lip press). In 35% of the cases only one of the two raters provided an annotation. Therefore, in the 45% that the two annotations were different, the third rater decided for one of the two annotations. In total, 4583 instances of facial or bodily expressive units were annotated. Figure 5 shows an exemplary drive with the continuous frustration rating (top), annotated expressive unit (middle) and the Affectiva-value of Action Unit four (Brow Lowerer, bottom). Figure 6 shows the frequencies with which expressive units occurred in different drives over all participants. It is

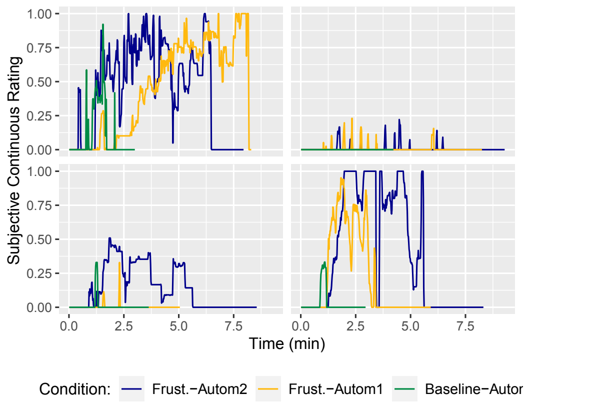


Fig. 4. Subjective frustration ratings of four exemplary participants' automated drives.

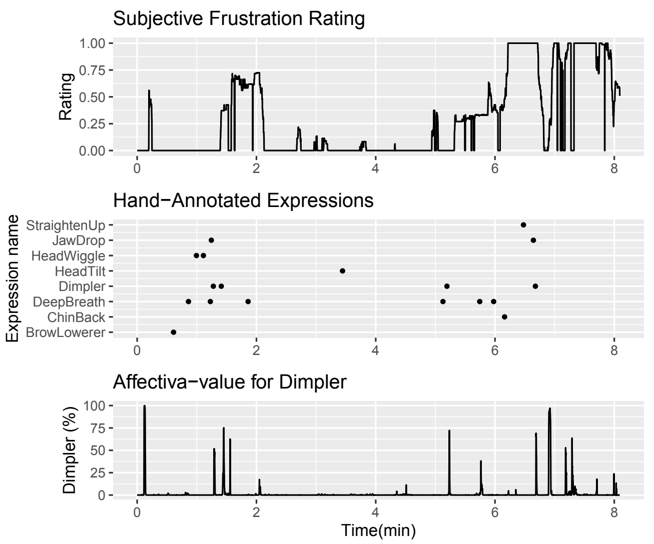


Fig. 5. Frustration rating, hand-annotated expressions and the automated detected value for action unit four for one drive.

visible that the hand-annotated expressions of Brow Lowerer, Brow Raiser, Deep Breath, Dimpler, Hands to Face, Jaw Drop, Lip Press, Smile and Tongue Out occur most often. These include previously described frustration-typical facial expressions [18], [19], [38]. There was a small, but significant correlation between expression frequency and frustration ratings ($\rho = 0.004, p > .05$). The same was true when only including previously described [18], [19], [38] expressions of frustration (Brow Lowerer, Brow Raiser, Dimpler, Lip Press and Smile) ($\rho = 0.08, p < .05$). Expression frequency was significantly higher in frustration-inducing than in baseline drives with a small effect size ($(W_{\text{Mann-Whitney}}) = 373000, p < .001, \hat{r}^{\text{rank}}_{\text{biserial}} = 0.23$), but this effect did not persist when only including previously described expressions of frustration ($(W_{\text{Mann-Whitney}}) = 36371.0, p > .05, \hat{r}^{\text{rank}}_{\text{biserial}} = 0.07$). Expression frequency was significantly different between participants with close to no effect ($\chi^2_{\text{Kruskal Wallis}}(40) = 119.61, p < .001, \hat{\epsilon}^2_{\text{ordinal}} = 0.06$), and was significantly different in pairwise tests only in 6 out of 780 possible pairwise tests ($p_{\text{Holm-corrected}} < .05$).

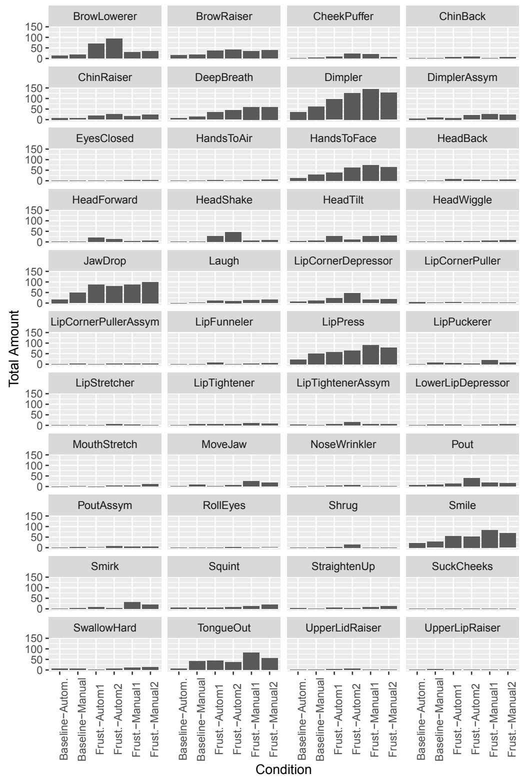


Fig. 6. Expression frequency per expression, split by drive.

# 3 Frustration Recognition Model

In this section, we present our Graph Convolutional Network (GCN) model to detect frustration levels using the introduced FRUST dataset. The model was trained with a graph representation of the face using the extracted landmarks.

## 3.1 Landmark-based Emotion Recognition

Landmark-based emotion recognition methods traditionally use deep neural networks or SVM to directly process the landmark position [39], [40]. Kim et al. [41] combined a Convolutional Neural Network with a LSTM layer for an emotion classification. The rise in popularity of Graph Neural Networks (GNNs) in image classification [42] and action recognition [43] has also inspired the use of GNNs for affect detection.

The advantage of GNN models is that they can depict hidden patterns from non-euclidean data structures. A skeleton-based emotion recognition GCN was developed by [44], which classifies emotions from the walking gait of a person. Rao et al. recognized facial expressions with GNNs, using facial landmarks as input to eliminate redundant information from the dataset [45]. In comparison to Rao et al. this work also considers temporal dependencies, which is highly relevant in dynamic processes like emotion emergence [46].

### 3.1.1 *Graph-Based Models for Spatio-Temporal Tasks*

The beginning stages of GNNs are marked by the Recurrent Graph Neural Networks, which were first applied by

Sperduti et al. in 1997 [47]. Here, information is repeatedly passed between nodes for the network to learn node representations. However, these models are computationally very expensive. Simultaneously, Convolutional Neural Networks (CNNs) were developing very fast and with it also the interest to simulate its methods to graph based models. This brought two streams of Graph Convolutional Networks: spectral-based and spatial-based approaches. The spectral-based approaches are based on graph signal processing theories and were first presented in 2013 by Bruna et al. [48]. The more efficient spatial-based GCNs were first studied in 2009 by Micheli et al. [49]. These convolutions rely on the graph's topology. Although unnoticed until recently, they are one of the main methods used to learn from graph-structured data. GCNs are very popular today and several frameworks have developed from them.

One example are Spatio-Temporal Graph Convolutional Networks (ST-GCNs), which were first presented by Yan et al. [43]. ST-GCNs aim to compute the dynamic structures and inputs of graphs by performing convolutions over neighboring nodes for spatial and temporal dependencies [43]. They have proven to be effective for a wide range of tasks where spatial and temporal dependencies play an important role, such as action recognition [43] or traffic forecasting [50]. One limitation of ST-GCNs is the high computational cost caused by the high number of parameters processed. Also, they are susceptible to vanishing gradients, therefore limiting the amount of stacked layers that can be used. Song et al. [51] addressed this problem by including residual links between layers aiming to reduce the difficulty of model training and decreasing computational cost. By this, the ResGCN architecture was introduced [51]. The architecture already showed promising results when recognizing walking patterns from skeleton data [52]. Deeper GCNs were then enabled by Li et al. [53], who transferred the concepts of residual connections and dilated convolutions to GCNs.

Song et al. [54] first introduced a multi-stream GCN to solve the problem of noisy skeletons and occlusion. This was later enhanced to reduce computational costs with a three-branch architecture [51]. The architecture depicted joints, velocity and bone features. These features were processed first in separate input branches and then merged into a main stream to reduce the computational cost. The promising results shown by the ResGCN architecture combined with the reduced computational cost inspired the use of a multi-branch ResGCN architecture in the following work.

#### 3.1.2 *Graph Convolutional Network*

The most elementary building block of GCNs are the graph representations of the data. Graphs are a set of nodes $\boldsymbol{V}$ connected with the respective edges $\boldsymbol{E}$ and can be described as $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$. In a GCN model, each node contains a set of features $\mathbf{X}$ conveying information. This information is exchanged between nodes using the message passing algorithm during training until a stable equilibrium is reached, which can be mathematically described as follows:

$$\mathbf{X}_t^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}_t^{(l)}\boldsymbol{\Theta}^{(l)}\right), \tag{1}$$

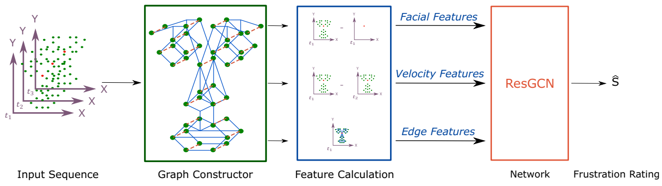


Fig. 7. Overview of the networks pipeline.

where $\sigma(\cdot)$ is an activation function and $\tilde{\boldsymbol{D}}$ is the diagonal degree matrix of $\tilde{\boldsymbol{A}}$. $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ denotes the graph structure with added self-connections. The advantage of employing facial landmarks to obtain a graph representation of the face is that the information is concentrated in the important parts of the face contributing to an expression, such as eyes, eye-brows or lips, while eliminating redundant information from the background. The advancements in skeleton-based recognition with the use of GNNs [55] have also shown that landmark representations are more robust against background and illumination changes.

### 3.2 Network Overview

We built a GCN that learns facial features from a set of landmarks. Our network is composed of three main modules depicted in Figure 7. The first module builds a facial graph using the x and y coordinates from the dataset. Similar to [51], our model uses three feature vectors providing information about the facial landmarks, velocity and edges. Thereupon, our next module calculates the necessary features for further analysis. Sequences of the extracted information are then fed to the network using a sliding window approach with adaptable window and stride size. The network is a combination of *ResGCN* layers adapted from [51]. The output of the network is a frustration rating for each sequence of data.

### 3.3 Pre-Processing

The 20Hz video recordings were first processed with the Affectiva module from iMotions [30] to extract 34 facial landmarks for each frame. The facial landmarks $\boldsymbol{w}_i$ are the x and y-coordinates of the key-points in the face. The key-points are then represented as the nodes $\boldsymbol{V} = \{\boldsymbol{v}_{ti} \mid t = 1, 2, \ldots, T, i = 1, 2, \ldots, N\}$ with $T$ being the number of frames and $N$ the number of nodes. Note that $\boldsymbol{v}$ is the feature of the node and does not exactly correspond to $\boldsymbol{w}$ as we consider multiple features based on $\boldsymbol{w}$.

Our model uses three input graphs with three different node features $\boldsymbol{X}(\boldsymbol{v}_{ti})$ containing the following groups of features: 1) node position, 2) motion velocity and 3) edge features. Each group of features is then processed in a separate network branch as depicted in Figure 10. All processing steps to obtain the former group of features were adapted to our problem from [51].

Following the notation of Song et al., the first set of features are formed by the original 2D-coordinates $\boldsymbol{w}_i = (x_i, y_i)$ and the relative coordinates $\boldsymbol{r}_i$. If the original coordinate feature set of a sequence is denoted as $\mathcal{D} = \{\boldsymbol{d} \in \mathbb{R}^{C \times T \times N}\}$, with $C$, $T$, $N$, being the number of coordinates, frames and nodes. Since we consider 2D-coordinates, $C = 2$. The relative coordinate feature set $\mathcal{R} = \{\boldsymbol{r}_i \mid i = 1, 2, \ldots, N\}$ is calculated as $\boldsymbol{r}_i = \boldsymbol{d}[:, :, i] - \boldsymbol{d}[:, :, c]$, $c$ represents the most central node, which in our case is the nasal bone (c.f. Figure 8). These two sets are concatenated to a single sequence, obtaining a 4D feature vector that serves as an input to the first branch. The goal of including relative coordinates as input features is to consider changes in head position independent of participants.

Apart form the landmark location, the movement of the landmarks over time also carries important information. Consequently, the second group of features examines the movement velocity between frames. For this, we consider the relative change in landmark position in the consecutive frame $\mathcal{F}$ and two consecutive frames $\mathcal{M}$, concatenating these two sets to obtain a 4D feature vector for each node. The two sets of motion velocities $\mathcal{F} = \{\boldsymbol{f}_t \mid t = 1, 2, \ldots, T-2\}$ and $\mathcal{M} = \{\boldsymbol{m}_t \mid t = 1, 2, \ldots, T-2\}$ were calculated the following way: $\boldsymbol{f}_t = \boldsymbol{d}[:, t+1, :] - \boldsymbol{d}[:, t, :]$ and $\boldsymbol{m}_t = \boldsymbol{d}[:, t+2, :] - \boldsymbol{d}[:, t, :]$.

Lastly, to get a complete representation of the facial structure the edge information encoding the lengths $\mathcal{L}$ and angles $\mathcal{A}$ of the edges is calculated. The length set $\mathcal{L} = \{\boldsymbol{l}_i \mid i = 1, 2, \ldots, N\}$ is determined by $\boldsymbol{l}_i = \boldsymbol{d}[:, :, i] - \boldsymbol{d}[:, :, i_{\text{adj}}]$, where $i_{\text{adj}}$ denotes the adjacent nodes. The adjacent nodes were assigned so that every edge was taken into account except the edges connecting the eyebrows and the eyes, the nasal bone and the side of the nose, and the two lip corners. The angle set $\mathcal{A} = \{\boldsymbol{a}_i \mid i = 1, 2, \ldots, N\}$ between node and adjacent node was calculated with the previously calculated length for the 2D coordinates:

$$\boldsymbol{a}_{iw} = \arccos\left(\frac{\boldsymbol{l}_{i,w}}{\sqrt{\boldsymbol{l}_{i,x}^2 + \boldsymbol{l}_{i,y}^2}}\right) \tag{2}$$

In our method, we used 27 nodes from the 34 extracted facial landmarks. Five landmarks describing face contour and two landmarks separating the lips were discarded due to their low reliability. However, the number of nodes can be easily changed in our network. In order to account for spatial and temporal dependencies in our graph $\boldsymbol{G}$ similar to [43] two subsets of edges $\boldsymbol{E}$ were used. The edges first follow the spatial configuration shown in Figure 8, where each node connects to the adjacent nodes $\boldsymbol{\mathcal{E}}_{\text{F}} = \{\boldsymbol{v}_{ti}\boldsymbol{v}_{ti_{adj}}\}$. Additionally, to model the temporal dependency, the nodes were connected to the same node of the consecutive frame $\boldsymbol{\mathcal{E}}_{\text{T}} = \{\boldsymbol{v}_{ti}\boldsymbol{v}_{(t+1)i}\}$.

Each frame was labeled with a frustration rating. In order to obtain a smooth rating without abrupt changes we applied a Butterworth low-pass filter with a cut-off frequency of 0.005. An excerpt of the results obtained after filtering is illustrated in Figure 9. Additionally, all node features $\boldsymbol{X}$ were standardized to range between -1 and 1. This allowed an equal weighting of their representation enhancing the learning process of the network.

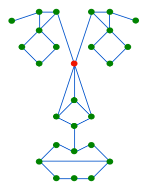

Fig. 8. Facial graph construction. Nodes are represented in green, while the central node is shown in red and edges connecting the nodes are illustrated in blue.

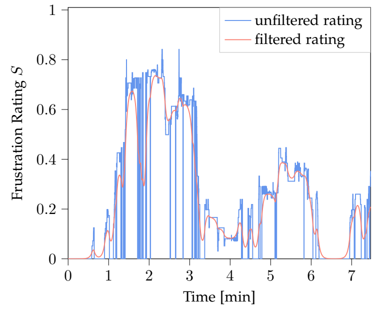


Fig. 9. Frustration rating $S$ before and after applying the low-pass filter.

## 3.4 Architecture

### 3.4.1 *Fast&Slow Model*

Each group of features is processed in a separate feature branch to reduce computational costs. The outputs of the branches are then concatenated and processed in a main model stream (Figure 10) following the general outline of the *ResGCN* network proposed by [51] with the adequate changes to fit our problem. Each branch starts with a batch normalization of the input data followed by one basic ResGCN module and two bottleneck modules. By applying a batch normalization at the beginning of the layer the input is standardized, which stabilizes and accelerates the learning process [56]. The main stream is composed of four bottleneck layers and one fully connected layer. The network is concluded with one Sigmoid layer to obtain a scalar $\hat{S} \in (0, 1)$, which describes the probability of frustration. The pipeline can be seen in Figure 10. The strength of ResGCNs modules is that the combination of spatio-temporal blocks [50] and bottleneck structures decreases the difficulty of model training by reducing the computational costs in parameter tuning and model inference [51].

Additionally, we enhanced our model by combining the information provided from high and low frequency data with a two pathway network Figure 11 similar to [57]. Both channels process the data separately in the ResGCN network. Before the fully connected layer, both 256 long feature vectors are concatenated. The fused information is passed trough three linear layers and then trough a Sigmoid layer to jointly predict the frustration rating $\hat{S}$. The channels take high frequency and low frequency data respectively. This allows for the high frequency channel to focus on fast movements and micro expressions, while the low frequency channel extracts the semantics of the data [57].

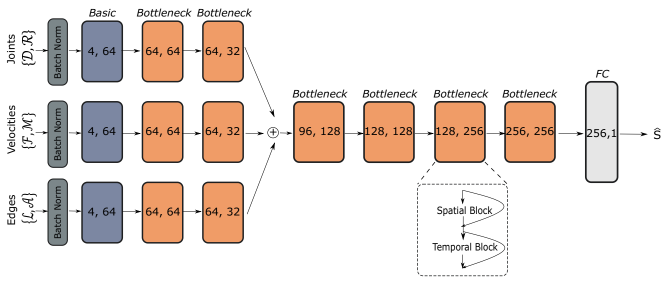


Fig. 10. Layer structure of one of the channels of the Fast&Slow network adapted from [51]

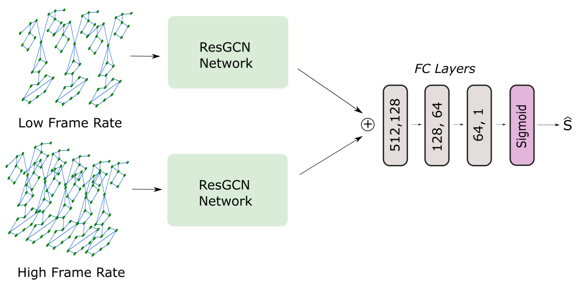


Fig. 11. Combination of a high frequency and low frequency channel adapted from [57]

### 3.4.2 LSTM Model

To compare our proposed network with an well-known baseline model, we employed an LSTM network with two stacked LSTM cells. The landmark, velocity and edge features are concatenated into one feature vector $\boldsymbol{X}$ and used as input to the LSTM model. The exact network structure can be seen in Fig. 12.

The two LSTM cells are followed by two ReLU and fully connected layer combinations and are finally processed by a Sigmoid layer that outputs the frustration rating $\hat{S}$.

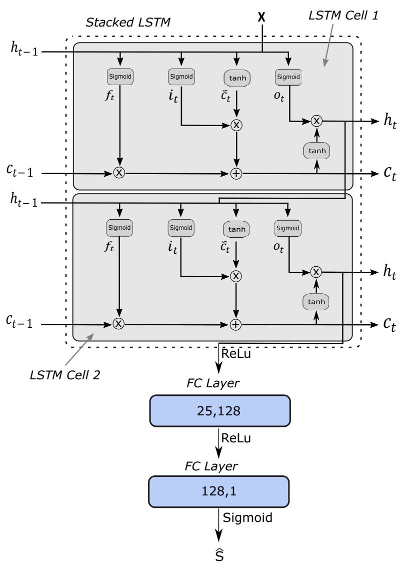


Fig. 12. LSTM layer structure adapted from [58]

TABLE 3
Parameters of the Sliding Window

| Model | Frequency [Hz] | Window [seconds] | Stride [seconds] |
|---|---|---|---|
| Fast&SLow | 20&5 | 7 | 2 |
| LSTM | 20 | 5 | 2.5 |

## 3.5 Training Details

Due to the non-comparable nature of facial landmarks in manual drives (head oriented towards windshield) and automated drives (head oriented towards UI), our network was trained using only manual drives. Consequently, each participant had three recordings, of which one was a baseline recording without induced frustration and two recordings correspond to different scenarios with induced frustration. The training parameters were set equal for both the Fast&Slow and LSTM network. Participants were divided into 24 for training, 4 for testing and 3 for validation, which results in $\approx$ 590k, $\approx$ 92k and $\approx$ 72k frames for training, validation and testing. A 5-fold cross-validation was performed for an optimal analysis, with 200 epochs of training. The validation dataset served as an early stopping criteria to avoid overfitting. Equal cross-validation splits were used for all analyzed models to facilitate the comparison between them. The used optimizer was the *Adam* optimizer, while the optimal learning rate was tested and set to $10^{-3}$. The batch size was set to 64. The proposed problem is a binary classification between the states "frustrated" and "not frustrated" with a continuous prediction label, hence *Binary Cross-Entropy* was chosen as the loss function. As the data was inserted into the network with a sliding window approach, further parameters that were optimized were the window and stride size. The optimal parameter setting found can be seen in Table 3.

Even though we included all participants in the data for completeness we decided to exclude all model recordings with a landmark detection rate from Affectiva lower than 80% from the study. Participants were completely removed from the analysis if they had more than one discarded recording, which was the case for six participants. Additionally, further six participants had to be excluded: one participant only had recordings of the automated driving condition, one participant had a very imbalanced frustration rating and four recordings had an inaccurate landmark detection as assessed by manual inspection of all videos. This resulted in 31 participants to train and test the model.

## 3.6 Evaluation Metrics

In the following we describe the evaluation metrics we chose in order to consider the subjectivity of the continuous frustration rating. During the cross-validation the optimal epoch was selected calculating the mean absolute error between the network prediction $\hat{S}_i$ and the ground truth $S_i$ frustration rating, for the total number of frames $n$ per subject:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{S}_i - S_i|. \tag{3}$$

The average MAE for each epoch was then computed. For the evaluation of the results using the test dataset, two main approaches were used. First, the cumulative prediction difference conveying the similarity between the predicted and real rating was computed. This was calculated according to Equation 5, taking the difference between predicted $\hat{S}$ and ground truth rating $S$ at frame $k$, then calculating the percentage of similarity over various allowed margins $m$ of difference. In order to interpret the results using this metric, the network was also evaluated using randomly assigned labels and compared to the obtained results. The advantage of using the cumulative prediction difference over other metrics is that this metric gives a more interpretable measure by giving information about the margin between the predicted and real rating. A frustration rating and especially the exact level of frustration is very subjective. Therefore, a metric that indicates the accuracy of the frustration pattern instead of the exact frustration level is needed to evaluate this model. In addition to the cumulative prediction difference the Root Mean Square Error (RMSE) was calculated.

$$\mathcal{C}(m) = \left\{ k \mid |\hat{S}_k - S_k| \leq m \; \forall k \in \{1, \ldots, n\} \right\} \tag{4}$$

$$F(m) = \frac{|\mathcal{C}(m)|}{n} \tag{5}$$

Thereupon, the area under the curve of the cumulative prediction difference was calculated using the trapezoidal rule, which gives an overall value of the performance of the model. In the following, this area will be referred to as Margin Area (MA) to avoid confusion with the Area Under the Curve usually calculated in classification problems. To also consider time shifts, we also conducted a visual inspection of the predicted and ground truth rating over time.

### 3.7 Model Results

Figure 13 presents the mean and standard variation of the cumulative prediction difference $F(m)$ of all five cross-validation folds. Table 4 shows the tabular quantification of the cumulative prediction differences. The Fast&Slow network provides the best results with a similarity of 79.4% with a variance of 7.7% allowing a margin of 0.1. In comparison, the LSTM network achieved a similarity of 69.3% with a variance of 8.6%. The high variance between folds is caused by the relatively small dataset available for training. Additionally, the networks have an RMSE of 0.10 for the Fast&Slow model and and 0.14 for the LSTM model. Best results were achieved with a comparably high window size of 7 seconds for the Fast&Slow and 5 seconds for the LSTM (Table 3). This underlines our finding that frustration is an emotion that evolves over a long period of time, hence, the network needs a large number of frames to attain the necessary information. Figure 15 illustrates some example plots of the predicted and ground truth ratings for the videos of different participants. The plots illustrate slight time shifts between the predicted and ground truth ratings. Additionally, poorly detected landmarks cause poor detection rates at specific time points. In order to carry out an in-depth analysis of the model performance, an ablation study on the input graphs was conducted. The model was trained once with the x- and y coordinates, once with the coordinates and the velocities of the input graphs, and finally with the coordinates and the edge information. Figure 14 compares the results when using different network branches.

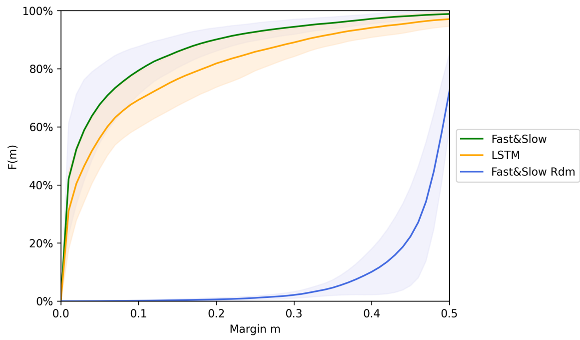


Fig. 13. Cumulative prediction differences comparing the Fast&Slow and LSTM network. The Figures also show the evaluation of the Fast&Slow network when using random (Rdm) labels. The curves show the average and the range of the five cross-validation folds.

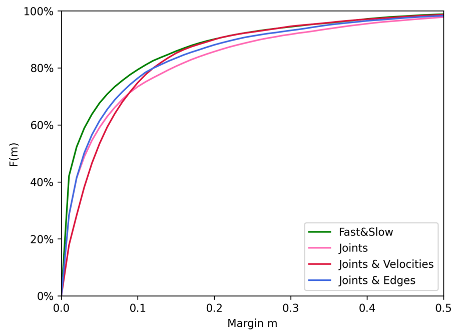


Fig. 14. Cumulative prediction differences comparing the Fast&Slow network including all three network branches, the joints and velocity branches together, and the joints and edge branches together. The curves show the average of the five cross-validation folds.

TABLE 4
Tabular description of the percentages of similarity reached with different marginal values for the Fast&Slow and LSTM network. 'MA' describes the margin area.

| Network | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | MA |
|---|---|---|---|---|---|---|
| **Fast&Slow** | **79.4±7.7** | **90.1±3.8** | **94.4±2.3** | **97.2±1.4** | **98.8±1.1** | **42.4±2.2** |
| LSTM | 69.3±8.6 | 81.8±7.3 | 89.0±4.4 | 94.1±2.9 | 97.1±2.2 | 39.1±2.6 |

## 4 Discussion & Conclusion

Datasets in which emotions occur naturally, are recorded under realistic circumstances and provide a subjective rating that can be used as one 'ground truth' label are crucial in order to develop well-working affect-aware systems. Such data are still rarely publicly available, and, to the best of our knowledge, not at all available for the emotion of frustration. As this emotion is so far less investigated than the six basic emotions despite its relevance in applications of affective computing, the FRUST dataset attempts to provide such a dataset. Previous work on emotion recognition has worked with either classical statistical or machine learning methods. Because both approaches have advantages, we

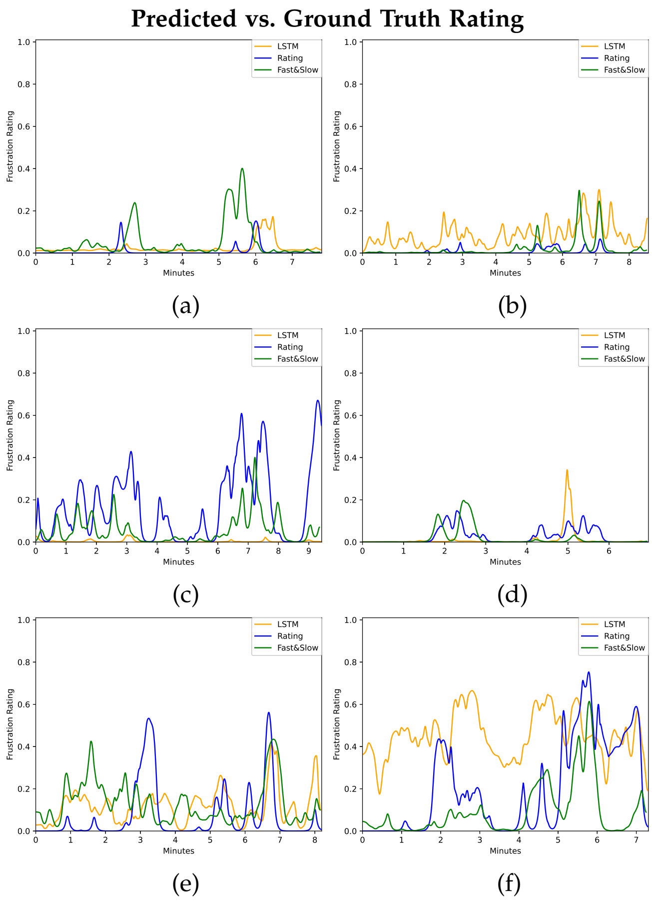


Fig. 15. Comparison of predicted and ground truth rating for Fast&Slow and LSTM network. The graphs represent the predictions from example videos taken from the test dataset.

provide both in this manuscript and their underlying data in the FRUST dataset. Even though it only contains data of 43 participants, the FRUST dataset has several advantages in comparison to freely accessible videos of acted or real-life situations on social media or elsewhere online. First, the context of data collection ensured that emotions in the dataset occurred naturally and were not acted. Second, the emotion-evoking situation is the same between participants following a standardized manner that allows a comparison of reactions between participants. Third, our dataset provides a continuous subjective frustration rating as valuable 'ground truth', which is not available in freely online accessible videos (e.g., on Youtube). The provided hand-annotations of expressions save researchers working with the dataset the tedious work these hand-annotations naturally bring with them. We also provide demographic information, which might be useful in finding sex- or age-typical patterns in expression of frustration. However, it needs to be mentioned that the dataset is likely skewed towards the local population, as all participants were German citizens.

In our study, frustration slowly built up over time, as shown by our dataset's subjective frustration ratings. The less likely it becomes to reach the aim of meeting friends at the cinema on time or join the online conference on time, the more frustrated participants became. When working with this dataset, therefore, it is important to consider this temporal feature.

A challenge common to any measure of subjective emotion data is the subjectivity of the frustration rating [59]. It was apparent that participants differed in the way they rated. Some participants rated spikes of frustration with long time of no frustration; others gave a continuous block of high frustration rating. We cannot know whether these variations are caused by the differences in how participants felt or by their interpretation of the task to rate their frustration. Also, absolute values of frustration ratings are difficult to compare, as 50% frustration might mean a different frustration intensity for different participants. Depending on how much a researcher decides to trust the participant's ability of providing comparable frustration intensity ratings, future users of this dataset may therefore decide to apply a z-transform to the subjective frustration rating. It is also a possibility that participants were frustrated but did not rate it as such, either because they were unaware of it or because they wanted to hide it from the experimenter [60]. Furthermore, the frustration rating might lag behind, as participants had to realize that something happened in the video before they could rate their frustration of that moment.

The hand-annotated expressions that have been shown most frequently include the previously described expressions of frustration of Brow Lowerer, Brow Raiser, Dimpler, Lip Press and Smile [18], [19], [38]. The results regarding expression frequency and subjective frustration indicate that expression frequency only should not be used in order to classify frustration.

The presented dataset was evaluated by our multi-pathway GCN architecture. In order to test the effectiveness of our model we conducted experiments with our suggested double pathway model and a LSTM model. Our model is based on the construction of a graph representing the facial structures at each frame. The model then learns spatial dependencies from edge connections between nodes and temporal dependencies from the connections between frames, while residual links allow the reduction of the computational cost. The architecture of the model builds upon three input branches containing facial landmarks, velocity and edge information that are later merged to form one main stream of information. All features are inferred from the facial landmarks and inserted in a graph format. Further enhancement of the model was obtained by combining high and low frequency data. The results show a promising automatic and frame-wise detection of frustration ratings with a similarity of 79.4$\pm$7.7% when allowing a prediction difference of 0.1. The example plots shown also give a sense of the accuracy of the detection pattern. Despite the limited amount of data and the subjective labels, it can be clearly seen that GCNs are able to predict complex frustration patterns with acceptable errors. To the best of our knowledge no publicly available dataset exists that can be interpreted with our model. Table 1, provides an overview of similar datasets. However, no emotion dataset exists with a continuous rating that could serve as an input to our spatiotemporal model. The DEFE [24] dataset provides similar input data, however, it recognizes six different emotions. Therefore, we provide a first baseline model for the automatic detection of frustration levels. For future work the information given by the body gestures of the participants could be included. Literature has shown that body expressions convey important affective information [61] and Narayanan et al. [44] already used skeleton sequences to successfully infer the emotional state from peoples' gait.

The FRUST dataset will be made publicly available after publication of this paper. By this, researchers can work with the collected data and use the presented model as benchmark. In summary, we show that it is possible to recognize a person's frustration in naturalistic data by direct and automated learning of coordinate data.

## Acknowledgments

This work was partly supported by the project F-RELACS which received funding from the German Federal Ministry of Education and Science in the program KMU-innovativ under the grant 16SV7930.

## References

[1] R. S. Lazarus, "Progress on a cognitive-motivational-relational theory of emotion." *American psychologist*, vol. 46, no. 8, p. 819, 1991.

[2] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the eeg during game play," *International journal of autonomous and adaptive communications systems*, vol. 6, no. 1, pp. 45–62, 2013.

[3] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[4] J. E. Richey, J. M. L. Andres-Bray, M. Mogessie, R. Scruggs, J. M. Andres, J. R. Star, R. S. Baker, and B. M. McLaren, "More confusion and frustration, better learning: The impact of erroneous examples," *Computers & Education*, vol. 139, pp. 173–190, 2019.

[5] P. Guchait and K. Namasivayam, "Customer creation of service products: role of frustration in customer evaluations," *Journal of services Marketing*, 2012.

[6] N. David Bowman, J. Keene, and C. J. Najera, "Flow encourages task focus, but frustration drives task switching: How reward and effort combine to influence player engagement in a simple video game," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–8.

[7] M. D. McHugh, A. Kutney-Lee, J. P. Cimiotti, D. M. Sloane, and L. H. Aiken, "Nurses' widespread job dissatisfaction, burnout, and frustration with health benefits signal problems for patient care," *Health affairs*, vol. 30, no. 2, pp. 202–210, 2011.

[8] A. Weidemann and N. Rußwinkel, "The role of frustration in human–robot interaction–what is needed for a successful collaboration?" *Frontiers in Psychology*, vol. 12, p. 707, 2021.

[9] A. Weidemann and N. Russwinkel, "Investigation of frustration," in *Proceedings of Mensch und Computer 2019*, 2019, pp. 819–824.

[10] K. Ihme, A. Unni, M. Zhang, J. W. Rieger, and M. Jipp, "Recognizing frustration of drivers from face video recordings and brain activation measurements with functional near-infrared spectroscopy," *Frontiers in human neuroscience*, vol. 12, p. 327, 2018.

[11] E. Bosch, K. Ihme, U. Drewitz, M. Jipp, and M. Oehl, "Why drivers are frustrated: results from a diary study and focus groups," *European transport research review*, vol. 12, no. 1, pp. 1–13, 2020.

[12] E. Bosch, M. Oehl, M. Jeon, I. Alvarez, J. Healey, W. Ju, and C. Jallais, "Emotional garage: A workshop on in-car emotion recognition and regulation," in *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2018, pp. 44–49.

[13] Jeon, Myounghoon and Walker, Bruce N and Yim, Jung-Bin, "Effects of specific emotions on subjective judgment, driving performance, and perceived workload," *Transportation research part F: traffic psychology and behaviour*, vol. 24, pp. 197–209, 2014.

[14] J. L. Deffenbacher, D. M. Deffenbacher, R. S. Lynch, and T. L. Richards, "Anger, aggression, and risky behavior: a comparison of high and low anger drivers," *Behaviour research and therapy*, vol. 41, no. 6, pp. 701–718, 2003.

[15] Y.-C. Lee, "Measuring drivers' frustration in a driving simulator," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 54, no. 19. Sage Publications Sage CA: Los Angeles, CA, 2010, pp. 1531–1535.

[16] S. Krüger, E. Bosch, K. Ihme, and M. Oehl, "In-Vehicle Frustration Mitigation via Voice-User Interfaces–A Simulator Study," in *International Conference on Human-Computer Interaction*. Springer, 2021, pp. 241–248.

[17] S. Zepf, T. Stracke, A. Schmitt, F. van de Camp, and J. Beyerer, "Towards real-time detection and mitigation of driver frustration using svm," in *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2019, pp. 202–209.

[18] M. Hoque and R. W. Picard, "Acted vs. natural frustration and delight: Many people smile in natural frustration," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 354–359.

[19] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Automatically recognizing facial indicators of frustration: a learning-centric analysis," in *2013 humaine association conference on affective computing and intelligent interaction*. IEEE, 2013, pp. 159–165.

[20] A. Belle, S.-Y. Ji, S. Ansari, R. Hakimzadeh, K. Ward, and K. Najarian, "Frustration detection with electrocardiograph signal using wavelet transform," in *2010 International Conference on Biosciences*. IEEE, 2010, pp. 91–94.

[21] M. Song, A. Mallol-Ragolta, E. Parada-Cabaleiro, Z. Yang, S. Liu, Z. Ren, Z. Zhao, and B. W. Schuller, "Frustration recognition from speech during game interaction using wide residual networks," *Virtual Reality & Intelligent Hardware*, vol. 3, no. 1, pp. 76–86, 2021.

[22] M. Song, Z. Yang, A. Baird, E. Parada-Cabaleiro, Z. Zhang, Z. Zhao, and B. Schuller, "Audiovisual analysis for recognising frustration during game-play: introducing the multimodal game frustration database," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, pp. 517–523.

[23] M. Macary, M. Tahon, Y. Estève, and A. Rousseau, "Allosat: a new call center french corpus for satisfaction and frustration analysis," in *Language Resources and Evaluation Conference, LREC 2020*, 2020.

[24] W. Li, Y. Cui, Y. Ma, X. Chen, G. Li, G. Zeng, G. Guo, and D. Cao, "A spontaneous driver emotion facial expression (defe) dataset for intelligent vehicles: Emotions triggered by video-audio clips in driving scenarios," *IEEE Transactions on Affective Computing*, 2021.

[25] W. Li, R. Tan, Y. Xing, G. Li, S. Li, G. Zeng, P. Wang, B. Zhang, X. Su, D. Pi *et al.*, "A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks," *Scientific Data*, vol. 9, no. 1, pp. 1–20, 2022.

[26] E. Bosch, D. Kaethner, M. Jipp, U. Drewitz, and K. Ihme, "Fifty shades of frustration: Intra- and interindividual variances in expressing frustration," *Transportation Research Part F: Traffic Psychology and Behaviour*, in review.

[27] M. Fischer, A. Richter, J. Schindler, J. Plättner, G. Temme, J. Kelsch, D. Assmann, and F. Köster, "Modular and scalable driving simulator hardware and software for the development of future driver assistence and automation systems," *New Developments in Driving Simulation Design and Experiments*, pp. 223–229, 2014.

[28] H. W. Krohne, B. Egloff, C.-W. Kohlmann, and A. Tausch, "Positive and negative affect schedule–german version," *Diagnostica*, 1996.

[29] D. Dupré, E. G. Krumhuber, D. Küster, and G. J. McKeown, "A performance comparison of eight commercially available automatic classifiers for facial affect recognition," *Plos one*, vol. 15, no. 4, p. e0231968, 2020.

[30] A. iMotions, "Affectiva imotions biometric research platform," 2015.

[31] L. Kulke, D. Feyerabend, and A. Schacht, "A comparison of the affectiva imotions facial expression analysis software with emg for identifying facial expressions of emotion," *Frontiers in psychology*, vol. 11, p. 329, 2020.

[32] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[33] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2012.

[34] F. Noroozi, D. Kaminska, C. Corneanu, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE transactions on affective computing*, 2018.

[35] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "Elan: A professional framework for multimodality research," in *5th International Conference on Language Resources and Evaluation (LREC 2006)*, 2006, pp. 1556–1559.
[36] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
[37] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.
[38] K. Ihme, C. Dömeland, M. Freese, and M. Jipp, "Frustration in the face of the driver: a simulator study on facial muscle activity during frustrated driving," *Interaction Studies*, vol. 19, no. 3, pp. 487–498, 2018.
[39] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2983–2991.
[40] V. M. Álvarez, C. N. Sánchez, S. Gutiérrez, J. Domínguez-Soberanes, and R. Velázquez, "Facial emotion recognition: a comparison of different landmark-based classifiers," in *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*. IEEE, 2018, pp. 1–4.
[41] D. H. Kim, M. K. Lee, D. Y. Choi, and B. C. Song, "Multi-modal emotion recognition using semi-supervised learning and multiple neural networks in the wild," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 529–535.
[42] L. Liu, T. Zhou, G. Long, J. Jiang, L. Yao, and C. Zhang, "Prototype propagation networks (ppn) for weakly-supervised few-shot learning on category graph," *arXiv preprint arXiv:1905.04042*, 2019.
[43] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal Graph Convolutional Networks for Skeleton-based Action Recognition," in *Thirty-second AAAI conference on artificial intelligence*, 2018.
[44] V. Narayanan, B. M. Manoghar, V. S. Dorbala, D. Manocha, and A. Bera, "Proxemo: Gait-based Emotion Learning and Multi-View Proxemic Fusion for Socially-Aware Robot Navigation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 8200–8207.
[45] T. Rao, J. Li, X. Wang, Y. Sun, and H. Chen, "Facial Expression Recognition with Multi-Scale Graph Convolutional Networks," *IEEE MultiMedia*, p. 1, 2021.
[46] D. Ong, Z. Wu, Z.-X. Tan, M. Reddan, I. Kahhale, A. Mattek, and J. Zaki, "Modeling emotion in complex stories: the stanford emotional narratives dataset," *IEEE Transactions on Affective Computing*, 2019.
[47] A. Sperduti and A. Starita, "Supervised Neural Networks for the Classification of Structures," *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 714–735, 1997.
[48] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral Networks and Locally Connected Networks on Graphs," *arXiv preprint arXiv:1312.6203*, 2013.
[49] A. Micheli, "Neural Network for Graphs: A Contextual Constructive Approach," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 498–511, 2009.
[50] B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," *arXiv preprint arXiv:1709.04875*, 2017.
[51] Y.-F. Song, Z. Zhang, C. Shan, and L. Wang, "Stronger, Faster and More Explainable: A Graph Convolutional Baseline for Skeleton-Based Action Recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1625–1633.
[52] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Gaitgraph: graph convolutional network for skeleton-based gait recognition," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2314–2318.
[53] G. Li, M. Müller, G. Qian, I. C. Delgadillo, A. Abualshour, A. Thabet, and B. Ghanem, "DeepGCNs: Making GCNs Go as Deep as CNNs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2021. [Online]. Available: http://arxiv.org/pdf/1910.06849v3
[54] Y.-F. Song, Z. Zhang, and L. Wang, "Richly Activated Graph Convolutional Network for Action Recognition with Incomplete Skeletons," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1–5.
[55] F. Yang, Y. Wu, S. Sakti, and S. Nakamura, "Make skeleton-based action recognition model smaller, faster and better," in *Proceedings of the ACM multimedia asia*, 2019, pp. 1–6.
[56] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.
[57] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "Slowfast networks for video recognition," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6202–6211.
[58] I. Pisa, I. Santín, J. L. Vicario, A. Morell, and R. Vilanova, "Ann-based soft sensor to predict effluent violations in wastewater treatment plants," *Sensors*, vol. 19, no. 6, p. 1280, 2019.
[59] K. Gardhouse and A. K. Anderson, "Objective and subjective measurements," *The Cambridge handbook of human affective neuroscience*, p. 57, 2013.
[60] T. F. Van de Mortel, "Faking it: social desirability response bias in self-report research," *Australian Journal of Advanced Nursing, The*, vol. 25, no. 4, pp. 40–48, 2008.
[61] M. Coulson, "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence," *Journal of nonverbal behavior*, vol. 28, no. 2, pp. 117–139, 2004.

**Esther Bosch** received the B.Sc. degree in Molecular Medicine at the University of Tübingen, Germany, in 2015 and the M.Sc. degree in Neural and Behavioral Sciences at the same university in 2017. Since 2018 she is pursuing a Ph.D. degree in Human Factors at the German Aerospace Center (DLR) in Braunschweig, Germany. She previously researched at the Max Planck Institute for Biological Cybernetics in Tübingen, Harvard University in Boston, TNO in Soesterberg (Netherlands), the Max Planck Institute for Intelligent Systems (Tübingen) and wrote her Master's thesis at Robert Bosch GmbH in Renningen. During her PhD, she absolved a research stay at the University of Chile, Santiago de Chile. Her special interest lays in (automated) recognition of a traveler's state with the aim to make innovative mobility solutions attractive to use.

**Raquel Le Houcq Corbí** received her B.Sc and M.Sc degree in Electrical Engineering and Information Technology from the Technical University Munich, Germany, in 2018 and 2021, respectively. She researched at the Academic Medical Centre Amsterdam (AMC), where she also wrote her bachelor's thesis. During her Master´s program, she worked as a software developer at Philips Medizin-Systeme Böblingen GmbH and as a research intern at Helmholtz Zentrum Munich. After completing her Master thesis at the German Aerospace Center (DLR) in Braunschweig, she continued to work there as a research assistant. Her interests lay in affective computing, pattern recognition and machine learning.

**Klas Ihme** studied Cognitive Science (B.Sc, 2007) at the University of Osnabrück, Germany, and the Middle Eastern Technical University in Ankara, Turkey. He graduated in Human Factors (M.Sc.) from Technical University Berlin, Germany, in 2010 and received his Ph.D. (Dr. rer. med.) from the University of Leipzig, Germany, in 2015. Since 2014 he works as a research scientist at the Institute of Transportation Systems of the German Aerospace Center in Braunschweig, Germany, where he leads the research group "Information Demand and Communication Design". His current research interests include user state recognition as input for user-adaptive automated systems.

**Stefan Hörmann** received his B.Sc. and M.Sc. degrees in Electrical and Computer Engineering in 2014 and 2017 from Technical University of Munich (TUM), Munich, Germany, respectively. Currently, he is pursuing his Ph.D. at the Institute for Human-Machine Communication at the Technical University of Munich, Germany. His current research interests include face recognition and face synthesis with emphasis on challenging scenarios.

**Meike Jipp** is Professor for "Transport Demand and Impact" at the Technical University of Berlin, Germany, and director of the Institute of Transport Research of the German Aerospace Center (DLR). In her research, she concentrates on human behavior in ground-based transportation systems. Meike Jipp studied psychology at Mannheim University, Germany, where she also completed her PhD thesis. She researched at the DLR-Institute of Flight Guidance, and became head of the "Human Factors"-Department at the DLR-Institute of Transportation Systems. Early in 2021, she took over her current positions in Berlin.

**David Käthner** studied Psychology at Technical University Chemnitz and graduated with a diploma in 2009. Since 2010, he has been working as a researcher at the Institute of Transportation Systems of the German Aerospace Center in Braunschweig, Germany. Working as data scientist and project manager in numerous national and European research projects, his expertise is in measuring, modelling and visualizing human behavior in complex socio-technical systems. He is greatly interested in the curation and distribution of high-quality open datasets.

Fifty Shades of Frustration: Intra- and Interindividual Variances in Expressing Frustration.

This is a published version of the manuscript:

Bosch, E., Käthner, D., Jipp, M., Drewitz, U., & Ihme, K. (2023). Fifty shades of frustration: Intra-and interindividual variances in expressing frustration. Transportation research part F: traffic psychology and behaviour, 94, 436-452. https://doi.org/10.1016/j.trf.2023.03.004.

Transportation Research Part F: Psychology and Behaviour 94 (2023) 436–452

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

# Fifty shades of frustration: Intra- and interindividual variances in expressing frustration

Esther Bosch [a,*], David Käthner [a], Meike Jipp [b], Uwe Drewitz [b], Klas Ihme [a]

[a] Institute for Transportation Systems, German Aerospace Center, Braunschweig, Germany
[b] Institute for Transportation Research, German Aerospace Center, Berlin, Germany

ARTICLE INFO

*Keywords:*
Facial expression of emotion
Interindividual variance
Emotion recognition
Affective computing
Empathic vehicles
Frustration recognition

ABSTRACT

Negative emotions like frustration can lead to risky driving behavior in manual driving and may hinder the acceptance of innovative automated vehicles. The option to automatically recognize such negative emotions of vehicle users with so called affect-aware systems has gained increasing attention within the last few years. This offers the possibility to adapt vehicle functions, such as the human–machine interface, in real-time. Emotional expressions in face and body form potential indicators for user frustration. Previous studies have investigated expressions of frustration in the context of driving and mobility, but have neglected situational and interindividual differences. In this paper, we examined the possibility to improve the recognition of frustration by considering individual differences. For this, a driving simulator study with 50 participants and a real-world driving study with 23 participants were conducted. An analysis of participants' facial expressions during frustrating driving situations confirms previously reported expressions of frustration (Brow Lowerer, Dimpler, Brow Raiser, Smile and Lip Press). In addition, the results also hint towards high variance between and low variance within participants for all other expressions, suggesting the existence of individual-typical expressions of frustration. Hence, future frustration-aware systems could benefit from considering these individual differences by using a universally trained algorithm that is then customized towards each individual.

## 1. Introduction

### 1.1. *In-vehicle frustration*

Frustration is a negative emotional state that occurs when goal-directed behavior is blocked (Lazarus, 1991). Persistent frustration fosters the experience of anger as well as aggressive behavior (Berkowitz, 1989; Jeon, 2015; Shinar, 1998) and can lead to experienced stress (Jeon & Zhang, 2013). By that, it can negatively influence the acceptance of human–machine systems (Picard & Klein, 2002). In addition, negative emotions, such as frustration, negatively affect cognitive skills used in driving (Jeon, 2015; Lee, 2010). In a system at capacity such as modern transportation, where most participants - especially car, bus, and truck drivers - want or need to reach their destinations as quickly as possible, many potential sources of frustration exist (Bosch et al., 2020). These include, for example, traffic jams on the highway caused by congestion, accidents, or construction sites; red traffic lights during urban rush hour; or slow tractors driving ahead on rural roads. Frustration can lead to aggressive behavior (Berkowitz, 1989), also on the road, and by that lead to

* Corresponding author.
*E-mail address:* esther.bosch@dlr.de (E. Bosch).

https://doi.org/10.1016/j.trf.2023.03.004
Received 24 January 2022; Received in revised form 1 March 2023; Accepted 2 March 2023
Available online 20 March 2023

accidents (Ma et al., 2018). The development of automated vehicles, which is intended to optimize traffic flow overall, promises to remedy this situation. However, autonomous vehicles are highly technical and thus complex systems for the normal user. The vehicles' intuitive design is challenging for engineers and designers. Therefore, users will likely experience frustration at many points when interacting with autonomous vehicles. Here, frustration can translate into a negative user experience (cf. Picard & Klein, 2002), which can negatively affect the evaluation and acceptance of automated vehicles. However, for a shift towards innovative and sustainable mobility solutions, travelers' acceptance of these new concepts is highly relevant. In summary, frustration experienced during driving has a negative impact on the overall safety of driving (Berkowitz, 1989; Jeon, 2015; Lee, 2010), as well as the user experience (Picard & Klein, 2002) and thus acceptance (Venkatesh et al., 2003) of automated vehicles. Reducing frustration is therefore highly desirable.

### *1.2. Frustration-Aware assistance systems*

Triggers for frustration are manifold (Bosch et al., 2020; Lawrence, 2006), vary between individuals (Ferreri & Mayhorn, 2022), and cannot always be predicted and therefore solved by design. For this reason, research towards automated recognition (Belle et al., 2010; Grafsgaard et al., 2013; Hoque & Picard, 2011) and subsequent mitigation (Oehl et al., 2019; Zepf et al., 2019) of frustration in human–machine interaction has increased within the last few years. One way to reduce frustration in cars would be to design frustration-aware assistance systems (Bruce, 1993; Krüger et al., 2021; Oehl et al., 2019; Picard & Klein, 2002; Stephan, 2015) that can detect 'pain points' by the current level of frustration of the driver (or other vehicle occupants), derive the driver's current needs from this and offer specific assistance based on this. This assistance then aims at either reducing the driver's level of frustration and bringing him or her to a different target state (e.g. relaxation, pleasure, or high attention) or at mitigating the negative consequences of frustration by providing support. This can be done by the initiation of intervention methods, like a voice assistant or ambient light (Braun et al., 2019; Krüger et al., 2021; Zepf et al., 2019). For example, the use of an empathic voice assistant that, on detection of anger, says 'Hey, are you alright? I can understand that you are a bit angry, sometimes I feel the same way. How about some music to take your mind off things?', has been shown to reduce negative emotions (Braun et al., 2019). Nass et al. (2005) examined a voice assistant and found that if the voice assistant's emotion matched the participant's emotions, fewer accidents happened and drivers paid more attention to the road. Harris and Nass (2011) conducted a driving simulator study in which frustration was induced and a voice assistant that reappraised the situation by saying that the other driver's actions were not intended offensively resulted in higher driving performance and less negative emotions. Further studies have found intervention methods like these to be effective for the increase of user acceptance (Grippenkoven et al., 2018; Herrenkind et al., 2019; Keller et al., 2019; Millonig & Fröhlich, 2018). Accordingly, recent research has tried to find relevant indicators for frustration that may be used as features for automated frustration recognition in frustration-aware systems. Many of these studies focused on frustration-typical facial expressions (Grafsgaard et al., 2013; Hoque & Picard, 2011; Kapoor et al., 2007), also in in-car settings (Ihme, Unni, et al., 2018; Lee, 2010; Malta et al., 2010), and have found Brow Lowerer, Dimpler, Brow Raiser, Smile and Lip Press to be often shown in frustration. While not studied for frustration yet, it has been found that bodily expressions can be relevant indicators of emotion, too (Kleinsmith & Bianchi-Berthouze, 2012; Noroozi et al., 2018; Wallbott, 1998).

### *1.3. Individual differences in expression of emotion*

To further improve in-vehicle frustration recognition, it is helpful to understand the occurrence of non-coincidental variance in frustration expression. Cohn et al. (2002) conducted two studies showing that it is possible to recognize individuals solely based on their facial expressions in response to emotional stimuli. This worked in two different contexts and over long-time intervals (12 and 4 months, respectively). One context was that participants watched a film alone and the other was a clinical interview. Gross (2008) describes that individuals differ on levels of emotion experience, behavioral responses, physiological responses and subsequent emotion regulation. Differences on all of these levels can lead to differences between individuals in expression of emotion. Barr et al. (2008) developed a taxonomy of individual differences in expression of emotions and found the two higher-order factors of emotional constraint and emotional expression. The corresponding first-order factors consisted of affect intensity, ambivalence about expression, disclosure of Negative emotion, disclosure of emotion, disclosure of Lack of affect, expression of Positive emotion, and Secret Keeping (Barr et al., 2008). Accordingly, Sangineto et al. (2014) built a personalized linear Support Vector Machine classifier to account for individual differences in expression of emotion. However, they take a black box approach, not considering the nature of differences between individuals. Also, the data used contain videos recorded in varying contexts. It is therefore impossible to differentiate among variance caused by individual or by situational differences. We argue that a descriptive approach that aims to study individual explanation of variance is an important basis to build generalizable systems. Here, we see the contribution of this paper.

### *1.4. Aim of the current work*

In summary, we aim to study the variance of expression of emotion in the applied context of in-vehicle frustration detection. Since a fine-tuning towards individuals would be feasible in the car, we study the expression of frustration in different individuals. Knowledge of the extent of this variance might be helpful to adapt future endeavors of (personalized) in-vehicle emotion recognition accordingly. The research question we aim to answer is 'Is it possible to recognize a person from one drive to the next based on their facial expressions?' A simulator study with 50 participants and a real-world driving study with 23 participants were conducted to answer this. The studies capture two different driving-related situations to ensure that our findings are generalizable over different driving-related contexts. According to Bosch et. Al. (2020), two situations that regularly cause frustration during driving are traffic and frustrating interactions with a human–machine interface (HMI). Accordingly, we chose these two as frustration inducing contexts. To include one context that is relevant nowadays and one that is more relevant in the future, the frustration induction by traffic was done in manual driving and the frustration induction by HMI in an automated driving context. In both studies, frustration was induced by blocking driving-related goals in different driving situations. Camera recordings captured the participant's face and body throughout the whole journey. Among other questions, a subjective frustration rating of the participants was collected after each drive.

## 2. Study 1 – Driving simulator study

### 2.1. Methods

#### 2.1.1. Summary

Study 1 was conducted in the high-fidelity driving simulator of the German Aerospace Center in Brunswick, Germany (Fischer et al., 2014) with 50 participants. Frustration was induced in two different contexts to foster the generalizability of our findings. Every participant experienced frustrating and baseline drives in both contexts. Subjective frustration ratings were collected after each drive on a 5-point-scale and after all drives as a continuous (i.e., highly time-resolved) rating after all drives as manipulation check. During all drives, cameras captured the participant's face and body in order to assess participants' expressions. Videos of the participants' face and body were annotated manually for facial and bodily expressions. For the data analysis, we identified the individual in one drive based on their expression behavior in another.

#### 2.1.2. Participants

Fifty participants recruited through the institute's participant pool took part in the study. Previous studies with similar scope and settings had comparable sample sizes (Hoque & Picard, 2011; Ihme, Unni, et al., 2018; Zhang et al., 2021). In total, nine participants were excluded from data analyses. Two of these nine participants became motion sick. For three participants simulator driving data were not saved, one had a condition of facial myoclonus and for three the frustration-rating data were partly missing. The $n = 41$ participants included in the analyses were aged 20 to 59 years (y) (M = 31.75 y, SD = 12.18 y, 13 female, 30 male). As reimbursement for their time, the participants received 5 € per commenced half hour for their participation.

#### 2.1.3. Materials

*2.1.3.1. Set-Up.* The experiment was conducted in a driving simulator virtual reality lab with 360° full view (Fischer et al., 2014). The participants sat in a realistic vehicle mock-up and controlled the mock-up car in the driving simulation (Virtual Test Drive, Vires Simulationstechnologie, Bad Aibling, Germany) via a standard interface with throttle, brake pedal, steering wheel and indicators. A user interface (UI) was displayed on a tablet (Microsoft Surface Pro 7, 12.3′) that was attached over the centre console of the car (needed for the frustration induction, see below). During all drives, the participant's face and body were recorded at 15 frames per second with one frontal camera and one that was placed above the screen displaying the UI. The latter was used to record the face when a participant oriented towards the UI in the automated condition. Another camera recorded the whole scene and was attached between the driver and the co-driver seat. All three cameras were network cameras from Axis, model M1065-L, recorded with a resolution of 1280 × 800 pixels.

*2.1.3.2. Stimuli.* We collected data in the simulator shown in Fig. 2 in two different driving modes (manual vs. automated) to include two different driving contexts. The participants read a story to immerse into the setting before all drives. In the manual driving mode, the participant read a story that told them they were supposed to meet friends at a cinema. It included the information that the usual driving time to the cinema was below ten minutes. They were told they would receive a 2 € reward if they reached their destination in time. A large clock in the car indicated the time left for timely arrival at the cinema. Tasks similar to this have successfully induced frustration in previous studies (Hoque & Picard, 2011; Ihme, Unni, et al., 2018; Malta et al., 2010). In the automated drives, the participants read a story that told them they were driving to a business meeting. Participants then solved a task on the in-car user interface displayed on the tablet. The participants, again, were told to receive a 2 € reward upon successful completion of their task.

*2.1.3.2.1. Frustration induction in manual driving mode. Baseline condition ('Baseline-Manual')*: The baseline condition consisted of a route which took five minutes to drive. In a simulated telephone call presented via loudspeakers, the participants were informed that the friends they were supposed to meet were running late. The scenario had minimal traffic and was designed in a way that the participants successfully reached the cinema within the given time.

*Frustration conditions ('Frust.-Manual1′ and'Frust.-Manual2′):* Frustration is induced when goal-directed behavior is blocked (Lazarus, 1991). Rendon-Velez et al., (2016) have found that time pressure increases the feeling of frustration in driving. Therefore, both goal-blocking events and time pressure were used in the experimental conditions. A telephone call at the beginning of the drive (again presented via loudspeakers) imitated the group of friends that were already waiting at the cinema, reminding the participant that they carried the entry tickets for all of them (time pressure). After 1 min of driving, a slow truck turned onto the road at a crossing, which lead to 1.3 min of driving behind the slow truck (goal blocking event). Right after the truck left the road at the next intersection, another slow vehicle (a car with horse trailer) turned onto the road and lead to another 1 min of slow driving (goal blocking event). Oncoming traffic was so dense that overtaking was impossible. Another call imitated the group of friends asking 'what's keeping you?' (time pressure). Subsequently, the participants reached an urban area which had three intersections with red lights. They were told to turn left at the third crossing by a voice imitating a navigation system. Five cars were already waiting in front of this third red light. The light was then green so shortly that only one person at a time could cross. In addition, the last waiting vehicle (a motorbike) was so slow to start driving that it missed a green phase. In total, the participants waited for 2 min in front of the left-turn lights (goal blocking event). After crossing the intersection, the cinema was located on the left side and the participants finished the drive. During the whole drive, the large clock counted down a time that was 30 s shorter than the track took to drive (time pressure). The task was to reach the cinema where friends were waiting in both frustration-inducing conditions. However, we exchanged car types of surrounding cars in the simulation and also the exact street course, meaning that the surrounding trees and also the turns in the street differed. We did this to deceive the fact that the same experimental condition was driven twice. As a result, one experimental condition took seven minutes to drive and the other ten. Frustrating events were the same in both experimental conditions and also took the same time.

*2.1.3.2.2. Frustration induction in automated driving mode. Baseline condition ('Baseline-Autom.'):* In the baseline condition, the participants were asked to visit a website, which could be accomplished easily. They were then asked to press the one button that appeared in different places of the UI. They were told to have no time pressure and to interact with the UI as natural as possible.

*Frustration conditions ('Frust.-Autom1′ and'Frust.-Autom2′):* In the frustration condition, the participants received a call from their 'boss', who told them that they were urgently needed for another, more important meeting and needed to turn around immediately to arrive on time. T.

he participants then had to change the destination of the navigation system. Through ambiguous naming of buttons, unclear icons, and unintuitive paths, this was hard to achieve within 7 min. In the second automation condition, a 'boss' called and asked the participant to very urgently join an online conference with clients. Again, the UI was so difficult to understand that it was hard to reach the goal of joining the online conference.

#### *2.1.3.3. Measures*

*2.1.3.3.1. Subjective frustration rating as manipulation check.* To assess whether frustration was successfully induced, the participants rated their frustration in two different manipulation checks. One was an emotion questionnaire was filled in after every drive. It first asked four distraction questions about gaze behavior in line with the cover story (see Supplementary Materials for the exact questions). Afterwards, the participants rated an emotion scale based on the German version of the positive and negative affect scale 'PANAS' (Krohne et al., 1996). It has a reliability of Raykovs ρ = 0.93 (Breyer & Bluemke, 2016) and is a commonly used method to acquire participant's emotions (see, for example, Barańczuk, 2018; Frison et al., 2019; Zhang et al., 2019). The exact emotions words used were 'active', 'distressed', 'interested', 'excited', 'upset', 'scared', 'inspired', 'proud','enthusiastic', 'ashamed', 'alert','nervous','determined','attentive','jittery','afraid' (from the original PANAS) and 'insecure', 'frustrated', 'angry', 'sad', 'surprised','relaxed' (our own addition) were rated on a 5-point scale from 'not at all' to 'extremely'. We decided to acquire this broad emotion spectrum to capture emotions that were possibly co-triggered, unintentionally, besides frustration. The dimensional scales of valence, arousal and dominance were also rated on a 5-point scale from negative to positive, excited to calm and influenced to independent, respectively.

The second manipulation check was obtained after all drives as additional subjective rating. For this, the participants watched the videos that were recorded during all drives of the whole scene (the participant's face was not visible; see Fig. 1) and rated their frustration with a joystick on a level from 0 to 100%. This rating was given continuously, i.e. the participant always held the joystick in the position that corresponded to their frustration level as experienced in the situation shown in the video. The joystick was moveable only in one direction and automatically returned to zero-position when not touched. The participants saw a visual feedback of their current rating, which was presented next to the video. They were asked to move the joystick according to the frustration level that they

**Fig. 1.** Screenshot of an exemplary video on which participants based their continuous frustration rating.

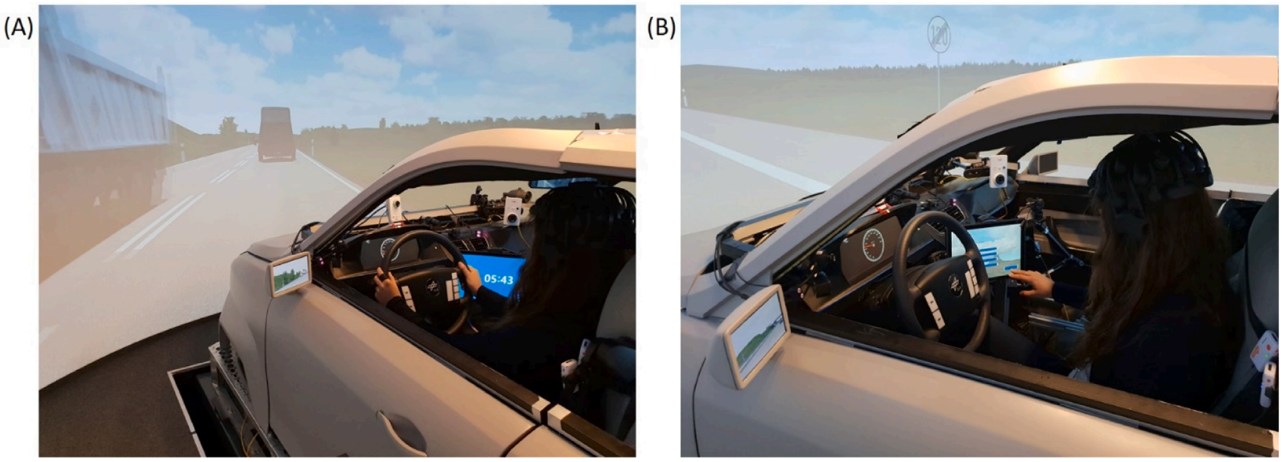


**Fig. 2.** Driving simulator with participant. Left. The manual driving mode; Right: the automated driving mode with interaction with the user interface.

felt in the situation shown in the presented video. By this, a continuous frustration rating for each drive and each participant was collected. We decided for this continuous measure in addition to the common method of questionnaires to receive a subjective rating not only once per drive, but for every timepoint during the drive.

*2.1.3.3.2. Facial expression annotation and preprocessing.* Three annotators manually annotated expressive units using the video recordings of all drives. The annotation was performed using the software ELAN (Wittenburg et al., 2006). More precisely, the annotators first came to accordance on all annotated expressive units and what they looked like, following Farnsworth (2019). Then, two annotators watched all videos of all participants in the software ELAN independently, where it is possible to add words under the video for freely selectable time windows. When an annotator discovered any deviation in facial expression from neutral, they added an annotation according to the previously chosen expressive unit names. In a second step, the third rater decided on all instances where the two ratings differed and chose one of the two annotations by an own judgement. For example, if annotator 1 annotated 'Smile' for a timepoint, and annotator 2 'Smirk', then annotator 3 checked the video at that timepoint and decided for either the 'Smile' or 'Smirk' as annotation. Annotations included the expressive units that were defined in the facial action coding system (Ekman, & Rosenberg, 1997), but were broadened to include head, body, and hand movements, as these have been shown to be relevant expressors of emotion, too (Kleinsmith & Bianchi-Berthouze, 2012; Noroozi et al., 2018; Wallbott, 1998). The categories that were annotated are shown in Table 1. Throughout the whole paper, the term 'expression' refers to facial movements in which several facial components are activated (e.g. an expression of joy). We use the term expressive unit as alternative to, but following the word 'action unit' (Ekman & Friesen, 1976), as we include head, body and hand movements. All words that were annotated (see Table 1) describe expressive units.

Expressive unit frequency was calculated by counting the absolute occurrence of expressive units in a first step and then divided by the length of the drive. For example, if 'Tongue Out' was shown 24 times and the drive was 8.62 min long, the expression frequency for 'Tongue Out' in this drive is 24/8.62 min = 2.78. Therefore, all occurrence frequencies are presented in occurrence per minute.

### *2.1.4. Procedure*

Participants arrived and filled in an informed consent and a data privacy statement. Before the start of the study, the participants were informed about the video recording, potential risks of driving in simulators (e.g., the experience of simulator sickness) according to the simulator safety concept and the rough duration of the experiment. The participants were informed that they could take a break or abort their participation at any time. All participants provided written informed consent to take part in the study and the video recording. The participants were told the cover story that the study investigated differences in gaze behavior between manual and automated driving modes. This was done to conceal the true aim of frustration induction and enable natural emergence of emotions. To reduce effects that came from unfamiliarity, all participants experienced the manual and automated driving scenarios before the start of the experiment until they said to be adapted to the simulator and the respective driving condition. This took five minutes on average. After the six drives, the participants were informed about the true goal of the experiment (evoking frustration) and the necessity to conceal this goal with a cover story. They then gave the continuous frustration rating for all six drives. The whole procedure took 2 h on average. The collected data was handled and saved in line with the European General Data Protection Regulation.

**Table 1**
Naming of the expressive units that were annotated.

| | |
|---|---|
| Lower Face | Blow Cheeks, Cheek Puffer, Chin Raiser, Dimpler, Asymmetric Dimpler, Jaw Drop, Laugh, Lip Corner Depressor, Lip Funneler, Lip Press, Lip Puckerer, Lip Stretcher, Pout, Asymmetric Pout, Smile, Smirk, Tongue Out, Upper Lip Raiser, Lip Corner Puller, Asymmetric LipCornerPuller, LipTightener, Asymmetric LipTightener, Lower Lip Depressor, Mouth Stretch, Nose Wrinkler |
| Upper face | Brow Lowerer, Brow Raiser, Roll Eyes, Squint, Upper Lid Raiser, Eyes Closed |
| Head | Head Back, Head Shake, Head Tilt, Head Wiggle, Chin Back, Head Forward, Swallow Hard, Move Jaw |
| Body | Deep Breath, Hands To Air, Hands To Face, Shrug, Straighten Up |

#### 2.1.5. Experimental design

In a 2 (driving mode: automated vs. manual) × 2 (frustration induction: frustration vs. baseline) within-subject design, each participant experienced six drives in total. Three of these were driven by the participants themselves (manual driving mode, see Fig. 2 A) and in three the car drove automatically (automated driving mode, see Fig. 2 B). Both driving modes consisted of one baseline drive and two frustration-inducing experimental drives each. The order of the drives was balanced by a balanced Latin square design for all participants, which means that every condition was driven in every position, and also the order of the drives was balanced (see for example Kim & Stein, 2009).

#### 2.1.6. Data analyses

*2.1.6.1. Manipulation check analysis.* A Spearman's rank correlation coefficient of the continuous subjective frustration rating with the emotion scale rating was calculated because the data was not normally distributed. We tested whether frustration induction was successful by comparing all conditions' frustration ratings against one another by a Friedman's test, as the assumptions for a repeated measures ANOVA were not fulfilled. Kendall's W was used for measuring effect size. The conditions were 'Baseline-Autom.','Baseline-Manual','Frust-Autom1′,'Frust-Autom2′,'Frust.-Manual1′ and'Frust.-Manual2′. The detailed plots included in the Supplementary Materials were plotted using the package 'ggstatsplot' (Patil, 2021).

*2.1.6.2. Expression frequency analysis.* In a first step, we created a plot that shows the expression frequency over all participants, per condition. We tested all conditions' expression frequencies against one another by a Friedman's test, as the assumptions for a repeated measures ANOVA were not fulfilled. Kendall's W was used for measuring effect size. We also created a bubble plot displaying all expression frequencies of all participants, colored by frustrating vs. non-frustrating drives. For the following analysis, we excluded the expressive units that are expected to be shown by all participants, as they are already known to be universally shown in frustration. These excluded expressions were 'Dimpler', 'Brow Lowerer', 'Brow Raiser', 'Smile', 'Lip Press' (Hoque & Picard, 2011; Ihme, Dömeland, et al., 2018; Ihme, Unni, et al., 2018; Sidney et al., 2005). If individuals are more similar to themselves than to others, it should be possible to recognize them by learning from one drive to another using non-universal expressive units. This was realized using a nearest neighbour approach using expressive unit frequencies from one drive to classify the individual in another. A similar approach has been taken by Cohn et al. (2002). For this, the Euclidean distances between all participants and all drives were calculated from vectors containing frequencies of all expressive units. More precisely, this means that a Euclidean space is created, in which each dimension corresponds to the expression frequency of one expressive unit. Each drive is one point in this Euclidean space. The distance between these points is calculated. The smaller the distance, the more similar the expressions that were shown in the two drives. A cumulative curve of the ranks (1st closest neighbour, 2nd closest neighbour, etc. in Euclidean space) in which the participant was successfully identified displayed how well this identification of individuals worked.

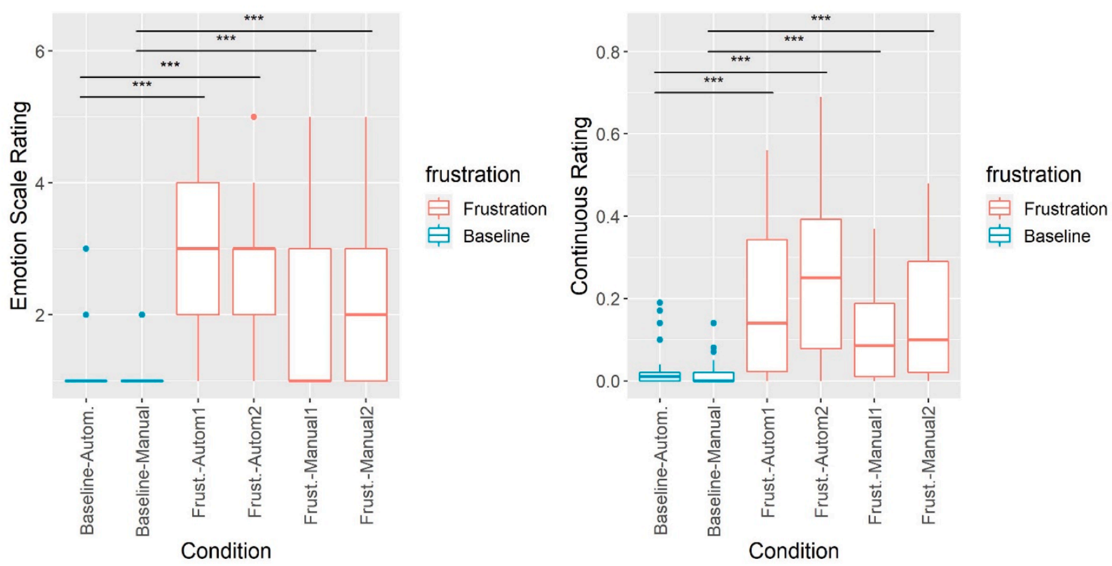


**Fig. 3.** Frustration ratings per condition over all participants. Left: The emotion scale frustration rating given after each drive. Right. Continuous frustration rating. The continuous frustration rating was averaged over each drive. Significance was tested by a Friedman's test. The lines with three asterisks above indicate significant differences between conditions of $p < 0.001$. In the automated (autom.) drives, participants drove in an automated vehicle and their task was to interact with the in-vehicle user interface. In Autom1 they were asked to change the navigation destination, in Autom2 to join an online conference, and in Baseline-Autom to make a search engine search, all on the user interface. The task was to meet friends at a cinema by manual driving in the 'Manual' conditions. Manual 1, 2 and baseline differed only by surrounding car types and the exact street course that were changed.

## 2.2. Results

### 2.2.1. Manipulation check

Frustration induction was successful as indicated by both subjective frustration ratings. Fig. 3 shows the emotion-scale rating of frustration after each drive (left) and the continuous frustration rating (right), both per condition. The emotion scale ratings of all other emotion words can be found in Fig. 12 in the Supplementary Materials. The Spearman's rank correlation coefficient of emotion scale rating per drive and mean continuous frustration rating per drive was 0.59, which is a high correlation according to Cohen (1988). The Friedman's test revealed that the frustration ratings – whether emotion scale or continuous – were significantly different between conditions, $\chi^2(5) = 122.07$, $p < .001$, W = 0.41 (emotion scale rating) and $\chi^2(5) = 116.63$, $p < .001$, W = 0.47 (continuous rating, see Fig. 3). Both are small effects according to Cohen (1988). Holm-corrected post hoc tests revealed that per driving mode, both subjective ratings were higher in the frustrating drives than in the baseline drives, but not different within baseline and frustration drives. The exact results are presented in Fig. 14 of the Supplementary Materials. Mean and median values can be found in Table 2. In summary, we assume that frustration was induced successfully in both the automated and manual driving modes. The emotions that seem to co-occur with frustration as rated with the 22 other emotion scale items (see Fig. 12 in the Supplementary Materials) were 'angry', 'upset' and 'ashamed'.

### 2.2.2. Expression frequency

Of all annotations, the two raters had the same annotation in 55% of the time, 10% were similar (e.g. Lip Suck and Lip Press), and 35% were identified by only one of the two raters. Therefore, in the 45% that the two annotations were different, the third rater decided for one of the two annotations. With a chance probability of $1/45 = 0.023$ for an annotated expression, this results in a Cohen's kappa of 0.55, which corresponds to a moderate inter-rater agreement (Landis & Koch, 1977).

Fig. 4 shows the expressive unit frequencies of every condition over all participants. Expressive unit frequencies are significantly different between conditions ($\chi^2$ (5) = 54.35, $p$ <.001, W = 0.29, $n_{pairs}$ = 38). This is a small effect according to Cohen (1988). Significant holm-adjusted pairwise Durbin-Conover tests are shown in Fig. 4. Fig. 5 shows the expressive unit frequencies of each participant, split by frustrating (red) and non-frustrating drives (blue). It becomes apparent that some expressive units have been shown by nearly all participants (Brow Lowerer, Brow Raiser, Deep Breath, Dimpler, Hands to Face, Jaw Drop, Lip Press and Smile) and others are only shown by a few of the participants.

*2.2.2.1. Identification of individuals.* If participants were individually typical in expressing frustration, their expressions should be more similar to themselves than to other participants in different drives. Therefore, individuals were identified in one drive based on their expressive unit frequencies in another drive by using a nearest neighbor approach,. This worked best within the same situations: the uppermost line of Fig. 6 shows that 37.5% of participants could be correctly identified with rank 1 (i.e., as their closest neighbor) from drive 'Frust.-Manual1′ to drive 'Frust.-Manual2′. When including rank two (cases, in which the participants were either their closest or second-closest neighbor), already 45% of participants were correctly classified. The second-upper line of Fig. 6 shows that participants are their first-closest neighbor in 25% of cases from drive 'Frust.-Autom1′ to 'Frust.-Autom2′ and their first or second-closest neighbor in 30% of cases. When trying to identify individuals across situations, recognition rate in the first-closest neighbor drops to values between 7.5 and 15%. Chance level is $1/41 = 2.4\%$.

## 2.3. Summary

Study 1 was conducted in a driving simulator and measured expressive units shown in baseline and frustration-inducing drives. The subjective measures confirmed that frustration was induced as planned. individuals can be correctly classified clearly above chance level from their expressive unit frequencies in one drive to another. These results support the option of individual-specific expressive units of frustration.

**Table 2**
Frustration ratings of emotion scale and continuous frustration rating in Study 1. The scale ranged from 1 to 5 in the Emotion Scale Rating with no anchor and from 0 to 1 in the Continuous Rating with an anchor at 0.

| | Condition | Mean | Median | Standard Deviation |
|---|---|---|---|---|
| **Emotion Scale Rating** | Baseline-Autom. | 1.13 | 1 | 0.41 |
| | Baseline-Manual | 1.15 | 1 | 0.36 |
| | Frust.-Autom1 | 2.91 | 3 | 1.17 |
| | Frust.-Autom2 | 2.67 | 3 | 1.14 |
| | Frust.-Manual1 | 1.96 | 1 | 1.15 |
| | Frust.-Manual2 | 2.22 | 2 | 1.28 |
| **Continuous Rating** | Baseline-Autom. | 0.02 | 0.01 | 0.04 |
| | Baseline-Manual | 0.01 | 0 | 0.03 |
| | Frust.-Autom1 | 0.18 | 0.14 | 0.16 |
| | Frust.-Autom2 | 0.26 | 0.25 | 0.2 |
| | Frust.-Manual1 | 0.11 | 0.09 | 0.11 |
| | Frust.-Manual2 | 0.15 | 0.1 | 0.14 |

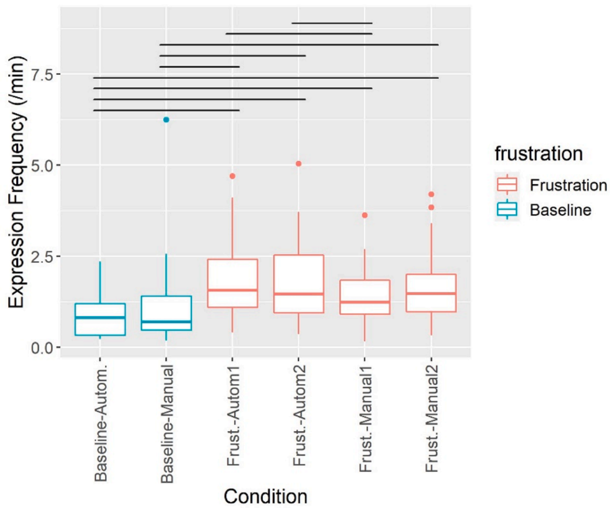


**Fig. 4.** Expression frequencies per condition. All conditions with lines above are significantly different with a significance level of $p<.001$ in pairwise holm-adjusted Durbin-Conover comparisons (which is not shown in the plot for clarity). In the automated (autom.) drives, participants drove in an automated vehicle and their task was to interact with the in-vehicle user interface. In Autom1 they were asked to change the navigation destination, in Autom2 to join an online conference, and in Baseline-Autom to make a search engine search, all on the user interface. The task was to meet friends at a cinema by manual driving in the 'Manual' conditions. Manual 1, 2 and baseline differed only by surrounding car types and the exact street course that were changed.

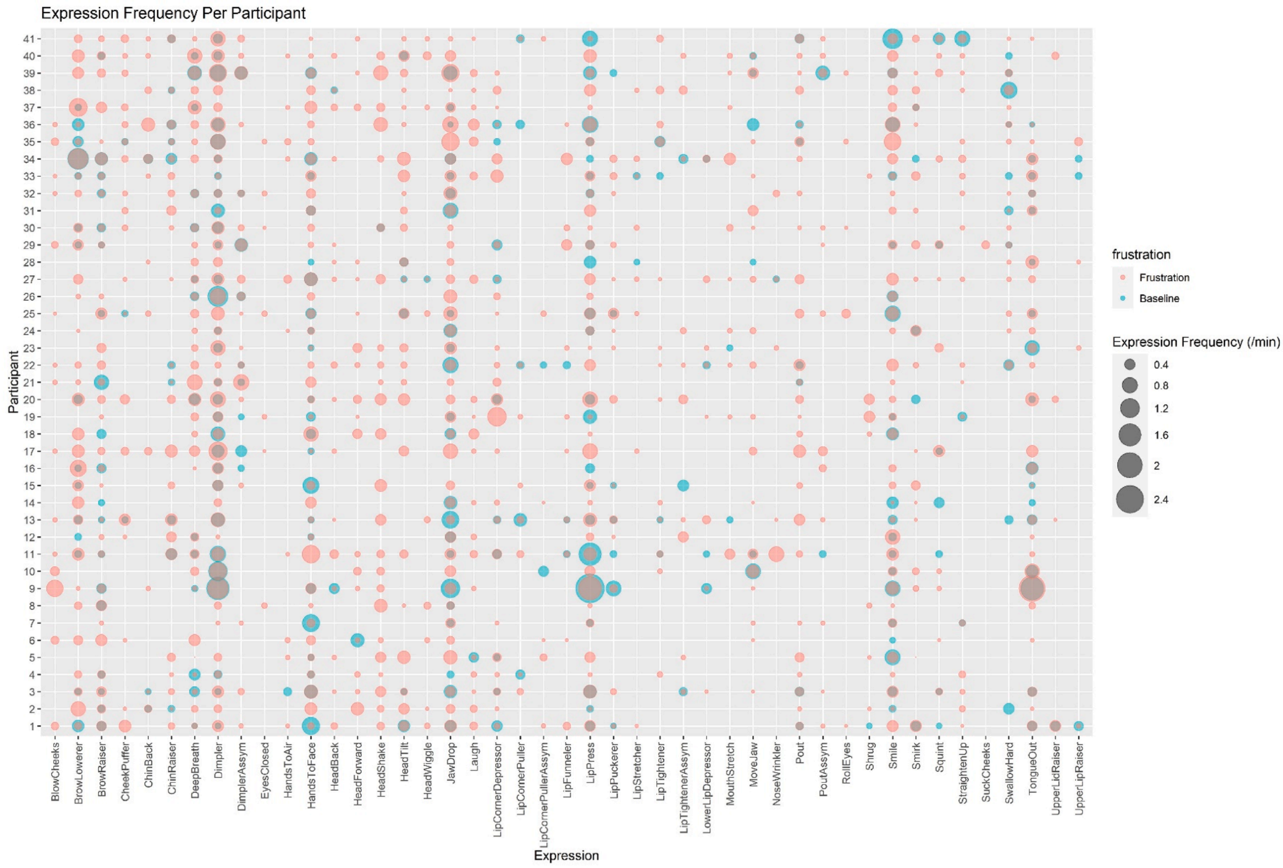


**Fig. 5.** Expression Frequency by participant, colors indicating whether the expression occurred in a frustration inducing or baseline drive. Vertical 'lines' of bubbles show the expressions that were shown in nearly all participants. The areas in between show the participant-typical expressions that were shown highly frequent in some, but not other participants.

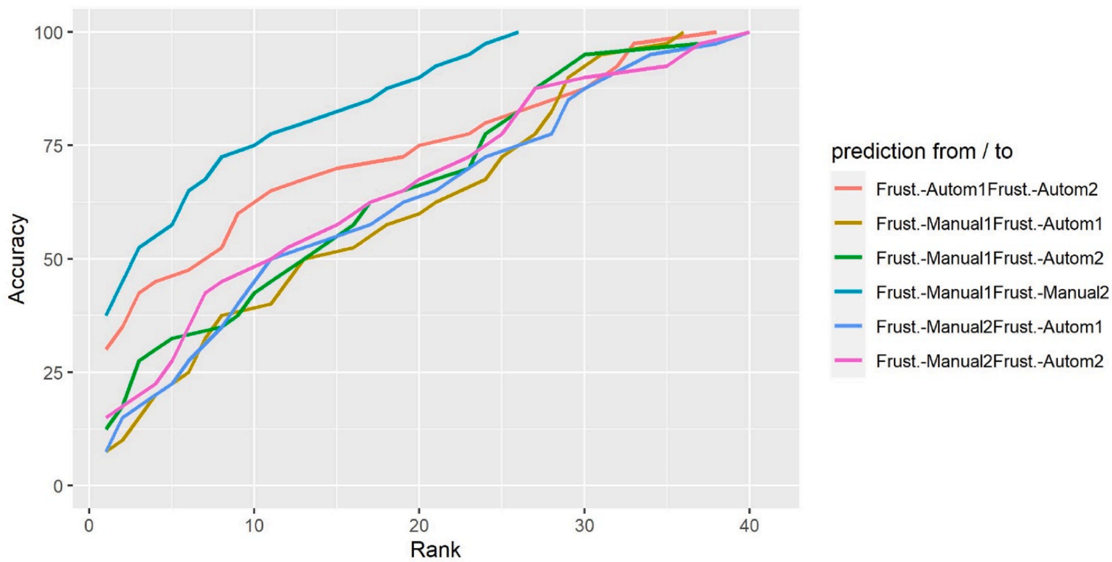


**Fig. 6.** Identification of individuals based on nearest neighbor in Euclidean space. A rank of 1 means that the vector of expressive units of participant 1 in drive 'automated 1′ is closest to themselves, a rank of 2 means that they are second-closest to themselves, etc. When using the vector of 'manual drive 1′ to identify individuals in 'manual drive 2′ (upper line), 42.5% of participants are already identified in rank 1 (i.e., closest to themselves).

## 3. Study 2 – Real-World driving study

### 3.1. Methods

#### 3.1.1. Summary

To test whether the results of the simulator study could be replicated in a real-world setting, we repeated a study as close as possible to Study 1 in a real car on a test track with 23 participants. As it was not feasible to use the manual driving mode frustration induction on the test track, Study 2 only used the automated driving mode. Every participant experienced baseline and frustrating drives which were all driven on the same test track. Subjective frustration ratings were collected after each drive on a 5-point-scale and as a continuous rating after all drives as manipulation check. The methods for Study 2 were the same as in Study 1, except for the driving mode (only automated driving) and that participants were brought to a test track before the start of the study, which took about 20 min. Also, the participants were different from the ones in Study 1. We include a shortened version of the methods again for clarity.

#### 3.1.2. Participants

Twenty-three participants were recruited through the institute's participant pool. The decision to recruit twenty-three participants was based on the tradeoff of measuring as many participants as possible within a feasible time of availability of the research car and the test track. One participant had to end the experiment early (for urgent private reasons). The $n = 22$ participants included in the analyses were aged 21 to 58 y (M = 40.8 y, SD = 10.8 y), of which 5 were female and 17 male. As reimbursement for their time, the participants received 5 € per commenced half hour for their participation.

#### 3.1.3. Materials

*3.1.3.1. Set-Up.* The experiment was conducted in a test vehicle of the German Aerospace Centre on a test track shown in Fig. 8 (comparable to SAE Level 4, SAE International, 2014). The participant sat in the driver seat and did not engage in any driving task. A security driver was present at all times on the co-driver seat with access to break and throttle. The car drove with a maximum speed of 30 km/h on a track of roughly 1.6 km that is shown in Fig. 8. The UI was displayed on a tablet (Microsoft Surface Pro 7, 12.3′) that was attached over the centre console of the car. During all drives, the participant's face and body were recorded at 15 frames per second with one frontal camera and one that was placed above the screen displaying the UI. The latter was used to record the face when a participant oriented towards the UI in the automated condition. Another camera recorded the whole scene and was attached between the driver and the co-driver seat. All three cameras were network cameras from Axis, model M1065-L, recorded with a resolution of 1280 × 800 pixels.

*3.1.3.2. Stimuli.* Frustration was induced through interaction with an in-car UI. The participants were told to receive a 2€ reward upon successful completion of their task. The participants read the same story before all three drives. In this story, they were asked to drive an automated car to a business meeting. The application used for interaction was exactly the same as in Study 1.

*Baseline condition ('Baseline'):* In the baseline condition, the participants were asked to visit a website, which could be accomplished easily. They were then asked to press the one button that appeared in different places of the UI. They were told to have no time pressure and to interact with the UI as natural as possible.

*Frustration condition ('Frust1′ and'Frust2′):* In the frustration condition, the participants received a call from their 'boss', who told them that they were urgently needed for another, more urgent meeting and needed to turn around immediately to arrive on time. The participants then had to change the destination of the navigation system. Through ambiguous naming of buttons, unclear icons, and unintuitive paths, this was hard to achieve within 7 min. In the second automation condition, a 'boss' called and asked the participant to very urgently join an online conference with clients. Again, the UI was so difficult to understand that it was hard to reach the goal of joining the online conference.

#### *3.1.3.3. Measures*

*3.1.3.3.1. Subjective frustration rating as manipulation check.* An emotion questionnaire was filled in after every drive. In a first step, the participants filled in a questionnaire that asked four distraction questions about gaze behavior in line with the cover story. Afterwards, the participants rated an emotion scale based on the positive and negative affect scale 'PANAS' (Krohne et al., 1996) and the dimensional scales of valence, arousal and dominance. This was the same questionnaire as in Study 1. Also, the continuous frustration rating was obtained after all drives in the same way as is Study 1. The conditions were 'Baseline', 'Frust1′ and'Frust2′.

*3.1.3.3.2. Facial expression annotation and preprocessing.* Three annotators manually annotated expressive units using the video recordings of all drives. This was done in exactly the same way as in Study 1 (see Table 1 for annotated expressive units). Also, the determination of expressive unit frequency worked as described in Study 1. Again, annotations included the expressive units that were defined in the facial action coding system (Ekman, & Rosenberg, 1997), but were broadened to include head, body, and hand movements. Expressive unit frequency was calculated by counting the absolute occurrence of expressive units in a first step and then divided by the length of the drive.

### *3.1.4. Procedure*

Participants arrived and first filled in an informed consent and a data privacy statement. They were then brought to a test track, which took about 20 min. Before the start of the study, the participants were informed about the video recording, potential risks of driving in an automated vehicle on a test track with safety driver (e.g., the experience of motion sickness) according to the vehicle safety concept and the rough duration of the experiment. The participants were informed that they could take a break or abort their participation at any time. All participants provided written informed consent to take part in the study and the video recording. The participants were told the cover story that the study investigated gaze behavior in automated driving. This was done to conceal the true aim of frustration induction and enable natural emergence of emotions. To reduce effects that came from unfamiliarity, all participants experienced an automated drive before the start of the experiment until they said to be adapted to the driving. This took five minutes on average. After the three drives, the participants were informed about the true goal of the experiment (evoking frustration) and the necessity to conceal this goal with a cover story. They then gave a continuous frustration rating for all three drives. The whole procedure took 2 h on average. The collected data were handled and saved in line with the European General Data Protection Regulation.

### *3.1.5. Experimental design*

Each participant experienced three drives in a within-participants design. In all three drives, the car drove in fully automated mode, i.e. the participant did not engage in any driving task (see Fig. 7). Both driving modes consisted of one baseline drive and two frustration-inducing experimental conditions each. The order of the drives was determined by a balanced Latin square design for all participants, which means that every condition was driven in every position, and also the order of the drives was balanced (see for example Kim & Stein, 2009). After each drive, participants gave a 5-point frustration rating. After all drives, participants provided a continuous frustration rating. Both served the purpose of a frustration induction manipulation check. During all drives, cameras captured the participant's face and body in order to assess participants' expressions.

### *3.1.6. Data analyses*

*3.1.6.1. Manipulation check analysis.* A Spearman rank correlation coefficient of the continuous subjective frustration rating with the emotion scale rating was calculated because the data was not normally distributed. We tested whether frustration induction was successful by comparing all conditions' frustration ratings against one another by a Friedman's test, as the assumption for a repeated measured ANOVA were not fulfilled. Kendall's W was used for measuring effect size. The conditions were 'Baseline', 'Frust1′ and

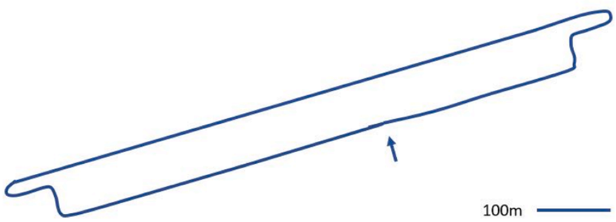


**Fig. 7.** Test track on which the automated car was driving. The arrow indicates start and end point of the track.

**Fig. 8.** Setup in the automated vehicle: the driver was interacting with a user interface while the car was driving fully automated in Study 2.

'Frust2′.

*3.1.6.2. Expression frequency analysis.* In a first step, we created a plot that shows the expression frequency over all participants, per condition. We tested all conditions' expression frequencies against one another by a Friedman's test, as the assumptions for a repeated measures ANOVA were not fulfilled. Kendall's W was used for measuring effect size.We also created a bubble plot displaying all expression frequencies of all participants, colored by frustrating vs. non-frustrating drives. The recognition of individuals from one drive to another was again realized using a nearest neighbor approach using expressive unit frequencies from drive 'Frust1′ to classify the individual in drive 'Frust2′. For this, the Euclidean distances between all participants and all drives were calculated from vectors containing frequencies of all expressive units. A cumulative curve of the ranks in which the participant was successfully identified displayed how well this identification of individuals worked.

### 3.2. Results

#### 3.2.1. Manipulation check

Frustration induction was successful as rated by the continuous rating, but not the emotion scale rating. Fig. 9 shows the emotion-scale rating (left) and the continuous frustration rating of frustration after each drive (right), both per condition. The emotion scale

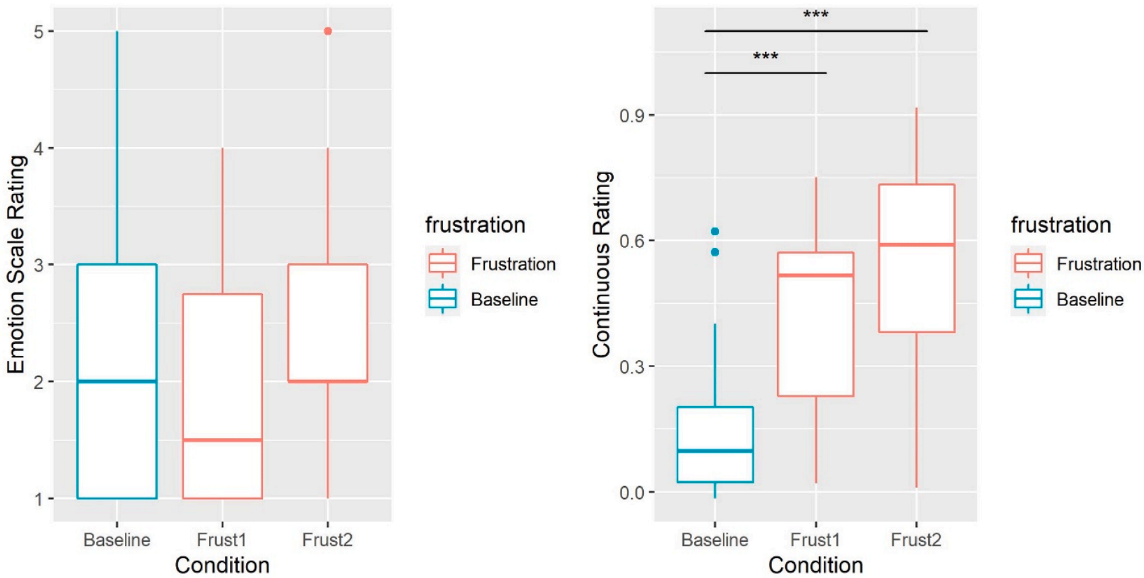


**Fig. 9.** Mean values over all participants of the continuous rating, per condition. The lines with three asterisks above indicate significant differences between conditions of $p < 0.001$. Participants drove in an automated vehicle and their task was to interact with the in-vehicle user interface. In Autom1 they were asked to change the navigation destination, in Autom2 to join an online conference, and in Baseline-Autom to make a search engine search, all on the user interface.

ratings of all other emotion words can be found in Fig. 15 in the Supplementary Materials. The Friedman's test revealed that the emotion scale frustration rating was not significantly and with no effect according to Cohen (1988) different between conditions, $\chi^2(2) = 2.33$, $p = .31$, W = 0.04. However, the continuous rating was significantly different between conditions $\chi^2(2) = 20.73$, $p < .001$, W = 0.47 (shown in Fig. 9), which corresponds to a small effect according to Cohen (1988). Post hoc tests were used with Holm correction applied (see Fig. 17 in the Supplementary Materials). The continuous frustration rating was significantly higher in both frustration-inducing drives than the baseline drive, but not different between the two frustration-inducing drives. Mean and median values can be found in Table 3. The results of each condition's comparison are presented in Fig. 16 of the supplementary materials. The emotion scale rating consisted of 23 other emotion words that were captured to check whether we, unintentionally, triggered other emotions in the experiment (see 3.1.3.3.1). None of the 23 other emotion words shows a clear pattern between the conditions (see Supplementary Materials, Fig. 15).

### *3.2.2. Expression frequency*

Of all annotations, the two raters (not the same as in Study 1) had the same annotation in 51.3% of the time, 27.4% were similar (e.g. Lip Suck and Lip Press), and 21.4% were rated by only one of the two raters. With a chance probability of 1/45 = 0.023 for an annotated expression, this results in a Cohen's kappa of 0.51, which corresponds to a moderate inter-rater agreement (Landis & Koch, 1977).

Fig. 10 shows the expressive unit frequencies of every condition over all participants. Expressive unit frequencies are significantly different between conditions ($\chi^2(2) = 20.82$, $p < .001$, W = 0.47, $n_{pairs} = 22$). Significant Holm-adjusted pairwise Durbin-Conover tests are shown in Fig. 10. Fig. 11 shows the expressive unit frequencies of all participants, split by frustration and baseline drives. It becomes apparent that some expressive units have been shown by nearly all participants (Brow Lowerer, Brow Raiser, Chin Raiser, Deep Breath, Dimler, Jaw Drop, Lip Press and Smile) and others are only shown by a few of the participants. Generally, expression frequency seems to be higher in frustration compared to baseline drives.

*3.2.2.1. Identification of individuals.* If participants were individually typical in expressing frustration, their expressions should be more similar to themselves than to other participants in different drives. Using a nearest neighbor approach, individuals were identified in one drive based on their expressive unit frequencies in the second. As can be seen in Fig. 11, this worked better for the direction of recognizing the individual in drive 'Frust2′ from information of drive 'Frust1′: 42.86% of participants were their closest neighbor from drive 'Frust1′ to drive 'Frust2′, 52.39% were their first or second-closest neighbor. From drive 'Frust2′ to 'Frust1′, only 23.81% of participants are their closest neighbor, but already 38.1% are their first- or second-closest neighbor. Chance level is 1/21 = 4.8%.

### *3.3. Summary*

Study 2 was conducted in a real car on a test track and measured expressive units shown in baseline and frustration-inducing drives. The continuous frustration rating as manipulation check confirmed that frustration was induced as intended, whereas the emotion scale rating did not reveal any difference between drives. The identity of the individual is correctly classified as closest neighbor in already 42.86% (23.81% in the reverse case) of participants, which is clearly above the chance level of 4.8%. This points towards a person-typicality of non-universal expressive units.

## 4. General discussion

The purpose of this work was to determine the interindividual variance in expression of in-vehicle frustration. Therefore, we conducted a driving simulator study and a real-world driving study that consisted of baseline and frustration-inducing drives. Subjective frustration ratings were obtained and facial and bodily expressive units were annotated for all drives. As a result, evidence for individual-typical, expressive units of frustration were found: Within both studies, it was possible to recognize the individual from one drive to another by previously shown facial expressions clearly above chance level. We can show that this is true in two different driving contexts (automated and manual driving) and is therefore most likely generalizable also to other driving contexts.

### *4.1. Frustration induction and expression*

Frustration induction was successful as confirmed by both frustration ratings in the simulator study and by the continuous frustration rating in the real-world driving study. Therefore, we can assume that participants really experienced frustration in the frustrating scenarios. The absence of differences among conditions in the emotion scale rating in Study 2 might be explained by the effect of novelty of the automated driving experience. We asked participants to rate their level of frustration in the situation as a whole, which was the novel and interesting situation of driving a real automated car, while the continuous rating forced participants to reconsider the exact situation and moments of frustration within the automated driving experience. In the emotion scale rating of all 22 rated emotions, positive emotions are generally rated higher than negative emotions in Study 2. This further supports the hypothesis that the participants did not differentiate between the generally exciting experience of driving an automated car and the particular driving condition. We therefore assume the continuous frustration rating to be a valid indicator of a successful manipulation towards frustration and non-frustration.

Previously described universal expressions of frustration were confirmed in this study, as the expressions that are shown among all

**Table 3**
Frustration ratings of emotion scale and continuous frustration rating in Study 2. The scale ranged from 1 to 5 in the Emotion Scale Rating with no anchor and from 0 to 1 in the Continuous Rating with an anchor at 0.

| | condition | Mean | Median | Standard Deviation |
|---|---|---|---|---|
| **Emotion Scale Rating** | Baseline | 2.32 | 2.00 | 1.32 |
| | Frust1 | 1.82 | 1.50 | 0.96 |
| | Frust2 | 2.36 | 2.00 | 1.09 |
| **Continuous Rating** | Baseline | 0.16 | 0.10 | 0.18 |
| | Frust1 | 0.41 | 0.52 | 0.22 |
| | Frust2 | 0.52 | 0.59 | 0.26 |

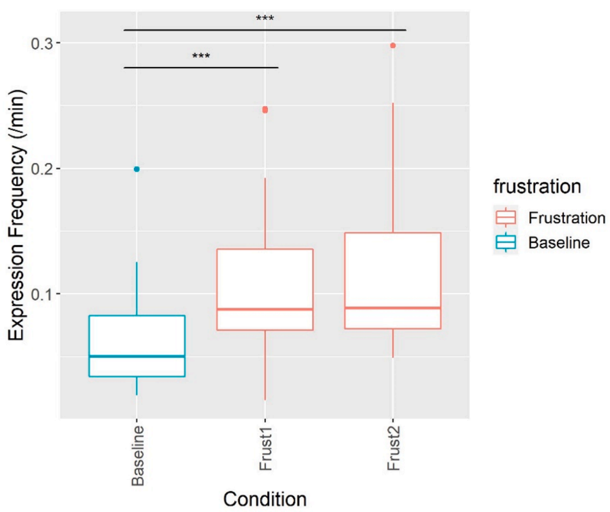


**Fig. 10.** Expression frequencies per condition. *** indicate a significance level of $p < .001$ in pairwise holm-adjusted Durbin-Conover comparisons.

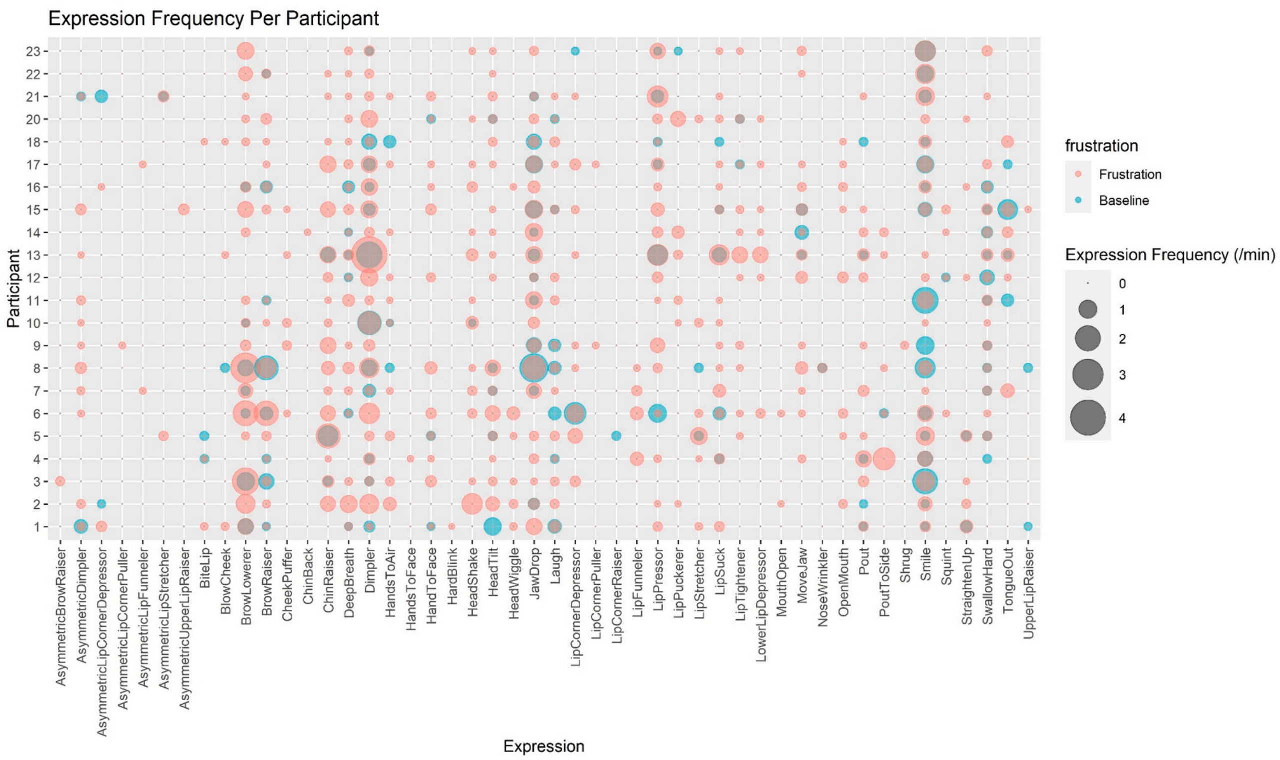


**Fig. 11.** Expression frequency per expression and per participant.

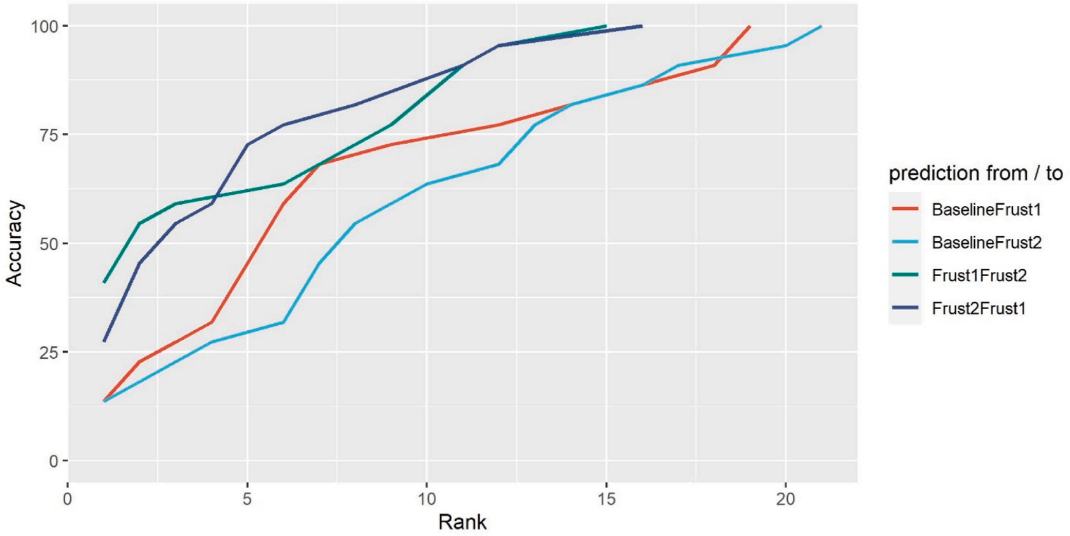


**Fig. 12.** Identification of individuals based on nearest neighbor in euclidean space. A rank of 1 means that the vector of expressive units of participant 1 in drive 'frust1′ is closest to themselves, a rank of 2 means that they are second-closest to themselves, etc. When using the vector of 'frust1′ to identify individuals in 'frust2′ (upper line), 42.86% of participants are already identified in rank 1 (i.e., closest to themselves).

participants coincide with the expressions that have previously been described as frustration-typical (see Section 1, Hoque & Picard, 2011; Ihme, Unni, et al., 2018; Sidney et al., 2005). In addition, we find individual-typical expressions of frustration. This can be explained by biological and learning reasons regarding the individual and the specifics of the situation according to Ekman (1992). According to Barrett (2017) and Hutchinson & Barrett (2019), each unique instance of emotion is influenced by context, culture and previous experiences. Therefore, one would expect expressive units that differ between participants, but are more similar within participants, as all these influencing factors are either constant within a participant (culture, previous experience) or kept constant in the experiment (context). Scherer et al. (2009) present a model that assumes emotion to be based on several appraisal steps. The major checks for these entail a judgement of relevance, implication, coping ability and normative significance for the individual (Scherer, 2009). Differences in the resulting emotion can be due to differences during these appraisal steps, that result from differences in social learning and cultural meaning systems (Scherer, 2009). Furthermore, Scherer explains individual differences in emotional reactions through differences in the individual's evaluation of and role within the situation, depending on their goals, values and coping potential (Scherer, 2009). Similarly, Kaiser and Wehrle (2001) attribute differences in emotion expression to differences in appraisal tendencies and coping styles. Overall, the practical meaning of these findings is that for recognition of emotions, universal expressions are a good 'starting point' that can be refined by learning additional, individual-specific expressions.

### *4.2. Limitations and future work*

For future research, it would be interesting to investigate the amount of constancy within an individual over a longitudinal study that considers expressive units shown by participants on different days. One limitation of this study is the moderate inter-rater agreement of the facial and bodily hand-annotations. However, previous research using hand-annotations in an in-vehicle context has similar inter-rater reliabilities (Schömig et al., 2018). As the automated recognition of facial and bodily expressions is still in development and far from perfect (see, for example, Dupré et al., 2020), the method of using three annotators is as close as we could get to a precise annotation of expressions. Between the drive and the emotion scale questionnaire, participants filled in four short distraction questions about gaze behavior. It is a possibility that these questions distracted the participants from their experienced emotion and that the reported emotions in the emotion scale rating are therefore slightly different than they would have been right after the drive. However, as the four distraction questions were very short, we argue this influence to be minimal. Furthermore, the continuous frustration rating was done at the end of all drives, which meant that participants had to remember their frustration level. However, the correlation between the continuous rating and the emotion scale rating is quite high. The emotion scale rating was based on the PANAS, which is a commonly used method to acquire participant's emotions (see, for example, Barańczuk, 2018; Frison et al., 2019; Zhang et al., 2019). It should also be mentioned that this study was focused on frustration and compared it to a baseline condition without emotion induction. The results need to be replicated in a study in which other emotions are studied in addition to frustration, for example happiness, to test whether other emotions are also characterized by individual-typical expressive units. We speculate this to be the case, as it has been shown to be the case in a different context, but otherwise similar study, by Cohn et al. (2002). From a theoretical perspective this is also likely, as also for other emotions, previous experience, appraisal steps and previous learning experience are constant within a person, but not between persons.

In summary, our findings can help to improve in-vehicle recognition of frustration by considering individual-typical expressions. A successful recognition of frustration enables the system to offer help or information accordingly, and by that mitigate frustration. This has been shown to be effective using a voice assistant, ambient light, notifications, or other intervention strategies (Braun et al., 2019;

Zepf et al., 2019). This is meaningful for transportation research, as frustrated drivers are more likely to take risky driving maneuvers (Jeon, 2015). A successful frustration mitigation can therefore contribute to safer roads as long as humans are still needed as drivers. In the future, the recognition and successful mitigation of frustration is relevant to promote a high ease of use and by that acceptance of newly introduced mobility solutions, like for example autonomous vehicles (Xu et al., 2018). Therefore, if frustration is recognized on time and help offered, a smooth interaction can be supported and, by this, the acceptance of new mobility solutions can be fostered.

## 5. Conclusion

The findings of this study indicate the relevance and possible individual-typicality of expressive units of frustration in the context of emotion recognition for emotion-aware manual and automated vehicles. For in-vehicle recognition of emotion, we therefore suggest to train frustration-aware systems on recognizing universal expressive units at first, which can then be refined per user on the user's specific, individual-typical expressive units of frustration. By this, it may be possible to recognize a user's 'pain points' even more reliably, which again enables to offer solutions in real-time. By this we hope to achieve less frustration in the vehicle and therefore safer roads, and to ameliorate user's acceptance of new mobility concepts like automated vehicles.

*CRediT authorship contribution statement*

**Esther Bosch:** Project administration, Conceptualization, Data curation, Methodology, Visualization, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **David Käthner:** Data curation, Methodology, Visualization, Formal analysis, Writing – review & editing. **Meike Jipp:** Conceptualization, Supervision, Writing – review & editing. **Uwe Drewitz:** Funding acquisition, Conceptualization, Writing – review & editing. **Klas Ihme:** Funding acquisition, Project administration, Conceptualization, Methodology, Visualization, Formal analysis, Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

This work was partly supported by the project F-RELACS which receives funding from the German Federal Ministry of Education and Science in the program KMU-innovativ under the grant 16SV7930. We thank Anne Rybizki for help with conducting the simulator study and the manual annotations of expressive units. We furthermore would like to thank Dr. Anja Katharina Huemer and Meng Zhang for fruitful conversations about statistical analysis.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trf.2023.03.004.

## References

Barańczuk, U. (2018). Emotion regulation mediates the effects of temperament traits and posttraumatic stress disorder symptoms on affect in motor vehicle accident survivors. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 528–535.

Barr, L. K., Kahn, J. H., & Schneider, W. J. (2008). Individual Differences in Emotion Expression: Hierarchical Structure and Relations with Psychological Distress. *Journal of Social and Clinical Psychology, 27*(10), 1045–1077. https://doi.org/10.1521/jscp.2008.27.10.1045

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23.

Belle, A., Ji, S. -Y., Ansari, S., Hakimzadeh, R., Ward, K., & Najarian, K. (2010). Frustration Detection with Electrocardiograph Signal Using Wavelet Transform. In *2010 International Conference on Biosciences* (pp. 91–94). IEEE. 10.1109/BioSciencesWorld.2010.19.

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin, 106*(1), 59.

Bosch, E., Ihme, K., Drewitz, U., Jipp, M., & Oehl, M. (2020). Why drivers are frustrated: Results from a diary study and focus groups. *European Transport Research Review, 12*(1), 52. https://doi.org/10.1186/s12544-020-00441-7

Braun, M., Schubert, J., Pfleging, B., & Alt, F. (2019). Improving driver emotions with affective strategies. *Multimodal Technologies and Interaction, 3*(1), 21.

Breyer, B., & Bluemke, M. (2016). Deutsche version der positive and negative affect schedule PANAS (GESIS panel).

Bruce, V. (1993). What the human face tells the human mind: Some challenges for the robot-human interface. *Advanced Robotics, 8*(4), 341–355. https://doi.org/10.1163/156855394X00149

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale, NJ, USA: Á/L. Erbaum Press.

Cohn, J. F., Schmidt, K., Gross, R., & Ekman, P. (2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*.

Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLOS ONE, 15*(4), e0231968.
Ekman, & Rosenberg. (1997). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA.
Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3–4), 169–200.
Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1*(1), 56–75. https://doi.org/10.1007/bf01115465
Farnsworth, B. (2019). *Facial Action Coding System (FACS): A Visual Guidebook*. https://imotions.com/blog/facial-action-coding-system/.
Ferreri, N. R., & Mayhorn, C. B. (2022). Identifying and understanding individual differences in frustration with technology. *Theoretical Issues in Ergonomics Science, 1–19*. https://doi.org/10.1080/1463922X.2022.2095458
Fischer, M., Richter, A., Schindler, J., Plättner, J., Temme, G., Kelsch, J., … Köster, F. (2014). Modular and scalable driving simulator hardware and software for the development of future driver assistence and automation systems. *New Developments in Driving Simulation Design and Experiments*, 223–229.
Frison, A.-K., Wintersberger, P., & Riener, A. (2019). Resurrecting the ghost in the shell: A need-centered development approach for optimizing user experience in highly automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 65*, 439–456.
Grafsgaard, J., Wiggins, J. B., Boyer, K. E., & Wiebe, E. N. (2013). Automatically recognizing facial expression: Predicting engagement and frustration. *Educational Data Mining*.
Grippenkoven, J., Fassina, Z., König, A., & Dreßler, A. (2018). Perceived safety: A necessary precondition for successful autonomous mobility services. *Proceedings of the Human Factors and Ergonomics Society Europe*.
Gross, J. J. (2008). Emotion and emotion regulation: Personality processes and individual differences. *15938583*.
Harris, H., & Nass, C. (2011). In *ACM* (pp. 749–752). https://doi.org/10.1145/1978942.1979050
Herrenkind, B., Brendel, A. B., Nastjuk, I., Greve, M., & Kolbe, L. M. (2019). Investigating end-user acceptance of autonomous electric buses to accelerate diffusion. *Transportation Research Part D: Transport and Environment, 74*, 255–276. https://doi.org/10.1016/j.trd.2019.08.003
Hoque, M., & Picard, R. W. (2011). Acted vs. *natural frustration and delight: Many people smile in natural frustration*.
Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science, 28*(3), 280–291.
Ihme, K., Dömeland, C., Freese, M., & Jipp, M. (2018). Frustration in the face of the driver: A simulator study on facial muscle activity during frustrated driving. *Interaction Studies, 19*(3), 487–498.
Ihme, K., Unni, A., Zhang, M., Rieger, J. W., & Jipp, M. (2018). Recognizing frustration of drivers from face video recordings and brain activation measurements with functional near-infrared spectroscopy. *Frontiers in Human Neuroscience, 12*, 327.
Jeon, M. (2015). Towards affect-integrated driving behaviour research. *Theoretical Issues in Ergonomics Science, 16*(6), 553–585.
Jeon, M., & Zhang, W. (2013). Sadder but Wiser? Effects of Negative Emotions on Risk Perception, Driving Performance, and Perceived Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 1849–1853. https://doi.org/10.1177/1541931213571413
Kaiser, S., & Wehrle, T. (2001). *The Role of Facial Expression in Intra-individual and Inter-individual Emotion Regulation*. https://www.researchgate.net/publication/2486501_The_Role_of_Facial_Expression_in_Intra-individual_and_Inter-individual_Emotion_Regulation.
Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies, 65*(8), 724–736.
Keller, C., Struwe, S., Titov, W., & Schlegel, T. (2019). Understanding the Usefulness and Acceptance of Adaptivity in Smart Public Transport. In H. Kromker (Ed.), *HCI in Mobility, Transport, and Automotive Systems: First International Conference, MobiTAS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings* (1st ed., pp. 307–326). Springer.
Kim, B. G., & Stein, H. H. (2009). A spreadsheet program for making a balanced Latin square design. *Revista Colombiana De Ciencias Pecuarias, 22*(4), 591–596.
Kleinsmith, A., & Bianchi-Berthouze, N. (2012). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing, 4*(1), 15–33.
Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der" Positive and negative Affect Schedule"(PANAS). *Diagnostica-Gottingen-, 42*, 139–156.
Krüger, S., Bosch, E., Ihme, K., & Oehl, M. (2021). In-Vehicle Frustration Mitigation via Voice-User Interfaces – A Simulator Study. In (pp. 241–248). Springer, Cham. 10.1007/978-3-030-78645-8_30.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior, 32*(3), 241–252. https://doi.org/10.1002/ab.20122
Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist, 46*(8), 819.
Lee, Y.-C. (2010). Measuring Drivers' Frustration in a Driving Simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(19), 1531–1535. https://doi.org/10.1177/154193121005401937
Ma, C., Hao, W., Xiang, W., & Yan, W. (2018). The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents. *Journal of Advanced Transportation*, *2018*.
Malta, L., Miyajima, C., Kitaoka, N., & Takeda, K. (2010). Analysis of real-world driver's frustration. *IEEE Transactions on Intelligent Transportation Systems, 12*(1), 109–118.
Millonig, A., & Fröhlich, P. (2018). Where Autonomous Buses Might and Might Not Bridge the Gaps in the 4 A's of Public Transport Passenger Needs. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 291–297). ACM. 10.1145/3239060.3239079.
Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. In G. van der Veer & C. Gale (Eds.), *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1973–1976). ACM. 10.1145/1056808.1057070.
Noroozi, F., Kaminska, D., Corneanu, C., Sapinski, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE Transactions on Affective Computing*.
Oehl, M., Ihme, K., Drewitz, U., Pape, A.-A., Cornelsen, S., & Schramm, M. (2019). In *ACM* (pp. 260–264). https://doi.org/10.1145/3349263.3351518
Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software*, *6*(61), 3167. 10.21105/joss.03167.
Picard, R. W., & Klein, J. (2002). Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting with Computers, 14*(2), 141–169. https://doi.org/10.1016/s0953-5438(01)00055-8
Rendon-Velez, E., van Leeuwen, P. M., Happee, R., Horváth, I., van der Vegte, W. F., & de Winter, J. C. (2016). The effects of time pressure on driver performance and physiological activity: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 41*, 150–169.
SAE International. (2014). *Automated Driving Levels of Driving Automation are Defined in New SAE International Standard J3016*. MI: SAE International Troy.
Sangineto, E., Zen, G., Ricci, E., & Sebe, N. (2014). We are not All Equal. In K. A. Hua (Ed.), *ACM Digital Library, Proceedings of the 22nd ACM international conference on Multimedia* (pp. 357–366). ACM. 10.1145/2647868.2654916.
Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion, 23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969
Schömig, N., Naujoks, F., Hammer, T., Tomzig, M., Hinterleitner, B., & Mayer, S. (2018). Experimental Induction and Measurement of Negative Affect Induced by Interacting with In-vehicle Information Systems. In M. Kurosu (Ed.), *Lecture Notes in Computer Science. Human-Computer Interaction. Theories, Methods, and Human Issues* (Vol. 10901, pp. 441–452). Springer International Publishing. 10.1007/978-3-319-91238-7_36.
Shinar, D. (1998). Aggressive driving: The contribution of the drivers and the situation. *Transportation Research Part F: Traffic Psychology and Behaviour, 1*(2), 137–160. https://doi.org/10.1016/s1369-8478(99)00002-9
Sidney, K. D., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Graesser, A. C. (Eds.) (2005). *Integrating affect sensors in an intelligent tutoring system*.
Stephan, A. (2015). Empathy for Artificial Agents. *International Journal of Social Robotics, 7*(1), 111–116. https://doi.org/10.1007/s12369-014-0260-0
Venkatesh, Morris, & Davis (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. 10.2307/30036540.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*(6), 879–896.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (Eds.) (2006). *ELAN: a professional framework for multimodality research*.

Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X., & Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. *Transportation Research Part C: Emerging Technologies, 95*, 320–334.

Zepf, S., Stracke, T., Schmitt, A., van de Camp, F., & Beyerer, J. (2019, December 16–19). Towards Real-Time Detection and Mitigation of Driver Frustration using SVM. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 202–209). IEEE. 10.1109/ICMLA.2019.00039.

Zhang, M., Ihme, K., & Drewitz, U. (2019). Discriminating drivers' emotions through the dimension of power: Evidence from facial infrared thermography and peripheral physiological measurements. *Transportation Research Part F: Traffic Psychology and Behaviour, 63*, 135–143.

Zhang, M., Ihme, K., Drewitz, U., & Jipp, M. (2021). Understanding the Multidimensional and Dynamic Nature of Facial Expressions Based on Indicators for Appraisal Components as Basis for Measuring Drivers' Fear. *Frontiers in Psychology, 12*, 622433. https://doi.org/10.3389/fpsyg.2021.622433

Multimodal Estimation of Frustrative Driving Situations Using a Latent Variable Model.

This is a postprint version of the manuscript published as:

Bosch, E., Klosterkamp, M., Guevara, A., Kaethner, D., Bendixen, A., & Ihme, K. (2022, September). Multimodal Estimation of Frustrative Driving Situations Using a Latent Variable Model. In 2022 13th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000011-000016). IEEE.
https://doi.org/10.1109/CogInfoCom55841.2022.10081636.

# Multimodal Estimation of Frustrative Driving Situations Using a Latent Variable Model

Esther Bosch*, Marie Klosterkamp†, Angelo Guevara§, David Kaethner*, Alexandra Bendixen‡, and Klas Ihme*
* Institute of Transportation Systems, German Aerospace Centre, Braunschweig, Germany
Email: esther.bosch@dlr.de
† Institute of Transportation, University of Kassel, Kassel, Germany
‡ Institute of Physics, Chemnitz University of Technology, Chemnitz, Germany
§ Department of Civil Engineering, Universidad de Chile and Instituto Sistemas Complejos de Ingeniería, Santiago, Chile

***Abstract*—Within recent years, it has become popular to use physiological and expression data to ameliorate inter-cognitive communication between human and machine. One emotion that is highly relevant for this is frustration, which occurs when a user's goal in using a system is failed to be met. This paper presents a latent variable model that estimates frustration in two different driving contexts by continuous subjective frustration rating, facial expressions and frontal alpha asymmetry in the electroencephalogram. We then compare this full model to models with less measurement variables to evaluate which measurements can be left out. Our results show that expression frequency and subjective frustration make important contributions to the model of experienced frustration. This paper presents a proof of concept for using a latent variable model to evaluate collected measures to estimate an experienced emotion. This method can inform researchers which measurements are most informative in different circumstances. Additionally, the method can be used to evaluate how well purely objective measurements (that are the only feasible measurements in most applied settings) perform in comparison to a model including subjective ratings.**

## I. Introduction

One goal of Cognitive Information Communication is to enable an efficient cooperation between artificial and natural cognitive systems [1]. Frustration is an emotion that is highly relevant in the mode of inter-cognitive communication (i.e. between a human and an artificially cognitive system), when 1) miscommunication can lead to frustration with the interaction and 2) the artificial system can simulate empathy and by that improve the cooperation. Generally, frustration occurs when a goal is failed to be met [2]. It is characterized by low valence, low dominance and high arousal by the valence arousal dominance model [3], [4], [5]. In contexts in which inter-cognitive communication occurs, frustration can be mitigated through its real-time detection and subsequent help that enables success at the given task. One of these contexts is transportation, in which different triggers of frustration occur [6] and negative emotions, such as frustration, negatively affect cognitive skills used in driving [7]. These numerous triggers of frustration and the link between frustration and aggressive behavior [8], such as speeding or risky overtaking maneuvers, suggest that frustration contributes to the number of aggression-related accidents, some of which are fatal. In future mobility systems, where the driver's skills may not be of importance anymore due to automated driving systems, a successful and frustration-free interaction with the human-machine interface becomes increasingly important [9]. The technology acceptance model states that usefulness and ease of use are the two most important predictors of user technology acceptance [10]. Therefore, frustration that is experienced when using future transportation can lead to a diminished acceptance for the system as a whole and, by that, against a decision to use this transportation mode. Especially due to the current need for change towards sustainable transportation modes, it is of interest to motivate the decision to use specific mobility solutions [11]. One way to facilitate this is to use in-vehicle frustration recognition and mitigation integrated into sustainable mobility concepts, like for example shared mobility.

Previous research has aimed to recognize frustration based on facial expressions that were captured by cameras. It was found that Brow Lowerer, Dimpler, Brow Raiser, Smile and Lip Press are often shown in frustration [12], [13], [14]. Another possible source for the assessment of frustration is neurophysiological data [15], [16]. A frequently used method for inferring frustration from electroencephalogram (EEG) is the calculation of the frontal Alpha Asymmetry Index (AAI) [17]. The AAI is based on the difference in the activation measured at frontal electrodes, most commonly F3 and F4, thereby comparing activity over the left and the right hemisphere of the brain [17]. Numerous studies have used the AAI as an indicator for emotion-related state and trait measures, analyzing mood inductions, alterations, and dispositional mood (e.g. [17], [18]). Regarding the recognition of emotion by EEG data, [19], [16] describe that low dominance is characterized by relative left frontal alpha band activation (negative AAI values when AAI = $F4-F3/F3+F4$) and [20] showed that also negative valence is characterized by relative left frontal alpha band activation (negative AAI values). According to these results, frustration as an emotion with low valence and low dominance would be characterized by negative AAI values.

The component process model of emotion [21] states that an emotion consists of an event's appraisal and a subsequent appropriate physiological response. Accordingly, previous studies have measured subjective as well as physiological data to learn what physiological patterns look like for various subjectively rated emotions. However, as both the subjective appraisal as well as the physiological response are part of the emotion, it is more accurate to estimate the actually experienced emotion in a model that uses both subjective and physiological data as estimation variables. This is possible with the use of a latent variable model as described in [22] and [23].

Latent variable models describe variables that cannot be observed directly. As explained in [24], the latent variable can be described by a structural equation: $x^* = h(x; \beta^s) + \epsilon^s$, where h is a (often linear) function, $x$ is a vector of explanatory variables, $\beta^s$ is a vector of parameters (to be estimated from data) and $\epsilon^s$ is the (random) error term. Information about latent variables is obtained from indirect measurements that are manifestations of the underlying latent variable. The relationship between a latent variable and measurements is described by measurement equations with the form $z = m(x^*, y; \beta^m) + \epsilon^m$, where $z$ is the reported value, $x^*$ is the latent variable, $y$ is a vector of observed explanatory variables, $\beta^m$ is a vector of $K_m$ parameters (to be estimated from data) and $\epsilon^m$ is the (random) error term.

Latent variable models have previously been used to combine driving simulator and physiological data to model how stress changes car-following behavior [25] and to model how a driver's cognitive effort impacts route choice decisions [26]. However, it is new to use them in the context of assessing a situation's frustrativeness based on multimodal data. Such latent variable models offer a great way to estimate experienced frustration by subjective as well as objective measures. To evaluate which measures are needed, the comparison to a full model that considers objective as well as subjective measurements is proposed in this paper. We will calculate a full model that uses a subjective frustration rating, facial expressions and frontal alpha asymmetry in the EEG. We will then compare the full model with models that each only contain two of the measurement variables to see if any of the measurements could be dropped. As context, we use two reasons for frustration that have been described in [6]: traffic-related causes and Human-Machine-Interface (HMI)-related causes for frustration.

## II. Methods

### A. Data Collection

*1) Experimental Design:* Each participant experienced six drives as described in [27] in a 2 (context: 3 HMI task vs. 3 driving task) x 2 (frustration induction: 2 frustration vs. 1 baseline drive) within-subject design in a driving simulator. Three of the drives were driven in manual driving mode (driving task context) and three in automated driving mode (HMI task context). Both driving modes had one baseline drive and two frustration-inducing drives each. The drives' order for each participant was determined by a balanced Latin square design.

*2) Participants:* Fourteen participants were recruited through the institute's participant pool. Of these, one participant was excluded from data analyses because of a lack of compatibility of the electrode placement with standard electrode locations due to a mismatch between head size and EEG system. Another was excluded post-hoc because the subjective frustration rating indicated that the induction of frustration had not been successful, as the subjective frustration ratings were zero for the whole duration of all drives. Of the N = 12 participants included in the analyses three were female and 9 male. Participants' age ranged from 21 to 59 years (M= 33.12, SD= 13.96). Participants were informed about all data recordings, potential risks of driving in simulators (e.g., the experience of simulator sickness) and the duration of the experiment. Participants could take a break or abort their participation at any time. All participants gave written informed consent to take part in the study. As reimbursement for their time, the participants received 5 € per commenced half hour. After the study, the true goal of the experiment (evoking frustration) was revealed and the necessity to conceal this goal with a cover story was explained. The collected data were processed according to European General Data Protection Regulations.

*3) Experimental Set-Up:* The data set was recorded in a 360-degree full-view driving simulator [28]. The participants sat in a vehicle mock-up and could use a conventional interface with throttle, brake pedal, steering wheel, and indicators to drive the mock-up car in the driving simulation (Virtual Test Drive, Vires Simulationstechnologie, Bad Aibling, Germany). On a tablet (Microsoft Surface Pro 7, 12.3') mounted to the car's center console, a user interface (UI) was shown (required for the frustration induction).

*4) Context:* To enable two different task contexts, we collected data in two different driving modes (manual vs. automated). In the manual driving mode ('driving task context'), participants were told to assume they were supposed to meet friends at a movie theater. The participants were informed that the average travel time to the movie theater was less than ten minutes. They were told that, if they arrived at their location on time, they would receive a 2 € prize. The time remaining for punctual arrival at the movie theater was displayed on a clearly visible clock. In the frustration-inducing drives, a timely reaching of the movie theatre was impeded by two subsequent slow trucks that could not be overtaken and a long waiting time at a red light. In the baseline drives, participants could easily arrive at the movie theatre on time, as there was little traffic. In earlier investigations, tasks comparable to the ones used in this study have been demonstrated to successfully elicit frustration [14], [15], [29]. In the automated driving mode ('HMI task context'), participants completed a task (joining an online conference or changing the destination) on the in-car UI shown on the tablet. Meanwhile, the car drove fully automated on a highway. The participants were told that if they completed their work successfully, they would get a reward of 2 €. Before the start of the experiment, all participants read the same story in all three automated drives. They were asked to imagine to drive to a business meeting in an autonomous car. In the frustration-inducing drives, completing the task on the user interface was very difficult to do in seven minutes due to vague button names, imprecise iconography, and confusing click-paths. In the baseline drive, the user interface was intuitive to use. Participants were then instructed to push a single button that appeared in various locations of the UI. They were assured not to be under any time constraints and asked to interact with the UI as relaxed as possible. In each of the modes (manual and automated), participants experienced one baseline drive and two frustration-induction drives. In the manual driving mode, we varied the car types and the track between the two frustration scenarios to disguise the fact that the same driving scenario was driven twice. Frustrating incidents were the same in both driving scenarios and took the same amount of time.

*5) Procedure:* On arrival, participants filled out an informed consent form as well as a data privacy declaration. The researcher told the cover story that the study was analyzing changes in gaze behavior between manual and automated driving modes. This was done to hide the true purpose of frustration induction and allow spontaneous emotion emergence. To lessen the impact of unfamiliarity, all participants practiced manual and automated driving modes before the start of the experiment until they were comfortable with the simulator and the driving conditions. Following the six drives, the participants were told of the study's true purpose. Then, they provided a continuous post-hoc frustration rating for each of the six drives. The entire process took 2 hours on average.

### B. Measurements

*Frontal Alpha Asymmetry*: The "CGX quick-30" mobile EEG system from CGX with a 29-channel layout was utilized to capture the participants' brain activity. The channel locations adhere to the extended international 10-20 layout. Processing of the EEG data was accomplished by the use of Matlab 2019b. After re-referencing to linked earlobes, two different filter designs were employed during the processing of the EEG data, one for the ICA set and the other for the analysis data set. The high-pass filter cut-off frequency for the ICA set was specified as 1.25 Hz with a 0.5 Hz transition bandwidth following [30]. For the analysis data set a 0.3 Hz cut-off frequency was deployed for the high-pass filter with a 0.2 Hz transition bandwidth as well as a lowpass filter with a 50 Hz cut-off frequency and a 10 Hz transition bandwidth following the filter settings of previous research [16], [19]. For both sets, a zero-phase Hamming window sinc FIR filter with a maximal passband deviation of 0.0022 (2 ‰) and a 53 dB (decibel) stopband attenuation was realized. We removed bad data from artifacts of individual channels due to for example the displacement of electrodes or inadequate scalp contact. Bad channels were removed without the additional removal of bad data segments. Channels were rejected if they were flat for more than five seconds, they correlated less than 80 % with the neighboring channels or the high-frequency noise passed a threshold of four standard deviations of the channels' activity. After the channel rejection, the data was epoched into one-second segments for further artifact removal processes as well as for the computation of the channel spectra. Epochs with a low signal-to-noise ratio were rejected when the channels' probability of activity or the kurtosis of the data was higher than the threshold of five standard deviations. In addition, a delta criterion of ± 250 µV amplitude deviation of the channels within an epoch was used to exclude noisy data from analysis. In addition to the rejection of artifacts within epochs or channels, an ICA was realized to remove artifacts such as muscle activity, eye movements, and blinks. The manual classification was performed in unison by two researchers coming into agreement about the rejection of artifact-laden components. The resulting rejection matrix was then transferred and applied to the analysis data set. The analysis was conducted within the frequency domain of the EEG data, for which a fast Fourier transformation was used to decompose the EEG data into frequencies for a fixed window size of one second, for which length of data points was zero-padded to the next power of two (512 pts), resulting in a periodogram with a 1 Hz resolution. The absolute power (in $\mu V^2$) of the frequencies was then approximated using the composite Simpson´s rule. For the analysis, the absolute power within each frequency was then divided by the total power of the spectrum (2 - 45 Hz) resulting in the relative power, which was transformed to a logarithmic scale (natural logarithm) to achieve a normal distribution for the resulting measure, following the recommendation of [17]. For the scores of the AAI, the relative spectral power for the alpha band (8 – 13 Hz) was calculated according to the process described above. The resulting measure at electrode F3 was then subtracted from the corresponding measure at F4 within the same epoch and divided by the overall activity of both electrodes ($F4 - F3/F3 + F4$).

*Subjective Frustration Rating*: In order to acquire a time-resolved indication of frustration, a continuous subjective assessment was collected when a participant had completed all drives. The participants assessed their frustration using a joystick on a scale of 0 to 100 percent while watching the videos that were recorded during all drives of the whole scene (the participant's face was not visible). When not touched, the joystick could only go in one direction and immediately returned to zero. The participants received a visual feedback of their current rating next to the video. They were instructed to move the joystick according to their level of frustration in the circumstances depicted in the video. This allowed for the collection of a continuous frustration rating for each drive and each participant. For the latent variable model, to match the EEG data, data points of the subjective frustration rating where all measurements were available were binned into one second bins.

*Frustration expressions*: All facial expressions were annotated by three trained annotators. A majority vote between these three annotations decided on the final annotation. Facial expressions previously described as typical for frustration (Brow Lowerer, Dimpler, Brow Raiser, Smile and Lip Press) [12], [13], [14] were extracted and their combined expressions frequency calculated per drive as measured in expression per minute.

To get an insight into the data and compare differences between conditions, the complete data used for the latent variable model was summarized as described in the following to calculate separate tests. Means were calculated per drive. Afterwards, the mean of both frustration induction drives was taken, so that every participant has four values per measure in total (one for frustration induction, one for baseline, one for HMI task context and one for driving task context). This data was checked for normality by Shapiro Wilk test and subsequently a paired t-test or a Wilcoxon signed rank test was performed to compare, per measure, differences between the two frustration conditions and differences between the two contexts. We decided for testing only main effects for the purpose of getting an insight into the data and accordingly adjusted all significance levels by Bonferroni-correction for multiple tests.

### C. Latent Variable Model

Figure 1 describes the relation between the latent variable, the experimental attributes and the indicators. The latent variable is depicted in an oval, to remark that it is not observed by the researcher, while attributes and indicators, which are observed, are depicted in rectangles. The arrows correspond to causal relations that are accounted for by structural or measurement equations.

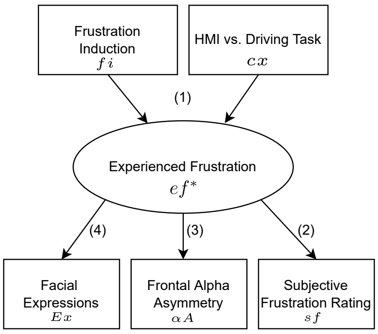


Fig. 1. Latent variable model between experienced frustration, experimental condition and measurement variables. Numbers in brackets refer to structural and measurement equations.

At each timepoint $t$ of the experiment, an individual $n$ experiences a level of frustration induction $fi$ in a context $cx$ consisting of a HMI or Driving task. Besides, measures of alpha asymmetry of the two frontal electrodes ($F4-F3/F4+F3$), facial expression frequency $Ex$ and subjective frustration $sf$ are recorded.

The structural equation 1 relates the latent variable experienced frustration $ef^*$, corresponding to the subjectively experienced frustration, with the attributes of treatment $fi$ and context $cx$. The $\beta^k$ are the parameters of interest to be estimated from data and $\epsilon^{ef}$ is an exogenous error term representing other factors affecting $ef^*$.

$$ef^*_{nt} = \beta^0 + \beta^{fi} fi_{nt} + \beta^{cx} cx_{nt} + \beta^{fi*cx} fi_{nt} * cx_{nt} + \epsilon^{ef}_{nt} \quad (1)$$

There are three measurement equations for the latent variable experienced frustration $ef^*$. The first comes the from subjective frustration $sf$ that the participants have been asked to report after all drives. The other measurement equation is obtained from the measured alpha asymmetry $\alpha A$ of the two frontal electrodes, $F4-F3/F4+F3$. The third measurement equation comes from the per-drive frequency of facial expressions indicative for frustration, $Ex$.

The measurement equations, in this case are specified as a linear equation, therefore take the form

$$sf_{nt} = \gamma_0^{sf} + \gamma_1^{sf} ef^*_{nt} + \epsilon^{sf} \quad (2)$$

$$\alpha A_{nt} = \gamma_0^{\alpha A} + \gamma_1^{\alpha A} ef^*_{nt} + \epsilon^{\alpha A} \quad (3)$$

$$Ex_{nt} = \gamma_0^{Ex} + \gamma_1^{Ex} ef^*_{nt} + \epsilon^{Ex} \quad (4)$$

where $\gamma^k$ is a vector of $k_m$ parameters (to be estimated from data) and $\epsilon^k$ is an exogenous (random) error term. All data was z-transformed before the models were calculated. The full model as shown in Figure 1 was compared against models that each only contain two of the three measurement variables. For this, we calculated the mean standard error per model and compared them to one another. For calculation of the latent variable model, due to the z-transformation of the data, $\hat{\gamma}_0^{sf}$, $\hat{\gamma}_0^{Ex}$ and $\hat{\gamma}_0^{\alpha A}$ were set to 0 and $\epsilon^{sf}$, $\epsilon^{Ex}$ and $\epsilon^{\alpha A}$ were set to 1.

TABLE I. PAIRWISE COMPARISONS OF MEASURED VARIABLES. $n_{pairs} = 12$ FOR ALL MEASUREMENTS.

| | Mean | Median | Standard Deviation | Wilcoxon Signed Rank (V) or Paired T-Test (t) |
|---|---|---|---|---|
| Subjective Rating Frustration vs. Baseline | 0.17 vs. 0.03 | 0.12 vs. 0.01 | 0.15 vs. 0.05 | $V = 18, p < 0.01$, $r = -0.84$ |
| Subjective Rating HMI vs. Driving | 0.13 vs. 0.07 | 0.06 vs. 0.02 | 0.16 vs. 0.09 | $V = 160, p = 0.08$, r = -0.26 |
| Frustration Expressions Frustration vs. Baseline | 1.14 vs. 0.83 | 0.98 vs. 0.61 | 0.68 vs. 0.71 | $V = 82, p < 0.05$, r = -0.33 |
| Frustration Expressions HMI vs. Driving | 1.16 vs. 0.87 | 0.88 vs. 0.85 | 0.89 vs. 0.48 | $V = 120, p = 0.1$, r = -0.25 |
| Frontal AAI Frustration vs. Baseline | 0.004 vs. 0.002 | 0.004 vs. 0.002 | 0.01 vs. 0.02 | $t(12) = -1.03$, p = 0.31, r = 0.18 |
| Frontal AAI HMI vs. Driving | 0.003 vs. 0.003 | 0.02 vs. 0.006 | 0.02 vs. 0.02 | $t(12) = 0.04$, p = 0.97, r = 0.006 |

## III. RESULTS

Figure 2 shows the data for subjective frustration rating, frustration expressions and frontal AAI by frustration induction and by context. The pairwise comparisons for all measured variables are shown in Table I. The subjective frustration rating is significantly higher in the frustrating drives than the baseline drives and is higher in the HMI task context than the driving task context. The frustration expressions were shown more frequently in frustrating drives than the baseline drives and were shown more frequently in the HMI task context than the driving task context. The frontal alpha asymmetry was similar in the baseline drives and the frustrating drives and very similar in the HMI task context and the driving task context (see Table I).

TABLE II. ESTIMATED FULL MODEL.

| | Estimate | Std. error | t value | Pr(> t) |
|---|---|---|---|---|
| $\hat{\beta}^0$ | -31.31 | 13.43 | -2.33 | 0.02 |
| $\hat{\beta}^{fi}$ | 29.76 | 12.78 | 2.33 | 0.02 |
| $\hat{\beta}^{cx}$ | 4.61 | 2.91 | 1.58 | 0.11 |
| $\hat{\beta}^{fi*cx}$ | 25.73 | 11.18 | 2.30 | 0.02 |
| $\hat{\gamma}_1^{sf}$ | 0.02 | 0.01 | 2.35 | 0.02 |
| $\hat{\gamma}_1^{\alpha A}$ | 0.00 | 0.00 | 0.25 | 0.80 |
| $\hat{\gamma}_1^{Ex}$ | 0.02 | 0.01 | 2.35 | 0.02 |

The estimation results of the full latent variable model are shown in Table II. Frustration induction $fi$ had a significant influence on experienced frustration $ef$ (see $\hat{\beta}^{fi}$ and $\hat{\beta}^{cx}$ in Table II). Also the interaction coefficient $\hat{\beta}^{fi*cx}$ is significant. Due to the z-transformation of the input data, we can interpret the estimated coefficients as the weights of the respective indicators. As the coefficients of subjective frustration ($\hat{\gamma}_1^{sf}$) and expression frequency ($\hat{\gamma}_1^{Ex}$) are significant and near equal, this shows that the latent variable has an important and comparable impact on both measurements. However, the same cannot be said about the frontal AAI, as $\hat{\gamma}_1^{\alpha A}$ is not significant. These results remain the same when normalizing the model to

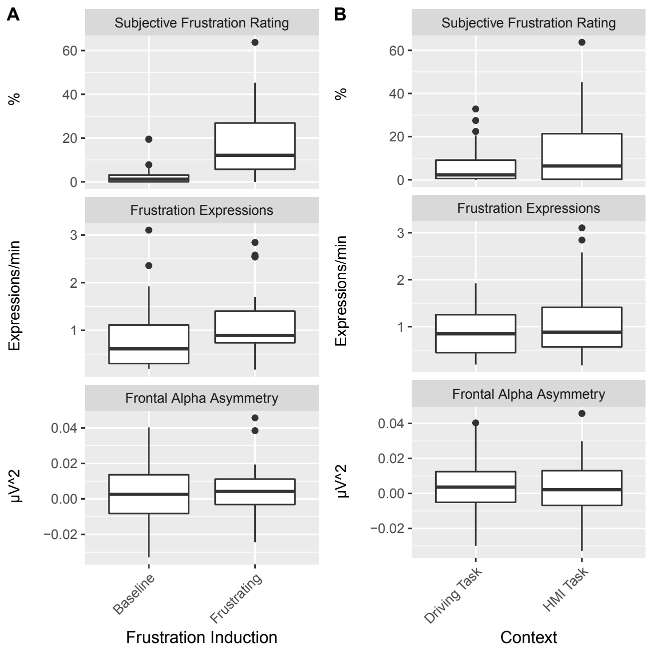


Fig. 2. Measurement Variables by Frustration Induction (A) and Task (B).

TABLE III. COEFFICIENTS AND MEAN STANDARD ERRORS OF THE LATENT VARIABLE MODELS.

| | Full Model ($sf$, $\alpha A$, $Ex$) | Model 2 ($\alpha A$, $sf$) | Model 3 ($sf$, $Ex$) | Model 4 ($\alpha A$, $Ex$) |
|---|---|---|---|---|
| Frustrating Tablet Task | 28.8 | 14.8 | 14 | 15.9 |
| Baseline Tablet Task | -26.7 | -11 | -13.1 | -18.2 |
| Frustrating Driving Task | -1.56 | -2.34 | -0.78 | 1.32 |
| Baseline Driving Task | -31.3 | -12.6 | -15.3 | -21.8 |
| Mean Standard Error | 5.76 | 2.71 | 1.74 | 3.87 |

expression frequency instead of subjectively rated frustration. The estimated latent variable results per model can be seen in Table III. In all models, frustration condition with the HMI task leads to the highest experienced frustration, frustration condition with the driving task leads to second-highest experienced frustration, and both baseline conditions lead to lowest frustration. The full model has the highest mean standard error (5.76), the model only containing subjective frustration and expression frequency has the lowest means standard error (1.74).

## IV. Discussion

The aim of this work was to evaluate different contexts' frustrativeness by a latent variable of experienced emotion by multimodal data and to compare which measurements are needed to achieve a model with the lowest standard errors. Specifically, we estimated how frustrating two different contexts were experienced by integrating subjective, EEG and facial expression data into a latent variable model. The presented latent variable model approximates experienced frustration based on subjective as well as objective data, which is chosen based on the component process model of emotion [21], in which an experienced emotion consists of an interplay between subjectively felt and physiologically experienced emotion. We compared the full model to models with less measurement variables to test which measurements could be left out to achieve similar results. In our case, the subjective rating and the expression frequency both seem to be informative measurements. This is in line with the significance of the measurements' coefficients of the full model. Leaving out the alpha asymmetry, however, seemed to improve the fit of the model. Prospectively, this method of model comparisons that includes objective and subjective data could help researchers to evaluate which measurements are needed in which situation. This could differ depending on the exact circumstances: for example, facial expressions might be shown more often when other people are around than when a person is alone [31], [32]. Ultimately, the goal in an applied context is a model that only depends on objective measurement variables, but gives similar estimation results as a model that also includes subjective measures, as subjective measures are still the most commonly used ground truth measure nowadays, but are impractical to be measured in most applied contexts of emotion recognition. As the estimated latent variable coefficients are similar in all models (also the model only including objective data), we can show that we can approximate the experienced emotion based on objective data only, with a quality close to that of the model that contains subjective data.

Of the measurement variables, subjective frustration rating and expression frequency behaved as expected: the subjective frustration rating was higher in the frustrating drives with a large effect size according to [33] and frustration expressions were shown more frequently in the frustrating drives, also with a medium effect size. In contrast, the frontal AAI was very similar in frustrating and baseline drives. This little difference in frontal AAI could be explained by the fact that all data of the complete drives were included in the analysis and frontal AAI is influenced most likely not only by experienced frustration during that time. For example, [34] find that frontal AAI is changed by reappraisal processes, which most likely happened in several moments of the drives.

According to the results of our model, frustration caused by the unclear user interface was more frustrating than the frustration caused by the traffic situations in our experiment. These results are in line with results found by [6], in which participants rated higher frustration for HMI- than for traffic-related causes. The addition of contextual information combined with this knowledge of how frustrating a certain situation typically is experienced might help future endeavors of recognizing frustration [35].

## V. Conclusion

This paper presented how latent variable models can be used to assess which behavioral and psychophysiological measures are feasible to replace the need for a subjective ground truth in future real-world applications of frustration recognition. This can be relevant in order to ensure a successful intercognitive communication between a vehicle and a traveler. On the long term, when integrated into sustainable (for example shared) mobility concepts, this improvement could lead to a more likely decision for sustainable mobility behavior.

## Acknowledgment

This work was partly supported by the project F-RELACS which received funding from the German Federal Ministry of Education and Science in the program KMU-innovativ under the grant 16SV7930. The third author was partially funded by ANID FONDECYT 1191104 and ANID PIA/BASAL AFB180003. We thank Helena Wasle for comments on the final manuscript.

## References

[1] P. Baranyi and Á. Csapó, "Definition and synergies of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[2] A. M. Colman, *A dictionary of psychology*. Oxford University Press, USA, 2015.

[3] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "Eeg-based bci emotion recognition: A survey," *Sensors*, vol. 20, no. 18, p. 5083, 2020.

[4] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[5] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of research in Personality*, vol. 11, no. 3, pp. 273–294, 1977.

[6] E. Bosch, K. Ihme, U. Drewitz, M. Jipp, and M. Oehl, "Why drivers are frustrated: results from a diary study and focus groups," *European transport research review*, vol. 12, no. 1, pp. 1–13, 2020.

[7] M. Jeon and W. Zhang, "Sadder but wiser? effects of negative emotions on risk perception, driving performance, and perceived workload," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2013, pp. 1849–1853.

[8] L. Berkowitz, "Frustration-aggression hypothesis: examination and reformulation." *Psychological bulletin*, vol. 106, no. 1, p. 59, 1989.

[9] M. Zöldy and P. Baranyi, "Cognitive mobility–cogmob," in *12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2021): Proceedings IEEE*, 2021, pp. 921–925.

[10] F. D. Davis, "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts," *International journal of man-machine studies*, vol. 38, no. 3, pp. 475–487, 1993.

[11] J. Majerova, "Cognitive rationality and sustainable decision based on maslow's theorem: A case study in slovakia," *Cognitive Sustainability*, vol. 1, no. 1, 2022.

[12] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Automatically recognizing facial indicators of frustration: a learning-centric analysis," in *2013 humaine association conference on affective computing and intelligent interaction*. IEEE, 2013, pp. 159–165.

[13] K. Ihme, C. Dömeland, M. Freese, and M. Jipp, "Frustration in the face of the driver: a simulator study on facial muscle activity during frustrated driving," *Interaction Studies*, vol. 19, no. 3, pp. 487–498, 2018.

[14] M. Hoque and R. W. Picard, "Acted vs. natural frustration and delight: Many people smile in natural frustration," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 354–359.

[15] K. Ihme, A. Unni, M. Zhang, J. W. Rieger, and M. Jipp, "Recognizing frustration of drivers from face video recordings and brain activation measurements with functional near-infrared spectroscopy," *Frontiers in human neuroscience*, vol. 12, p. 327, 2018.

[16] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the eeg during game play," *International journal of autonomous and adaptive communications systems*, vol. 6, no. 1, pp. 45–62, 2013.

[17] E. E. Smith, S. J. Reznik, J. L. Stewart, and J. J. Allen, "Assessing and conceptualizing frontal eeg asymmetry: An updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry," *International Journal of Psychophysiology*, vol. 111, pp. 98–114, 2017.

[18] M. Palmiero and L. Piccardi, "Frontal eeg asymmetry of mood: A mini-review," *Frontiers in Behavioral Neuroscience*, vol. 11, p. 224, 2017.

[19] T. J. Schuster, "Elektroenzephalographische erfassung des emotionalen nutzerzustands in der simulierten mensch-maschine interaktion," Ph.D. dissertation, Universität Ulm, 2014.

[20] D. Huang, C. Guan, K. K. Ang, H. Zhang, and Y. Pan, "Asymmetric spatial pattern for eeg-based emotion detection," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2012, pp. 1–7.

[21] K. R. Scherer, "The dynamic architecture of emotion: Evidence for the component process model," *Cognition and emotion*, vol. 23, no. 7, pp. 1307–1351, 2009.

[22] M. Ben-Akiva, J. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa, and A. Polydoropoulou, "Integration of choice and latent variable models," *Perpetual motion: Travel behaviour research opportunities and application challenges*, vol. 2002, pp. 431–470, 2002.

[23] M. Castro, C. A. Guevara, and A. Jimenez-Molina, "A methodological framework to incorporate psychophysiological indicators into transportation modeling," *Transportation research part C: emerging technologies*, vol. 118, p. 102712, 2020.

[24] M. Bierlaire, "Estimating choice models with latent variables with pandasbiogeme," *Rep. TRANSP-OR*, vol. 181227, 2018.

[25] E. Paschalidis, C. F. Choudhury, and S. Hess, "Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour," *Analytic methods in accident research*, vol. 22, p. 100089, 2019.

[26] S. Agrawal and S. Peeta, "Hybrid route choice model incorporating latent cognitive effects of real-time travel information using physiological data," *Transportation research part F: traffic psychology and behaviour*, vol. 81, pp. 223–239, 2021.

[27] E. Bosch, D. Käthner, M. Jipp, U. Drewitz, and K. Ihme, "Fifty shades of frustration: Intra- and interindividual variances in expressing frustration," *Transportation Research Part F: Traffic Psychology and Behaviour*, under review.

[28] M. Fischer, A. Richter, J. Schindler, J. Plättner, G. Temme, J. Kelsch, D. Assmann, and F. Köster, "Modular and scalable driving simulator hardware and software for the development of future driver assistence and automation systems," *New Developments in Driving Simulation Design and Experiments*, pp. 223–229, 2014.

[29] L. Malta, C. Miyajima, N. Kitaoka, and K. Takeda, "Analysis of real-world driver's frustration," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 109–118, 2010.

[30] M. Klug and K. Gramann, "Identifying key factors for improving ica-based decomposition of eeg data in mobile and stationary experiments," *European Journal of Neuroscience*, vol. 54, no. 12, pp. 8406–8420, 2021.

[31] R. E. Kraut and R. E. Johnston, "Social and emotional messages of smiling: an ethological approach." *Journal of personality and social psychology*, vol. 37, no. 9, p. 1539, 1979.

[32] M.-A. Ruiz-Belda, J.-M. Fernández-Dols, P. Carrera, and K. Barchard, "Spontaneous facial expressions of happy bowlers and soccer fans," *Cognition and Emotion*, vol. 17, no. 2, pp. 315–326, 2003.

[33] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[34] D. Choi, T. Sekiya, N. Minote, and S. Watanuki, "Relative left frontal activity in reappraisal and suppression of negative emotion: Evidence from frontal alpha asymmetry (faa)," *International Journal of Psychophysiology*, vol. 109, pp. 37–44, 2016.

[35] D. Bethge, T. Kosch, T. Grosse-Puppendahl, L. L. Chuang, M. Kari, A. Jagaciak, and A. Schmidt, "Vemotion: Using driving context for indirect emotion prediction in real-time," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 638–651.

# 3

# Research Question 2: What contexts lead to in-vehicle frustration?

Why drivers are frustrated: results from a diary study and focus groups.

This is a published version of the manuscript:

Bosch, E., Ihme, K., Drewitz, U. et al. Why drivers are frustrated: results from a diary study and focus groups. Eur. Transp. Res. Rev. 12, 52 (2020).
https://doi.org/10.1186/s12544-020-00441-7. Postprint.

Bosch *et al. European Transport Research Review* (2020) 12:52
https://doi.org/10.1186/s12544-020-00441-7

European Transport Research Review

ORIGINAL PAPER Open Access

# Why drivers are frustrated: results from a diary study and focus groups

Esther Bosch*, Klas Ihme, Uwe Drewitz, Meike Jipp and Michael Oehl

## Abstract

**Introduction:** Designing emotion-aware systems has become a manageable aim through recent developments in computer vision and machine learning. In the context of driver behaviour, especially negative emotions like frustration have shifted into the focus of major car manufacturers. Recognition and mitigation of the same could lead to safer roads in manual and more comfort in automated driving. While frustration recognition and also general mitigation methods have been previously researched, the knowledge of reasons for frustration is necessary to offer targeted solutions for frustration mitigation. However, up to the present day, systematic investigations about reasons for frustration behind the wheel are lacking.

**Methods:** Therefore, in this work a combination of diary study and user focus groups was employed to shed light on reasons why humans become frustrated during driving. In addition, participants of the focus groups were asked for their usual coping methods with frustrating situations.

**Results:** It was revealed that the main reasons for frustration in driving are related to traffic, in-car reasons, self-inflicted causes, and weather. Coping strategies that drivers use in everyday life include cursing, distraction by media and thinking about something else, amongst others. This knowledge will help to design a frustration-aware system that monitors the driver's environment according to the spectrum of frustration causes found in the research presented here.

**Keywords:** Emotion-aware systems, Frustration, Emotions in driving

## 1 Introduction

When watching car advertisements, we see empty roads, happy faces, relaxed people, and children enjoying their rides. When comparing this to what we experience every day on the road, reality looks different: We get up too late, rush to work, and get annoyed about slow vehicles in front of us while crying children on the back seat take our last hope of a relaxed start into the day. This is one out of endless examples that may result in an emotion that can crucially influence driver's focus of attention and also well-being: frustration. Frustration is defined as the emotion that occurs when a goal is blocked to be reached [5, 15]. It is to be differentiated from anger, which is directed towards someone that is responsible for an undesirable event [28]. As stated in the frustration-aggression-hypothesis, frustration is the emotion that often precedes anger and aggression (cf. [3, 5, 21]). Previous research clearly shows the impact that frustration has on driver behaviour and attention. However, to the best of our knowledge, no study systematically has investigated frustrating events during driving and their relative amount of occurrence yet. Hence, this work aims to shed light on the spectrum of reasons for which drivers become frustrated, based on subjective reports of the same. In order to study this, the two complementary methods of a diary study (showing how often which frustrating situation occurs within a given time) and focus groups (showing which frustrating events mainly stay in memory) were employed. In the focus group study, we additionally investigated user's daily coping strategies with frustrating situations on the road.

* Correspondence: esther.bosch@dlr.de
German Aerospace Center (DLR), Institute of Transportation Systems, Lilienthalplatz 7, 38108 Braunschweig, Germany

## 2 Previous work

### 2.1 Frustration during driving

Frustration in driving is critical in manual as well as automated driving. In manual driving, frustrated drivers were less aware of potential distractions, their mental state, and potential dangers in the driving environment [16]. Earlier research has shown that frustrated drivers exhibit more aggressive driving styles [4, 12, 25]. In automated driving, frustration might turn out as a central challenge: every user of modern complex interfaces knows how frustrating it can be to try to use an interface that does not behave as expected [14]. Therefore, even during automated driving frustration can occur and may decline acceptance and comfort. In conclusion, frustration is an emotion that can strongly affect road safety, user experience and comfort in manual and automated driving.

### 2.2 Frustration-aware systems

The concept of designing frustration-aware systems has emerged due to the above mentioned effects of frustration on user experience and road safety [18, 22]. The aim of such a system would be recognition of frustration and successful mitigation of the same. Three main steps are necessary to design such a system: 1) recognizing that a driver or passenger is frustrated, e.g. by means of physiological measurements or video recordings, 2) detecting the reason for frustration and 3) offering help that is tailored towards the specific situation. Several researchers investigated the first step of recognizing frustration [1, 11, 19] and the third step of mitigating frustration (e.g., [6, 13]), but the intermediate step of detecting its reasons remains elusive [17]. In addition, knowledge of these environmental factors can crucially improve the first step of detecting frustration [19]. As a prerequisite for this step, knowledge of the frustration-inducing events' spectrum is required. Therefore, on the way towards designing frustration aware-systems it is a necessary step to gain insight into reasons for which people get frustrated in the vehicle. With increasing automation accompanied by developments in driver monitoring, the extent of available sensor technology in modern vehicles is growing. This is an advantage for recognizing not only frustration itself, but also reasons for the same – inside and outside the vehicle. By understanding the spectrum of reasons for frustration, a frustration-aware system can be equipped with the knowledge of which information is relevant to scan in order to recognize sources of frustration.

### 2.3 Coping strategies

The last step of designing a frustration-aware system is to offer help that is specifically tailored towards the situation at hand. To do so in a user-centred way, investigation of user's everyday coping strategies with frustrating in-car situations is of interest. Various strategies for coping with negative emotions like frustration have been proposed previously [7, 20]. One example is Gross [9], who suggests to differentiate coping strategies into the categories of attentional deployment, response modulation, cognitive change, situation modification and situation selection. Situation selection is described as 'approaching or avoiding certain people, places, or objects in order to regulate emotions'. This is close to situation modification, which are 'active efforts to directly modify the situation so as to alter its emotion impact'. Attentional deployment describes the process of directing attention towards or away from an emotional situation, e.g., looking for distraction. Cognitive change is defined as changing the cognitive steps necessary to elicit an emotion. Response modulation aims to modify the response to an emotion after it is already fully felt. While coping strategies have been widely studied, no research so far has investigated which coping strategies are used in frustrating in-car situations in everyday life.

### 2.4 Methods to investigate causes for emotions

Focus groups and diary studies are two common methods of psychological qualitative analysis [2, 8, 24]. An advantage of diary studies is that participants can report about their feelings during everyday situations in real-time. Bias of emotion cause or intensity due to memory is very unlikely. The benefit of a focus group study is that in-depth discussion lead to reflection on emotion causes and intensity. In comparison, the diary study is likely to reflect the amount of day-to-day occurrences of frustration. In contrast to that, the focus group study reveals which frustrating situations stay in memory on the long term. Both methods have been used previously to investigate emotions on in the road. Underwood et al. [27] used a diary study in which participants wrote down situations in which they felt anger over a period of two weeks. The participants reported situations after each car journey they took with help of a microcassette recorder and also rated their anger on a Likert scale. Huemer et al., [10] used focus groups to identify anger provoking events in cycling. They validated these findings by using a diary study. A common timeframe often used for diary studies is one week [2, 24]. Similarly, the research presented here has used a one week - diary study to investigate the spectrum and frequency of frustrating driving situations, and a focus group study to identify these situations in-depth and with a focus on long-term remembered frustrating situations.

## 3 Diary study

This study set out to explore reasons of frustration by means of a diary study. By collecting data for one week after every car ride, real-life occurring frustrating situations were captured together with their frequency.

## 3.1 Methods

Diary data was acquired in February 2019 using paper questionnaires which were distributed among the authors' networks. The participants were asked to fill out the questionnaire at the end of each car drive for seven days in a row to report their daily frustration experiences during driving.

### 3.1.1 Participants

Of 80 questionnaires that were distributed, 51 German-speaking participants (22 women, three unspecified gender) returned the questionnaire. The participants' age range was 20 to 73 years with a mean of 40.9 years (standard deviation [SD] = 12.5 years). On average, they drove 17,054 km per year (SD = 11,845 km), and had their driving license for 24 years (SD = 12.5 years). Figure 1 shows the distributions of demographic data.

### 3.1.2 Questionnaire

The German paper questionnaire for the diary study was custom-designed for the study and had fourteen pages. The first page contained the declaration of consent. On page two, the questionnaire's aim was explained and it was clarified that situations that cause frustration can occur before and during the ride. The definition of frustration was given as 'emotion that arises through a goal that is blocked to be reached.' For each frustrating situation that occurred to them during the week of data acquisition, participants were asked to fill in frustration intensity, a short description of the frustrating situation, and the importance of the blocked goal. Frustration intensity and the importance of the blocked goal were rated on a 5-point Likert-scale (from 1 = 'not at all', 'a little bit', 'somewhat', 'very' to 5 = 'extremely').

On page two and three, four examples of how to fill in the questionnaire were given (for example, 'I wanted to check the weather on my smartphone but the browser crashed all the time' as very frustrating and a little bit important goal).

On page four, participants were asked to provide personal information (age, gender, km/year, year of driver's license acquisition).

On pages five to twelve, participants were asked to report frustrating situations, the level of frustration intensity and the importance of the goal.

On page 13, it was asked on how many days of the week they drove and how often they remembered to fill in the questionnaire.

On the last page, the participants were asked to recall a maximum of three previous frustrating situations in driving that they experienced before the diary study.

### 3.1.3 Data analyses

For data analyses, the paper questionnaires were transliterated and three independent raters decided whether or not the reported situations described the emotion of frustration according to the definition of the feeling that arises when a goal is blocked to be reached [28]. The three raters also decided on categories for reasons for frustration with an inter-rater-reliability of 99.1%. In case of disagreement they solved the disagreement by discussion and agreed on a category together.

Subsequently, the correlation between the importance of the goal and frustration intensity was calculated. Data was separated between situations that happened during the week of data acquisition (pages five to twelve of the questionnaire) and recalled situations (page 14 of the questionnaire). In the following, the amount each category was mentioned was counted, and the mean frustration intensity per category was calculated in R (R Core [26]).

## 3.2 Results

The participants drove on 6.2 days out of seven on average and remembered to fill in the questionnaire on 5.9 out of seven days. 346 situations were described in total, out of which 88 situations were recalled, i.e., from their previous driving history. The raters excluded 161 situations because they described emotions other than frustration. The raters

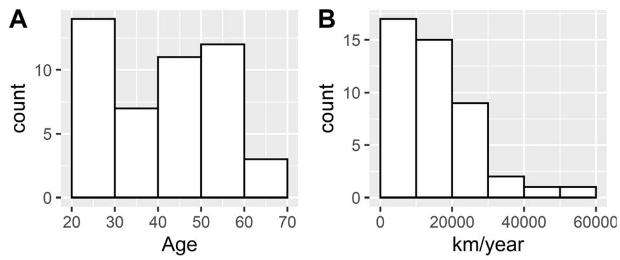


**Fig. 1** Demographic data of participants. **a** Age distribution, **b** mileage distribution, **c** Sex distribution (NA for participants who did not identify their gender)

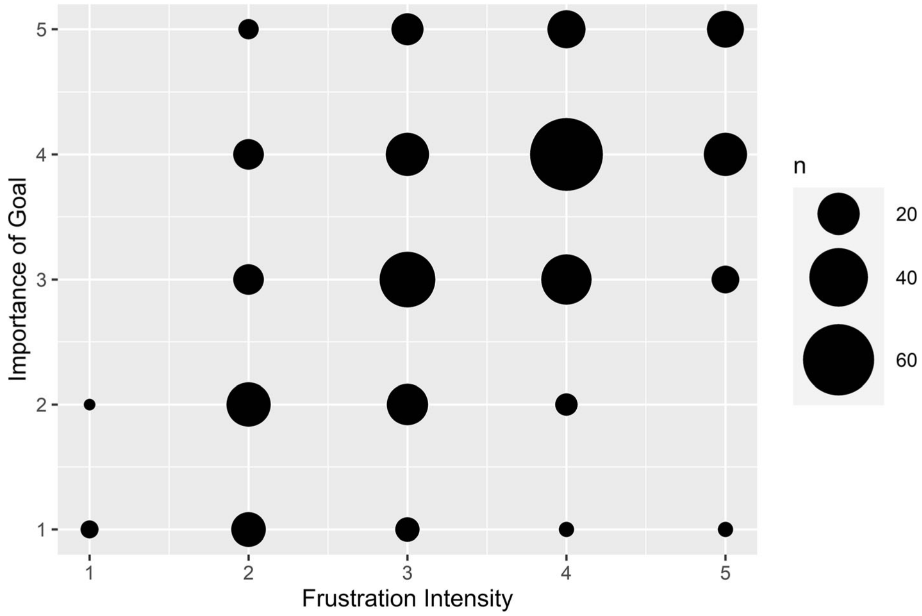


**Fig. 2** Bubble plot that shows the correlation of importance of goal and frustration intensity. n is the amount of described situations

considered 136 out of the 161 situations as ‘anger’. The other situations were rated as shame (6 situations), fear (3 situations), scare (2 situations), overextension (2 situations), disgust (1 situation), sadness (1 situation), worry (1 situation). Consequently, 185 frustrating situations were available for analyses. These situations happened in 128 drives, which results on average in 1.44 situations per drive. The categories and their respective subcategories for both the diary study and the focus group study were:

| Category | Subcategory |
|---|---|
| Traffic | Finding parking<br>Dense traffic |
| In-car situations | Other passengers/social environment<br>Human Machine Interface<br>Technical defects<br>Events before the start of ride |
| Weather | – |
| Self-inflicted | – |
| Other | – |

A Spearman’s correlation between the importance of the blocked goal and frustration intensity resulted in a rho of .48, $p < .001$ (Fig. 2). In this figure, a perfect correlation would show only large bubbles on the diagonal line between ‘Frustration Intensity’ and ‘Importance of Goal’. Also, the bigger bubbles on the right-hand side of the plot indicate that generally more situations have been rated with a high frustration rating.

The reasons for frustration were categorized into four categories and eight subcategories. Categories, subcategories and examples of each can be found in Table 1, the amount each category was mentioned in Fig. 3 (situations that occurred during the week of data acquisition) and Fig. 4 (recalled situations). For situations that occurred during the week of data acquisition, most situations were sorted into the category ‘traffic’ (54.5%), which consisted of the subcategories dense traffic (31.3%), red lights (9%), finding parking (7.5%), construction sites (4.5%), unnecessary traffic rules (1.5%) and unclear traffic management (0.7%). The category that occurred second-most was in-car situations (16.4%) which consisted of the subcategories social environment (6.7%), Human-Machine-Interface (3%), technical defects (3%), events before the start of ride (3%) and wrong information about traffic (0.7%).The third category was weather (13.4%), which did not have a subcategory. The smallest category with 9.7% and no subcategories was the self-inflicted category. Interestingly, the amounts these categories were mentioned are very similar for the recalled situations (Fig. 4). The intensity ratings for each subcategory do not show any clear differences between subcategories and are shown in Fig. 5 (situations during the week of data acquisition) and Fig. 6 (recalled situations).

**Table 1** Frustrating situations sorted by categories and subcategories with examples

| Category | Subcategory | Example |
|---|---|---|
| Traffic | | |
| | Finding parking | ‚no free parking spots' |
| | | ‚my usual parking spot was taken' |
| | Dense traffic | ‚I had to wait for two red light cycles' |
| | | ‚standing three hours because of traffic jam' |
| | Red lights | ‚I had to wait for 9 min at a closed train gate' |
| | | ‚many red lights and high traffic density' |
| | Construction sites | ‚roadworks and road constriction' |
| | | ‚long roadworks with speed limits' |
| | Unclear traffic management | ‚missing lane change from the center lane to the left lane' |
| | Unnecessary traffic rules | ‚many trucks and speed limits for no clear reason' |
| in-car | | |
| | Human-Machine-Interface | ‚setting up the navigation system was so complicated I had to stop on the right hand side' |
| | | ‚Android Auto updated, all settings were changed. I had to leave the highway and change back all settings' |
| | Events before start of ride | ‚co-driver complains about driving style' |
| | | ‚badly cognizable pedestrian because of bad weather conditions' |
| | Social environment | ‚argument with my son' |
| | | ‚car passengers linger at the roadhouse' |
| | Technical defect | ‚breakdown of the car, damaged beyond repair' |
| | | ‚flat tire on the highway' |
| | Wrong information about traffic | ‚Route diversion isn't displayed in the navigation system' |
| self-inflicted | | |
| | | ‚forgot my chip-card to get into the parking garage' |
| | | ‚I got caught in a speed trap' |
| weather conditions | | |
| | | "Having to drive slowly because of snow' |
| | | "bad sight because of blinding lights of oncoming vehicles in the snow' |
| others | | |
| | | ‚something clatters in the trunk' |
| | | ‚only bad music in the radio and everywhere the same music' |

# 4 Focus group study

## 4.1 Aim

The focus group study was conducted to further investigate reasons for which drivers become frustrated. In comparison to a diary study, focus groups enable discussions between participants, and make it possible to remind participants to stay with situations that are suitable according to our definition of frustration [23]. In addition to finding out reasons for frustration, the focus groups were designed to find out methods to cope with frustrating situations.

## 4.2 Methods

Seven focus groups consisting of four to six people each were conducted in July 2019. The participants were drawn from the Institute's participant data base.

### *4.2.1 Participants*

In total, 37 participants (14 women) participated. The mean age was 48.4 years with a range from 19 to 74 years (*SD* = *20.0 years*). The distribution of participants' demographic data is shown in Fig. 7. They gave their written informed consent, were native German speakers and had a valid driver's license.

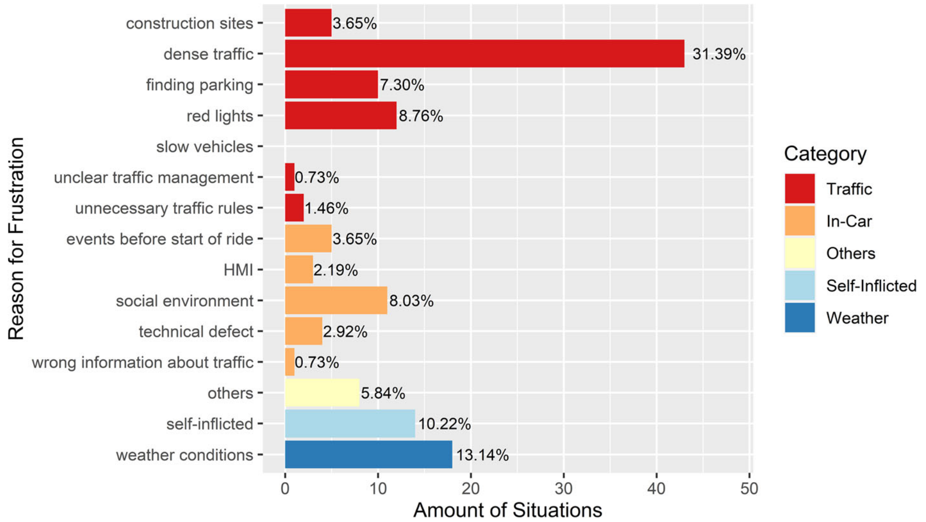


**Fig. 3** Amount of frustrating situations that occurred during the week of the diary study. Categories from both the diary study and the focus group study are shown on the y-axis for comparison

#### *4.2.2 Procedure*
Two female moderators lead the focus groups. Each session started with an explanation of the definition of frustration, especially in distinction to anger. In the following, the participants had time to brainstorm frustrating moments that they had experienced during driving. The results were collected on a pin board. Double mentions were kept, too. For each situation, a rating on the frustration intensity was given on a scale from one to five by the person who brainstormed the situation. Subsequently, the participants indicated how they usually cope with their frustration in the given moment. Each focus group took about 2 h.

#### *4.2.3 Data analyses*
The focus groups' data (audio recordings and photo protocols) were transliterated and frustration situations and coping strategies were categorized (inter-rater-reliability: 95.9%). For the 19 cases that they disagreed, a third rater gave a category and the majority vote won. Frustrating events were categorized into the same categories and subcategories as for the diary study. Also coping strategies were categorized into subcategories and categories. Categories were chosen according to Gross [9]. In his work, Gross divided emotion regulation into the categories of situation selection (for example, avoiding places/people that cause an emotion), situation modification (i.e., changing the situation to change the emotion it elicits), attentional deployment (like distraction), cognitive change (like re-evaluation of the situation) and response modulation (e.g. taking a deep breath to calm down). Subsequently, the amount categories were mentioned was counted and their mean frustration intensity or helpfulness rating calculated.

### 4.3 Results
#### *4.3.1 Frustrating events*
In total, 107 frustrating situations and 116 coping strategies were collected. The category that was mentioned most often was traffic (63.3%). Other categories included in-car causes for frustration like other passengers or the Human-Machine-Interface (13.1%), self-inflicted causes like starting too late (11.2%) and weather conditions (e.g., bad sight because of snow [7.5%]). These four high-level categories were divided into 13 subcategories (Fig. 8 for amounts and Fig. 9 for frustration intensity). Examples of situations with their categories and subcategories can be found in Table 2, the amount that each category was mentioned in Fig. 8.

The situations that were with more than 5% difference mentioned more often in the diary study (excluding recalled situations) were 1) dense traffic (31.3% vs. 16.8%), 2) social environment (6.7% vs. 1.9%), and 3) weather conditions (13.4% vs. 7.5%). The situations that were with more than 5% mentioned more often in the focus group study were slow vehicles (6.5% vs. 0), unclear traffic management (8.4% vs. 0.7%) and unnecessary traffic rules (6.5% vs. 1.5%).

#### *4.3.2 Frustration intensity*
Descriptive statistics indicate that frustration was highest for 'HMI' and 'wrong information about traffic' and lowest for 'weather conditions' and 'others' (Fig. 9).

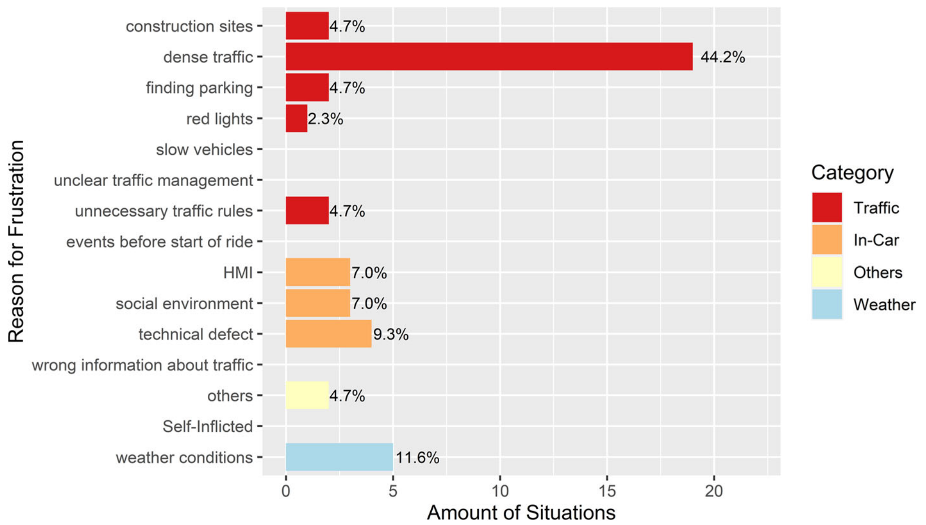


**Fig. 4** Amount of recalled frustrating situations

#### *4.3.3 Coping strategies*

Participants described their usual coping strategies with frustrating situations (Table 3 and Fig. 10) and how helpful they evaluate them (Fig. 11). According to the coping strategies proposed by Gross, [9], 35 coping strategies mentioned were categorized as attentional deployment, 32 as response modulation, 21 as cognitive change, 17 as situation modification and 11 as situation selection. In the subcategories, the strategy mentioned the most was cursing (18.3%), followed by distraction by media (17.4%), thinking about something else (11.3%), prevention strategies (9.6%), thinking differently about the situation (7.8%), breathing or relaxing (7%), accepting the situation (6.1%), adapting one's own driving style (6.1%), looking for a solution (5.2%), leaving the situation or taking a break (3.5%), distraction by others (2.6%), talking to someone about the situation (2.6%), changing one's aims (1.7%) and smoking (0.9%).

## 5 Discussion

This study was designed to determine the spectrum of reasons for frustration in driving and possible mitigation methods for the same. For this, the two methods of a diary study and a focus group study were employed. In comparison, the diary study reveals more information about day-to-day occurrences of frustration. In contrast,

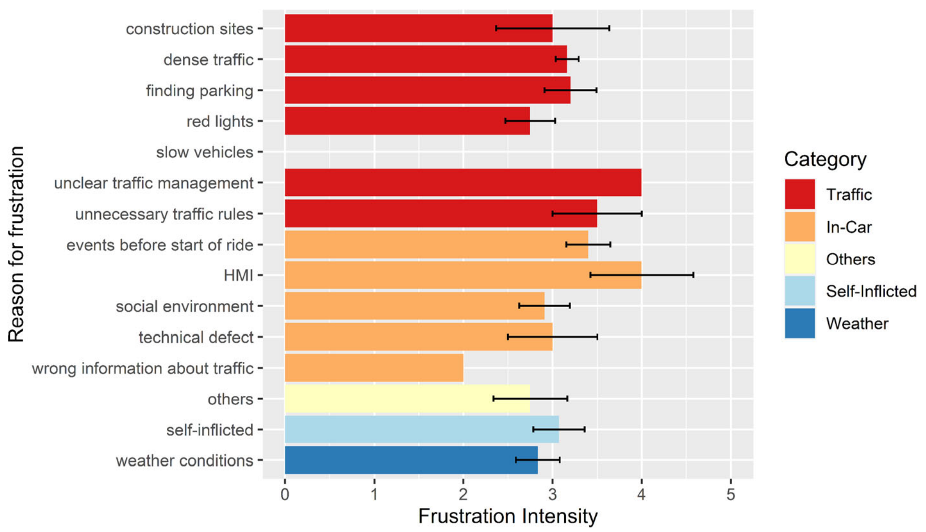


**Fig. 5** Frustration intensity sorted by category. Error bars indicate standard error of the mean

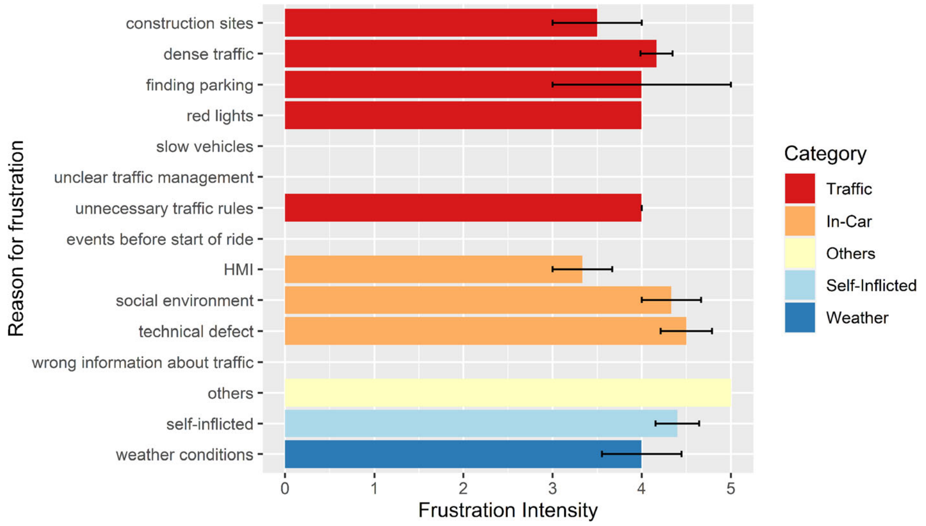


**Fig. 6** Frustration intensity of recalled situations, sorted by category. Error bars indicate standard error of the mean

the focus group study explored which reasons for frustration stay in memory on the long term.

### 5.1 Reasons for frustration

The high correlation between frustration intensity and importance of the goal supports the definition of frustration as the emotion that arises when a goal is blocked to be met [28]. Furthermore, the situations that lead to frustration during driving were collected and their frequency counted: traffic (diary: 54.5%, focus groups: 63.3%), in-car (diary: 16.4%, focus groups: 13.1%), self-inflicted (diary: 9.7%, focus groups: 11.2%), weather (diary: 13.4%, focus groups: 7.5%) and others (diary: 6.0%, focus groups: 4.7%). Interestingly, some differences in the results from the two employed methodological approaches occurred. Especially the subcategories of dense traffic, social environment and weather conditions seem to occur more often in day-to-day-life than they are remembered. Vice versa, situations that are more often named from memory than they occur in everyday life are slow vehicles, unclear traffic management and unnecessary traffic rules. These differences might be due to the fact that some situations are frustrating in the moment but less remembered on the long term. This might have various reasons. For example, users might show increased acceptance for frustrating events to which they can relate better or the reasons of which they understand better. By this, situations that are frustrating in a situation might be remembered less on the long term (dense traffic, social environment and weather conditions). On the other hand, if reasons are unclear, frustrating situations are increasingly remembered on the long term (slow vehicles, unclear traffic management and unnecessary traffic rules).

Unfortunately, this study could only assess frustrating situations and coping methods in manual driving. For automated driving, especially the cases mentioned in the in-car category would most likely be of interest. Most of these in-car situations could occur as likely or even more likely in automated driving. In a study with a similar goal – finding emotional triggers during a 50 min

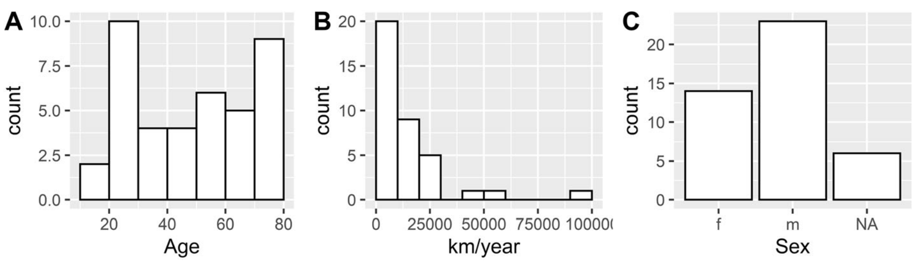


**Fig. 7** Demographic data of participants. **a** Age distribution, **b** Mileage distribution, **c** Sex distribution (NA for participants who did not identify their gender)

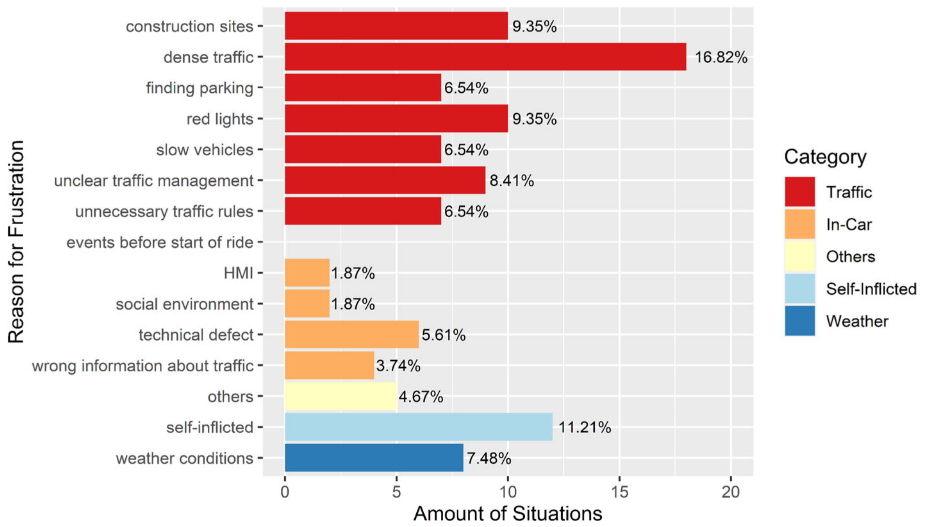


**Fig. 8** Amount that each subcategory was mentioned in the focus groups

car drive - [29] found that traffic, driving task, Human-Machine-Interface, and navigation most frequently lead to negative emotions. This is in line with our findings. To sum up, situations that lead to frustration most frequently were related to traffic, in-car situations and self-inflicted causes.

### 5.2 Coping with frustration

Of Gross' [9] categories, attentional deployment (distraction, thinking about something else) and response modulation (cursing, breathing/relaxing) were mentioned as being used most often. In contrast to this, the categories of situation selection (prevention strategies) and situation modification (looking for a solution) were rated as most helpful. As a concrete example, when asking participants about their own coping strategies, 'cursing' was mentioned the most often and rated as least helpful on average. A strategy that was rated as very helpful but only six times mentioned as actually used is 'look for a solution'. This is an important gap giving room for effective intervention.

For a frustration-aware system, this could mean that after having recognized frustration and its cause, methods of distraction or a voice assistant helping with modification of the emotional response are most interesting to develop. The exact character of the same is a promising next step for further research.

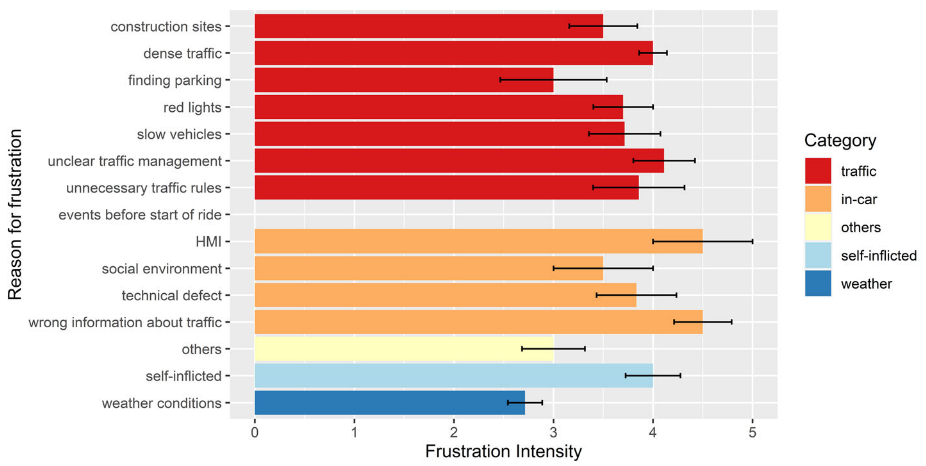


**Fig. 9** Frustration intensity sorted by subcategories. Error bars indicate standard error of the mean

**Table 2** Frustrating situations sorted by categories and subcategories with examples

| Category | Subcategory | Example |
|---|---|---|
| traffic | | |
| | Construction sites | 'roadworks' |
| | | 'track width of roadworks' |
| | Dense traffic | 'traffic jam due to an accident' |
| | | 'slow traffic flow in the rush hour' |
| | Finding parking | 'didn't find a parking spot' |
| | | 'parking in big cities' |
| | Red lights | 'bad traffic light circuit' |
| | | 'no green wave on the main street' |
| | Slow vehicles | 'stuck behind a truck on a curvy road' |
| | | 'slow car on the road' |
| | Unclear traffic management | 'too many road signs' |
| | | 'badly signposted diversion road' |
| | Unnecessary traffic rules | 'unnecessary speed limit' |
| | | 'traffic light circuit led to a long latency at night' |
| in-car | | |
| | HMI | 'drive in a rental car with a lane departure warning system that constantly warned me in a roadwork section' |
| | | 'infotainment-system hangs while driving' |
| | Social environment | 'co-driver constantly instructs me while driving' |
| | | 'co-driver criticizes my driving mode' |
| | Technical defect | 'malfunctions of the car' |
| | | 'car didn't recognize the car key' |
| | Wrong information about traffic | 'obsolete traffic news' |
| | | 'suddenly blocked road' |
| others | | |
| | | 'bad roads (potholes)' |
| | | 'too expensive fuel' |
| self-inflicted | | |
| | | 'got caught in a speed trap' |
| | | 'got lost while driving' |
| weather | | |
| | | 'bad view and difficult driving conditions' |
| | | 'too much heat in the car' |

### 5.3 Implications

The findings of the current research help to determine what information a frustration-aware assistant needs to know about the driver's context. With increasing availability of sensors in the vehicle, information coming from these can be used not only for recognition of frustration, but also its reasons. Based on the current study, the development of a frustration-aware system can be enriched by 1) knowledge about where to gather information regarding causes of driver frustration and 2) likelihoods of these causes. In combination with the information of measured frustration level, the system can offer help or mitigation methods tailored towards the specific situation.

### 5.4 Limitations

The generalizability of this study is subject to limitations. First, the diary study was distributed the

**Table 3** Examples for mentioned coping strategies

| Category | Subcategory | Example |
|---|---|---|
| Attentional deployment | | |
| | Distraction by media | 'turn on music' |
| | | 'listen to a podcast' |
| | Think about something else | 'count to ten' |
| | | 'try to enjoy the landscape' |
| | Distraction by others | 'distraction through talking to someone' |
| | | 'talk to co-driver (to distract myself)' |
| Response modulation | | |
| | Curse | 'curse once and call the other an idiot, after that I'm relaxed' |
| | | 'yell out of the window' |
| | Breathing/Relaxing | 'massage my earlobes' |
| | | 'taking a deep breath' |
| | Smoke | 'smoked a cigarette' |
| Cognitive change | | |
| | Think differently about situation | 'I took a step back in thought to get an overview' |
| | | 'remind myself that coming home safe is more important than this takeover |
| | Accept situation | 'see the situation more relaxed' |
| | | 'I decided to wait' |
| | Talk to someone about situation | 'I talked to my co-driver about the situation' |
| | | 'talk about the situation with a friend on the phone' |
| | Change aims | 'communicate that I will be too late' |
| | | 'set a new time frame' |
| Situation modification | | |
| | Adapt own driving style | 'switch on ACC to 80 (instead of the 100 that is allowed) if streets are crowded' |
| | | 'drive slowly' |
| | Look for solution | 'looked for a solution of the problem' |
| | | 'ask an expert for help' |
| | Leave situation / take a break | 'look for an alternative route' |
| | | 'took a break' |
| Situation selection | | |
| | Prevention strategies | 'leave my house on time' |
| | | 'avoid places that repeatedly lead to frustration' |

among authors' networks. This might have led to a biased group of participants. Second, a diary study during a longer term than one week might show different results. It might be interesting to repeat the study with a larger subject number. Third, data was acquired only during manual drives. When planning to use frustration-aware systems in automated driving, validation of the presented study results is necessary. Currently, this is a challenging task considering the spread of automated vehicles. Last, diary studies and focus groups acquire data by asking participants about their emotion. The answers can be dependent on factors other than the primary cause for frustration (for example, higher frustration if previous events were frustrating that day). Also, reports on frustration given in retrospective (after each ride) might differ from immediate reactions.

## 6 Conclusion and outlook

The results of this study indicate that reasons for experiencing frustration in driving are related to traffic (construction sites, dense traffic, finding parking, red

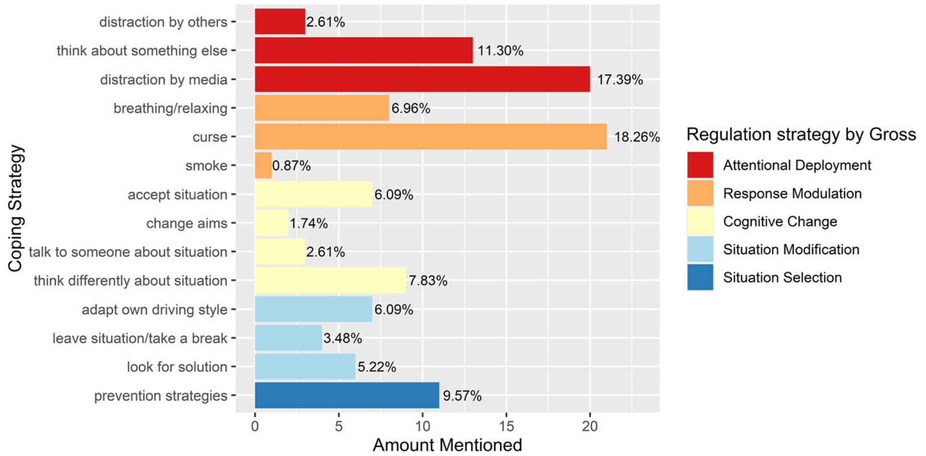


**Fig. 10** Coping methods and the amount that they were mentioned

lights, slow vehicles, unclear traffic management, unnecessary traffic rules), in-car causes (events before the start of ride, Human-Machine-Interface, social environment, technical defects, wrong information about traffic), self-inflicted causes and weather conditions. The reasons most feasible to target are probably the ones in the in-car category. By recognition of the frustration's time of occurrence combined with tracking the driver's current focus of attention (e.g., by eye tracking), the cause for frustration could, e.g., be differentiated between events before the start of drive (= frustration is recognized right when the passenger gets into the car) and Human-Machine-Interface (= frustration occurs while driver interacts with Human-Machine-Interface). According to the reasons of frustration, a frustration-aware system could offer help through the personal assistant or the Human-Machine-Interface, and algorithms could be trained towards personal preferences of the user. This help could be inspired by the coping strategy results: finding a solution is the most helpful option, but if that is not possible other ways of mitigating frustration are distraction by others or media or thinking about something else, amongst others. To sum up, this study helped to shed light on reasons for frustration and coping strategies employed by vehicle users. Further studies are needed to verify the research presented here, including hypothesis-based experiments that, e.g., could test for differences between different user groups (for example by age, driving experience, or cultural differences). A future questionnaire could additionally ask participants whether they felt like they were driving differently due to frustration, e.g., unsafely in terms of speeding or decreased time headways. Also, the format of a mobile

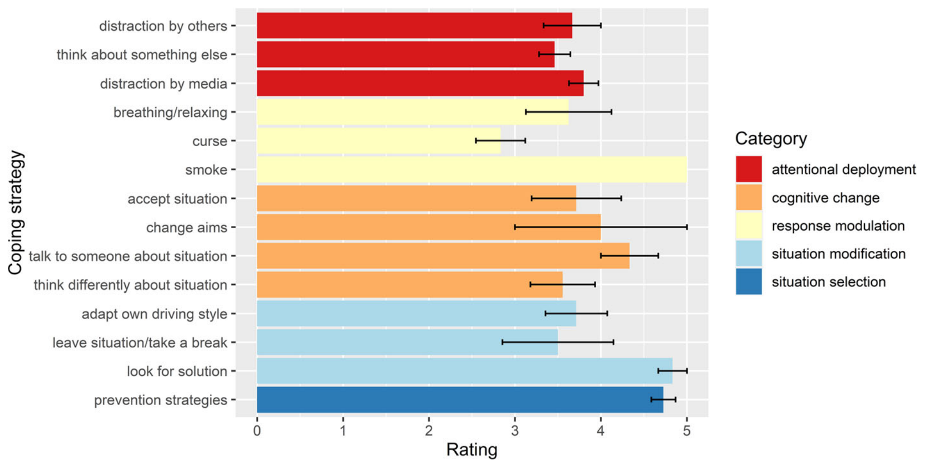


**Fig. 11** Rating of helpfulness of coping strategies. Error bars indicate standard error of the mean

application could facilitate the conduct of the study with a larger number of participants or over a longer time period (e.g. by sending regular reminders). Future research aiming at the design of frustration-aware in-vehicle assistants could build on the knowledge presented here to improve the detection of causes for frustration and the design of optimal coping strategies.

## 7 Supplementary information

**Supplementary information** accompanies this paper at https://doi.org/10.1186/s12544-020-00441-7.

**Additional file 1.**

#### Acknowledgements

The authors gratefully acknowledge the support of Meike Kühne, Sarah Klages and Alexander Behrens during data collection and categorization.

#### Author's contributions

Diary study: EB conceived, designed, acquired, analysed and interpreted the data, all with supervision of KI. UD helped with conception, design and interpretation of the data. Focus Group study: MO conceived, designed and acquired the data in consultation with EB and KI. EB did data analysis and interpretation with help of MO and KI. JM helped with conception and interpretation of the work of both studies. EB drafted the paper and KI, MO and JM substantively revised it. All authors read and approved the final manuscript.

#### Funding

This work was partly supported by the project F-RELACS which receives funding from the German Federal Ministry of Education and Science in the program KMU-innovativ under the grant 16SV7930. Open access funding provided by Projekt DEAL.

#### Availability of data and materials

The dataset supporting the conclusions of this article is included within the article and its additional file.

#### Competing interests

There are no competing interests.

Received: 12 May 2020 Accepted: 17 August 2020
Published online: 22 September 2020

#### References

1. Aslam, S., Verhoeven, G., Gouweleeuw, K., & Zwart, N. (2019). *Classification of disappointment and frustration elicited by human-computer interaction: Towards affective HCI*.
2. Berg, J., Levin, L., Abramsson, M., & Hagberg, J.-E. (2015). "I want complete freedom": Car use and everyday mobility among the newly retired. *European transport research review*, *7*, 31.
3. Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist*, *45*, 494.
4. Deffenbacher, J. L., Deffenbacher, D. M., Lynch, R. S., & Richards, T. L. (2003). Anger, aggression, and risky behavior: A comparison of high and low anger drivers. *Behaviour research and therapy*, *41*, 701–718.
5. Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*.
6. Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B., et al. (2010). *Emotion on the road—Necessity, acceptance, and feasibility of affective computing in the car. Advances in Human-Computer Interaction*, (p. 2010).
7. Folkman, S., & Lazarus, R. S. (1990). *Coping and emotion. Psychological and biological approaches to emotion*, (pp. 313–332).
8. Geis, I. (2019). Foreign countries, foreign customs: An analysis of short-distance mobility of new immigrants in the Rhine-Main region in Germany. *European transport research review*, *11*, 23.
9. Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, *2*, 271–299.
10. Huemer, A. K., Oehl, M., & Brandenburg, S. (2018). Influences on anger in German urban cyclists. *Transportation Research Part F: traffic psychology and behaviour*, *58*, 969–979.
11. Ihme, K., Unni, A., Zhang, M., Rieger, J. W., & Jipp, M. (2018). Recognizing Frustration of Drivers From Face Video Recordings and Brain Activation Measurements With Functional Near-Infrared Spectroscopy. *Frontiers in human neuroscience*, *12*, 327.
12. Jeon, M., Walker, B. N., & Yim, J.-B. (2014). Effects of specific emotions on subjective judgment, driving performance, and perceived workload. *Transportation Research Part F: traffic psychology and behaviour*, *24*, 197–209.
13. Johnson, M., & McKnight, S. (2009). Warning drivers about potential congestion as a means to reduce frustration-driven aggressive driving. *Traffic Injury Prevention*, *10*, 354–360.
14. Lazar, J., Jones, A., Hackley, M., & Shneiderman, B. (2005). Severity and impact of computer user frustration: A comparison of student and workplace users. *Interacting with Computers*, *18*, 187–207.
15. Lazarus, R. S. (1991). *Emotion and adaptation. Oxford University Press on Demand*.
16. Lee, Y.-C. (2010). *Measuring drivers' frustration in a driving simulator*, (p. 19).
17. Li, S., Zhang, T., Liu, N., Zhang, W., Tao, D., & Wang, Z. (2020). Drivers' attitudes, preference, and acceptance of in-vehicle anger intervention systems and their relationships to demographic and personality characteristics. *International Journal of Industrial Ergonomics*, *75*, 102899.
18. Löcken, A., Ihme, K., and Unni, A., eds (2017). *Towards designing affect-aware Systems for Mitigating the effects of in-vehicle frustration. ACM*.
19. Malta, L., Miyajima, C., Kitaoka, N., & Takeda, K. (2011). Analysis of real-world driver's frustration. *IEEE Transactions on Intelligent Transportation Systems*, *12*, 109–118.
20. Mikolajczak, M., Nelis, D., Hansenne, M., & Quoidbach, J. (2008). If you can regulate sadness, you can probably regulate shame: Associations between trait emotional intelligence, emotion regulation and coping efficiency across discrete emotions. *Personality and Individual Differences*, *44*, 1356–1368.
21. Miller, N. E. (1941). I. The frustration-aggression hypothesis. *Psychological review*, *48*, 337.
22. Oehl, M., Ihme, K., Drewitz, U., Pape, A.-A., Cornelsen, S., and Schramm, M., eds (2019). *Towards a frustration-aware assistant for increased in-vehicle UX: F-RELACS. ACM*.
23. Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research & Method in Education*, *29*, 23–37.
24. Raux, C., Ma, T.-Y., & Cornelis, E. (2016). Variability in daily activity-travel patterns: the case of a one-week travel diary. *European Transport Research Review*, *8*, 26.
25. Shinar, D. (1998). Aggressive driving: The contribution of the drivers and the situation. *Transportation Research Part F: traffic psychology and behaviour*, *1*, 137–160.
26. Team, R. C. (2019). R Core Team (2019). R: A language and environment for statistical computing. In *R Found. Stat. Comput. Vienna, Austria URL http://www.R-project. org/., page R Foundation for Statistical Computing*.
27. Underwood, G., Chapman, P., Wright, S., & Crundall, D. (1999). Anger while driving. *Transportation Research Part F: traffic psychology and behaviour*, *2*, 55–68.
28. Wirtz, M. A., & Strohmer, J. (2017). *Dorsch-Lexikon der Psychologie. 18., überarbeitete Auflage*. Bern: Hogrefe.
29. Zepf, S., Dittrich, M., Hernandez, J., and Schmitt, A., eds (2019). *Towards empathetic Car interfaces: Emotional triggers while driving. ACM*.

## Publisher's Note

# 4 Discussion

Frustration-aware assistance systems hold the potential to make traveling safer and increase acceptance of new and possibly more sustainable mobility solutions. However, they will only be helpful if they work in a reliable manner and might change things for the worse if working poorly. This dissertation presented four original research papers that contribute to the successful development of such frustration-aware assistance systems by improving the reliability of frustration recognition. I examined the data collected in Study 1 from three different perspectives: First, I collected, prepared, and described the data and its characteristics and challenges when working with the data. Second, I examined and quantified inter-individual differences in frustration expression. For this, I used not only the data from Study 1, but also from Study 2. Third, I evaluated the method of using a latent variable model to compare the influence different measurement modalities have on estimating experienced frustration. Finally, in Study 3, I examined causes for which drivers become frustrated during driving in a diary study. In the remainder of this chapter, I will discuss how the presented research contributes to the advancement of developing FAAS, give an outlook for future directions of the development of emotionally aware assistance systems. Last, I will discuss ethical implications of FAAS, and finally discuss a vision of how our current transportation system can be changed for the better by developing emotionally aware assistance systems.

## 4.1 Data Sources and Estimators - Traveler

The first part of this dissertation sheds light on how frustration can be measured by answering research question 1: 'How can we classify frustration by multimodal measurements?'. For this, we conducted a driving simulator study with 50 participants (Study 1) and a real-car driving study with 23 participants (Study 2). We induced frustration with traffic-related and Human-Machine-Interface (HMI)-related tasks that were difficult to achieve. We measured facial and bodily expressions, EEG data, ECG data, and skin conductance during all drives. We collected subjectively experienced frustration by a Likert-scaled questionnaire after every drive. In addition, participants saw their drives' video recordings after they were done with the experiment. Based on this, they gave a continuous frustration rating with a joystick and visual feedback of their current rating. We compared expression frequencies during baseline and frustration-inducing drives and confirmed previously described frustration-typical expressions (Grafsgaard et al., 2013; Hoque and Picard, 2011; K. Ihme, A. Unni, et al., 2018) of Brow Lowerer, Dimpler, Brow Raiser, Smile, and Lip Press. In addition, we used the nearest neighbor approach to recognize participants from one drive to another out of all other participants, who worked with an accuracy of 37.5% in the simulator study (chance level: 2.4%) and 42.86% in the real-car driving study (chance level: 4.8%). We, therefore, concluded that participants showed expressions of frustration that were typical per person, which we called individual-typical expressions. This individual-typical expression of emotion has previously been described for other emotions (Cohn et al., 2002). These individual-typical expressions provide a content-related explanation for previous studies on automated emotion recognition that used black-box approaches more successfully when training towards an individual person (DMello and Kory, 2015; Kosch et al., 2020). In practice, frustration recognition could work in a generalized way at first. With a few ‘training’ drives, a FAAS could learn an individual’s typical expressions and, therefore, improve recognition performance over time.

Furthermore, I modeled experienced frustration according to the component process model of emotion (Scherer, 2009), which states that an emotion is accompanied by psychological and behavioral responses. For this, I used a latent variable model (Ben-Akiva et al., 2002). Such models have previously been used to model how stress changes car-following behavior (Paschalidis, Choudhury, and Hess, 2019) and to model how a driver’s cognitive effort impacts route choice decisions (Agrawal and Peeta, 2021). Castro, Guevara, and Jimenez-Molina (2020) use a latent variable model to predict transportation mode choices based on physiological and regularly reported subjective data. The latent variable model has the advantage that it is possible to model a non-measurable latent variable, like experienced frustration, with several surrogate measurement modalities. Subsequently, it is possible to build a model that leaves out one measurement modality, respectively, and to compare these against the full model. This comparison gives information on how important each measurement modality is for building the full model. By this, I showed that subjective rating and expression frequency of frustration-typical expressions are informative additions to a latent variable model of experienced frustration. Frontal alpha asymmetry, as measured by EEG, did not improve the latent variable model and therefore did not add new information to the measurement of frustration. A possible explanation for this is that the context of a driving simulator produces too much noise in the EEG signal to be utilized in such an applied setting. Previous studies that used more controlled laboratory conditions did find a difference in AAI in frustration vs. non-frustration (Reuderink, Mühl, and Poel, 2013). Therefore, EEG measurements are informative for researchers that test new human-machine interfaces in experimental laboratory studies. For the development of applied systems like a FAAS, this means that the use of expressions as an indicator of frustration is more informative than the measurement of AAI. In addition, measurement by a camera is less intrusive for the user. Future FAAS can use these insights to get in-situ sensing capabilities by using camera data from a smartphone attached to the windshield (Bethge, Kosch, et al., 2021). This way, the FAAS could be easily integrated into everyday navigation apps. Generally, Paper 3 shows that latent variable models are suitable for evaluating which measurement methods are needed in which context.

If a measurement (e.g., EEG) shows significant coefficients in the full model containing all measurements, the latent variable (i.e., the experienced frustration) has an essential impact on the measurement and is, therefore, useful in estimating the latent variable in future scenarios. On the other hand, if one decides to trust a combination of measurements in a context, then it is possible to estimate an unknown latent variable, like for example experienced frustration. This method, of course, is the ultimate aim of a FAAS.

I showed that the methods used for frustration induction in Studies 1 and 2 were successful in inducing driving- or vehicle-interaction-related frustration. For future research on frustration recognition, this means that using dense traffic, slow vehicles, and shortly timed red lights are suitable for inducing frustration in a manual driving context. For automated driving, HMI with unclear icons and paths work well. In comparison, the HMI was rated as more frustrating than the frustrators used in the manual driving use cases. I also found that participants' frustration rating might be influenced by experiencing something new and exciting during frustration induction. This influence seemed to be present in Study 2, where participants rated higher frustration in the frustration-inducing drives than in the baseline drives only in the frustration rating that was given in a calm setting after the drives in the automated car were experienced. The frustration rating given after every drive (on-site, in the automated car) did not differ between frustration-inducing and baseline conditions. Interestingly, the continuous frustration rating given post-hoc by all participants revealed that frustration built up slowly, over time, for most participants. This slow buildup is in contrast to other emotions, which are caused by one event and have a clear 'spike' after the emotion-inducing event. For example, most studies regarding the six basic emotions (Ekman, 1992) induce emotions through film clips or audiobooks (Maryam Fakhrhosseini and Myounghoon Jeon, 2017). However, this continuous buildup of frustration was not true for all participants. Some participants did rate frustration in spikes or not at all. We cannot know whether this difference in frustration rating is due to actual differences in experienced emotion or to the participant's way of rating frustration. The fact that frustration built up slowly implicates challenges for its real-time measurement, as sudden spikes of emotion, and therefore the associated measurements, are generally easier to detect than gradual changes in measurements.

Overall, the results regarding the recognition of frustration support researchers that aim to assess frustration in real-time through time-series data such as camera or EEG data. In comparison to questionnaires, which are commonly used at the end of every experimental condition (Breyer and Bluemke, 2016), such measurements enable to detect exact moments of frustration in real-time. They provide an implicit frustration measure while increasing the temporal accuracy compared to post hoc questionnaires. Implicit and continuous frustration measures like EEG or expression data have two key advantages: (1) In experimental settings, researchers can measure frustration as the additional dependent variable in real-time and avoid that participants solely remember their most frustrating experiences, and report only these in questionnaires. (2) In real-world settings, currently used machine learning methods lead to a generalized frustration recognition model that does not consider individual differences in perceived frustration and their thresholds. Through constant frustration measures, existing machine learning models can be refined to match individual frustration levels while deducing new potential sources for frustration, effectively including them into existing prediction models and making adaptive FAAS more robust.

## 4.2 Data Sources and Estimators - Context

The second part of this dissertation researches when opportune moments for mitigation by FAAS are by presenting contexts that usually lead to experienced frustration in driving. These presented contexts answer research question 2 of this dissertation: 'What contexts lead to in-vehicle frustration?' For this, we conducted a diary study in which participants reported frustrating events during one week of driving. These I found to be within the categories of traffic, in-car causes, self-inflicted causes, and weather conditions. This finding is consistent with the results from Bethge, Kosch, et al. (2021), who conducted a study where participants drove and reported their emotions every 30 seconds while contextual data such as cell phone data, calendar data, and weather and traffic information were recorded. They found that especially weather and traffic conditions co-occurred with negative emotions (Bethge, Kosch, et al., 2021). In a follow-up study, Bethge, Coelho, et al. (2023) found that using contextual data is more robust than using facial expressions for detecting drivers' emotions. To develop a complete picture of contextual causes of frustration, additional studies will be needed that investigate individual differences in elicitors of frustration. A study similar to that of Bethge, Kosch, et al. (2021), but focused on frustration only, could give important insights and especially more data on the bandwidth of possible causes of frustration, as well as a better statistical indication of the occurrence of each elicitor.

In practice, contextual frustration elicitor recognition could first work with the generally trained algorithm. Then, to improve the measurement, users could agree to customize their frustration context algorithm by labeling frustrative situations for the first few hours of driving. By this, individual-typical frustrative situations and the individualized extent of frustration could be extrapolated. The customization of expressions for frustration would be possible at the same time. Users of technological systems are more likely willing to share personal information if the system is personalized to their interests (Chellappa and Sin, 2005) or when providing it for a global benefit for the wider public (Ziefle, Halbey, and Kowalewski, 2016). However, a continuous camera measurement in-vehicle might cause privacy concerns. One solution to this is that the cameras recognizing expressions of frustration are only switched on when these frustrating situations are recognized, and detected frustration can be confirmed or dismissed. This combination of contextual and expressive data would pave the way for reliable frustration detection.

## 4.3 Mitigation Strategy and Intervention Execution

In this paragraph, I will discuss what a FAAS does upon detecting frustration and how the resulting frustration mitigation could affect the real world. Krüger et al. (2021) have used the same setup and frustration induction as Study 1 of this dissertation but used a voice assistant that empathized with the participant in the frustration-inducing conditions in one half of the participants. Based on Gross (1998), Krüger et al. (2021) used the mitigation strategies of situation modification and cognitive change to mitigate frustration in half of the participants. While the sample size was too small to reveal differences between the groups with and without the empathic assistant, a trend showed that participants of the empathic voice assistant group reported less frustration than the control group without a voice assistant. Although the voice assistant showed a mitigation trend, there is a gap in how mitigation can be efficiently communicated with users.

Hence, this paragraph describes how mitigation can be achieved via a communication modality. The human sensory system reacts differently to notifications with varying urgency. For example, Glatz et al. (2018) found that drivers perceive verbal commands and auditory icons differently depending on the auditory design. Since frustration can implicitly lead to situations where immediate driver attention is required, FAAS can communicate the mitigation modality that matches the sensory perception of the user. Although the previous study found that voice assistants lead to mitigation, future research must ensure that voice assistants communicate such mitigation efficiently. For example, verbal commands can be used when the user is frustrated and might not pay attention to the cues given by the voice assistant. Instead of using auditory icons (Glatz et al., 2018), the voice assistant can use verbal commands to gain the driver's attention. At the same time, cues that lead to the highest attention levels should be selective since users will get used to them. Thus, users will stop paying attention to the mitigation strategy. Isolating mitigation strategies and investigating their design is a relevant future research field. Especially when assessing mitigation strategies with physiological sensing (e.g., EEG), research can provide novel insights into isolated mitigation modalities' sensory perception and granularities.

## 4.4 Ethics

A point that is important to consider is that emotions are highly personal. In spaces where a recognized emotion could become visible to others, it is highly relevant that this information is concealed. Furthermore, according to the General Data Protection Regulation (GDPR), data cannot be stored permanently without the user's consent. Therefore, measured data must not be saved in the long term. With the same conclusion, K. Ihme, Bohmann, et al. (2021) developed a data privacy concept regarding a FAAS. It was based on the principle of 'privacy by design' (privacy is technically integrated when a system is developed) and 'privacy by default' (default settings are set to protect the privacy of users). Participants that rated the concept in an online survey evaluated the indicated measures as rather sufficient or sufficient on average.

A FAAS that accidentally leaks information about recognized emotions could easily lead to backlashes. Also, a FAAS that mis-recognizes moments of frustration could lead to acceptance changes towards the worse very quickly. It is, therefore, essential to ensure that a FAAS functions well before it is deployed in real-world systems. Furthermore, a FAAS needs to be designed in a way that avoids discrimination of gender, age, and race. Currently, most data that is collected to train frustration classification algorithms work with participants that are WEIRD – western, educated, industrialized, rich, and democratic (Linxen et al., 2021). No one should be systematically disadvantaged from using FAAS because the system does not detect their frustration correctly.

In addition, technological developments and assistance in systems where the human plays a safety-critical role hold the risk of over-reliance on the system. For example, in the case of a FAAS, a driver that has not been classified as frustrated might be less cautious when knowing that they are usually warned in cases their attention and emotions might diminish cognitive function. This could lead to more risky driving maneuvers or over-reliance on the system in general, for example, to warn in every critical situation or correct driving mistakes (e.g., lane keep assistant). On the other hand, travelers could experience loss of control if a FAAS is, for example, used to decide when a human driver can take back driving control from a highly automated driving car. Whether frustration is wrongly or correctly detected, in these cases, a human traveler could feel overruled by the system if they are not allowed to drive in case of frustration. As a result, frustration could be enhanced instead of mitigated in these cases. A system that can influence emotions has the potential to be misused in several contexts. For example, customers could be manipulated into buying products, or citizens could be spied on by authorities. An extreme example is given in George Orwell's 1984 (Orwell, 1954):

*"It was terribly dangerous to let your thoughts wander when you were in any public place or within range of a telescreen. The smallest thing could give you away. A nervous tic, an unconscious look of anxiety, a habit of muttering to yourself – anything that carried with it the suggestion of abnormality, of having something to hide. In any case, to wear an improper expression on your face (to look incredulous when a victory was announced, for example) was itself a punishable offence. There was even a word for it in Newspeak: facecrime, it was called."* — George Orwell, 1984

The possibilities of an automated emotion recognition system should be treated with caution when considering the following statement from Edward Snowden (Snowden, 2019):

*"Technology doesn't have a Hippocratic oath. So many decisions that have been made by technologists in academia, industry, the military, and government since at least the Industrial Revolution have been made on the basis of "can we," not "should we." And the intention driving a technology's invention rarely, if ever, limits its application and use."* — Edward Snowden, Permanent Record

## 4.5 Limitations and Future Work

In order to employ FAAS in the real world, a few steps are still necessary. First, broader data sets than the one presented in Paper 1 are needed to include culturally different people. Second, this thesis concentrated on post-hoc recognition, but it is necessary to build algorithms that detect frustration in real time. Third, by building upon the findings of Paper 2, algorithms that adapt to individuals in real-time are promising to improve frustration recognition further. Next, Paper 3 enables a comparison of which measurement methods are needed in which context. Further research is needed to evaluate which measurement methods are best to use in slightly different in-vehicle contexts, for example in public transport.

The scope of this thesis was limited to post-hoc analyses and did not yet produce a closed-loop FAAS. More research on successful mitigation strategies in general but also tailored towards different contexts and individuals is necessary to achieve this. Then, the next step is to produce a FAAS that recognizes frustration and reasons for frustration in real-time, initiates respective mitigation strategies, and evaluates whether frustration was reduced. The first step in this direction has been taken by Krüger et al. (2021) (see Section 4.3) and shows that participants that experienced the frustration mitigation voice assistant reported less frustration than the participants that did not. First, an evaluation of a closed-loop FAAS is needed to evaluate its utility in practice fully.

Furthermore, more modalities for the measurement of frustration exist. Previous studies have used ECG and skin conductance data to recognize frustration (Ding et al., 2020; Egger, Ley, and Hanke, 2019). The additional use of these data sources could further improve the recognition of in-vehicle frustration. I recorded ECG and skin conductance data in Studies 1 and 2, but found no differences between frustrating and non-frustrating drives in these measures. Research finds differences in ECG and skin conductance data between emotional conditions (DMello and Kory, 2015; Schmidt et al., 2019). For example, W. Yang et al. (2018) detect moments of frustration during gameplay based on ECG, skin conductance, electromyogram, respiration, and acceleration data with an F1-score of 68.5% when classifying time segments into frustration vs. no frustration. Belle et al. (2010) detect moments of frustration and no frustration based on only Heart Rate Variability (HRV) data with an accuracy of 81.45% and a sensitivity of 78.1%. However, experience sampling research that collected affective moments in everyday-life situations has found that physiological responses can be very different in similarly labeled emotions (Hoemann et al., 2020). DMello and Booth (2022) discuss the results of the MOSAIC Challenge, in which the U.S. Intelligent Advanced Research Project Agency (IARPA) funded three teams whose results were rigorously tested for accuracy and robustness in measuring 'psychological, physiological and physiological aspects of an individual.' None of the three teams met the demanded criteria, and several results were null (DMello and Booth, 2022). Therefore, I argue that a driving study's situation was too complex to measure frustration with a two-dimensional measure like ECG or skin conductance.

The analyses of the facial expression frequencies are based on hand annotations of facial expressions. These cannot be perfect, as facial expressions are not always shown in clear categories. We, therefore, used two independent annotators and a third that decided on one of the two annotations. Even though the inter-rater agreement was only moderate according to Landis and Koch (1977), this is similar to previous research in the field (Schömig et al., 2018). Human annotations outperform computational annotations nowadays (Dupré et al., 2020).

Analyzing facial expressions in a time-dependent way without using a black-box approach remains challenging, as long time stretches without any facial expression exist. Such sparse data is challenging to handle with any statistical method. The method of using expression frequencies solves this problem but needs to consider the time-series nature of the data. An approach considering this has been taken in Paper 1 by using a Long short-term memory model. This method, on the other side, is a black-box approach. Future work could find an analysis method that is time-dependent and makes results transparent at the same time.

The subjective continuous frustrating rating was collected post-hoc in Studies 1 and 2. This 'remembered' frustration level might have been different from the experienced emotion. Therefore, we also acquired the widely-used PANAS questionnaire in a modified version that includes a 'frustration'- item (we called it the emotion scale rating). By this, we compared the rating after every drive with the continuous rating given after all drives and found a high correlation between both ratings. A continuous frustration rating during the drive would have changed the participant's task from 'solving the actual task' to 'solving the actual task and rate frustration.' We, therefore, would have measured a frustration rating, not of the actual task alone, but a frustration rating that could also contain frustration induced by giving the continuous frustration rating. This is why we decided to use the emotion scale rating and the post-hoc continuous rating. In Study 3, the method of a diary study and a focus group study were used. Participants were asked to report their current experience of frustrating moments and to extrapolate of what they expect to be problems in automated driving. Of course, it is not possible at this moment to investigate the causes of frustration when automated driving vehicles belong to a common everyday life. This research would best be repeated in the future when automated vehicles are integrated into people's everyday life.

Individuals differ in how they experience frustration, how they express it, in which circumstances they feel frustration and which mitigation methods work in which contexts. Therefore, every step of a FAAS should be tailored to individuals rather than using a one-size-fits-all solution. For example, Bethge, Kosch, et al. (2021) find that a classifier based on context data for recognition of emotions during driving works much better when using a personalized classifier. Individuals are, therefore, highly different in which situations they experience as frustrating. Moreover, physiological changes can be very different per person (Hoemann et al., 2020). The paper on individual-typical expressions of this dissertation underlines these findings.

An important feature to consider when further developing FAAS is that it will work implicitly in the optimal case. Frustration would be detected and mitigated without the user noticing that he or she was on the way to becoming frustrated. For this, it would be necessary to detect frustration before the user feels it. Also, mitigation would need to be implicit enough not to be noticed as an intervention by the FAAS. This is, of course, a challenging task.

## 4.6 Conclusion

This dissertation took one step in the direction of the development of a frustration-aware assistance system. More specifically, I showed that it is possible to recognize frustration through facial expressions and that this can be improved by considering individual-typical expressions. I presented a method that enables the investigation of optimal emotion recognition methods in different contexts. Lastly, I found common frustrators in driving. These can be used to improve the recognition of frustration by considering contextual data.

The presented research shows that frustration can be detected in-vehicle reasonably well. Based on this, future research is promising for interfaces that adapt to sensed user frustration. The mitigation strategies mentioned in Section 4.3 can be connected with the frustration sensing modalities to present a closed-loop FAAS. For this, previously mentioned mitigation strategies, such as empathic voice assistants, can be employed in-situ when detecting moments of frustration. Consequently, I anticipate improvements in driver experience through reduced frustration levels.

One use case for a system that recognizes frustration is to provide users with the option to reflect when, where, and in which transportation modes of their journeys they are frustrated. By this, they can decide on the frustration-free instead of the fastest route. When collected over several participants, continuous data on travelers' frustration could also indicate points of frustration in a transportation system. These 'pain points' could inform transportation planners, scientists, politicians, and citizens of points that are challenging for the user.

If we imagine a world with fully functioning FAAS, frustration-related accidents could be reduced, and therefore mobility by car could be safer. When deployed in new mobility systems, a FAAS could increase user acceptance by decreasing frustration that mainly occurs when unfamiliar with a system. Ultimately, it might be possible to use FAAS not only in the context of transportation but also in different contexts, like in meetings, mobile interaction at the workplace, or to evaluate the customer experience. If slightly adjusted, FAAS could also help in mental health clinical settings. On the one hand, a FAAS could help patients who have trouble recognizing emotions. On the other hand, patients that are especially sensitive to frustration could be helped by specialized FAAS that could be tailored to their specific needs. Another use case could be to improve mental health and productivity at the workplace. This could be done by recognizing moments in which employees are unable to solve tasks so that help can be provided.

If fully functioning, such a FAAS could make 'Vision Zero' a reality – a vision to eliminate all fatalities and severe injuries in transportation (Tingvall, 1997). Moreover, in the future, a FAAS could help, especially in introducing new mobility concepts like automated vehicles. This development, in turn, is crucial to make a zero-emission vision for transportation reality.

# Bibliography


Agrawal, Shubham and Srinivas Peeta (2021). "Hybrid route choice model incorporating latent cognitive effects of real-time travel information using physiological data". In: *Transportation research part F: traffic psychology and behaviour* 81, pp. 223–239. ISSN: 1369-8478.

Barr, Leah K., Jeffrey H. Kahn, and W. Joel Schneider (2008). "Individual Differences in Emotion Expression: Hierarchical Structure and Relations with Psychological Distress". In: *Journal of Social and Clinical Psychology* 27.10, pp. 1045–1077. ISSN: 0736-7236. DOI: 10.1521/jscp.2008.27.10.1045.

Belle, Ashwin et al. (2010). "Frustration Detection with Electrocardiograph Signal Using Wavelet Transform". In: *2010 International Conference on Biosciences*. IEEE, pp. 91–94. ISBN: 978-1-4244-5929-2. DOI: 10.1109/BioSciencesWorld.2010.19.

Ben-Akiva, Moshe et al. (2002). "Integration of choice and latent variable models". In: *Perpetual motion: Travel behaviour research opportunities and application challenges* 2002, pp. 431–470.

Berkowitz, Leonard (1989). "Frustration-aggression hypothesis: examination and reformulation". In: *Psychological bulletin* 106.1, p. 59. ISSN: 1939-1455.

Bethge, David, Luis Falconeri Coelho, et al. (2023). "Technical Design Space Analysis for Unobtrusive Driver Emotion Assessment Using Multi-Domain Context". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. ISSN: 2474-9567.

Bethge, David, Thomas Kosch, et al. (2021). "VEmotion: Using Driving Context for Indirect Emotion Prediction in Real-Time". In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. Ed. by Jeffrey Nichols, Ranjitha Kumar, and Michael Nebeling. New York, NY, USA: ACM, pp. 638–651. ISBN: 9781450386357. DOI: 10.1145/3472749.3474775.

Bořil, Hynek, Pinar Boyraz, and John H. L. Hansen (2012). "Towards multimodal driver's stress detection". In: *Digital signal processing for in-vehicle systems and safety*. Springer, pp. 3–19.

Bosch, Esther, Raquel Le Houcq Corbí, et al. (2022). "Frustration Recognition Using Spatio Temporal Data: A Novel Dataset and GCN Model to Recognize In-Vehicle Frustration". In: *IEEE Transactions on Affective Computing*, pp. 1–12. DOI: 10.1109/TAFFC.2022.3229263.

Bosch, Esther, Klas Ihme, et al. (2020). "Why drivers are frustrated: results from a diary study and focus groups". In: *European transport research review* 12.1, pp. 1–13.

Bosch, Esther, David Kaethner, et al. (n.d.). "Fifty Shades of Frustration: Intra- and Interindividual Variances in Expressing Frustration". In: *Transportation Research Part F: Traffic Psychology and Behaviour* ().

Bosch, Esther, Marie Klosterkamp, et al. (2022). “Multimodal Estimation of Frustrative Driving Situations Using a Latent Variable Model”. In: *2022 13th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, pp. 000011–000016.

Braun, Michael, Bastian Pfleging, and Florian Alt (2018). “A survey to understand emotional situations on the road and what they mean for affective automotive UIs”. In: *Multimodal Technologies and Interaction* 2.4, p. 75. ISSN: 2414-4088.

Braun, Michael, Jonas Schubert, et al. (2019). “Improving driver emotions with affective strategies”. In: *Multimodal Technologies and Interaction* 3.1, p. 21. ISSN: 2414-4088.

Braunagel, Christian, Wolfgang Rosenstiel, and Enkelejda Kasneci (2017). “Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness”. In: *IEEE Intelligent Transportation Systems Magazine* 9.4, pp. 10–22. ISSN: 1939-1390.

Breyer, Bianka and Matthias Bluemke (2016). “Deutsche version der positive and negative affect schedule PANAS (GESIS panel)”. In.

Bruce, Vicki (1993). “What the human face tells the human mind: some challenges for the robot-human interface”. In: *Advanced Robotics* 8.4, pp. 341–355. ISSN: 0169-1864. DOI: 10.1163/156855394X00149.

Castro, Marisol, C. Angelo Guevara, and Angel Jimenez-Molina (2020). “A methodological framework to incorporate psychophysiological indicators into transportation modeling”. In: *Transportation research part C: emerging technologies* 118, p. 102712. ISSN: 0968-090X. DOI: 10.1016/j.trc.2020.102712.

Ceaparu, Irina (2004). “Social and psychological influences on computer user frustration”. In: *Media access: Social and psychological dimensions of new technology use*, p. 91. ISSN: 08058411.

Chaplin, Tara M. (2015). “Gender and Emotion Expression: A Developmental Contextual Perspective”. In: *Emotion review* 7.1, pp. 14–21. ISSN: 1754-0739. DOI: 10.1177/1754073914544408.

Chellappa, Ramnath K. and Raymond G. Sin (2005). “Personalization versus privacy: An empirical examination of the online consumer’s dilemma”. In: *Information technology and management* 6.2, pp. 181–202. ISSN: 1573-7667.

Cohn, J. F. et al. (2002). “Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification”. In: *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pp. 491–496. DOI: 10.1109/ICMI.2002.1167045.

Demaree, Heath A. et al. (2005). “Brain lateralization of emotional processing: historical roots and a future incorporating “dominance””. In: *Behavioral and cognitive neuroscience reviews* 4.1, pp. 3–20. ISSN: 1534-5823.

Ding, Yi et al. (2020). “Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning”. In: *Ergonomics* 63.7, pp. 896–908. DOI: 10.1080/00140139.2020.1759699.

DMello, Sidney K. and Brandon M. Booth (2022). “Affect Detection from Wearables in the “Real” Wild: Fact, Fantasy, or Somewhere Inbetween?” In: *IEEE Intelligent Systems* 37, p. 6.

DMello, Sidney K. and Jacqueline Kory (2015). "A Review and Meta-Analysis of Multimodal Affect Detection Systems". In: *ACM Computing Surveys* 47.3, pp. 1–36. ISSN: 0360-0300. DOI: 10.1145/2682899.

Dupré, Damien et al. (2020). "A performance comparison of eight commercially available automatic classifiers for facial affect recognition". In: *PLOS ONE* 15.4.

Egger, Maria, Matthias Ley, and Sten Hanke (2019). "Emotion Recognition from Physiological Signal Analysis: A Review". In: *Electronic Notes in Theoretical Computer Science* 343, pp. 35–55. ISSN: 15710661. DOI: 10.1016/j.entcs.2019.04.009.

Ekman, Paul (1992). "An argument for basic emotions". In: *Cognition & Emotion* 6.3-4, pp. 169–200. ISSN: 0269-9931.

Ekman, Paul and Wallace V. Friesen (1978). "Facial action coding system". In: *Environmental Psychology & Nonverbal Behavior*.

Ekman, Paul, Wallace V. Friesen, et al. (1987). "Universals and cultural differences in the judgments of facial expressions of emotion". In: *Journal of personality and social psychology* 53.4, p. 712. ISSN: 1939-1315.

Fan, Jing et al. (2018). "EEG-Based Affect and Workload Recognition in a Virtual Driving Environment for ASD Intervention". In: *IEEE transactions on bio-medical engineering* 65.1, pp. 43–51. DOI: 10.1109/TBME.2017.2693157.

Glatz, Christiane et al. (2018). "Use the Right Sound for the Right Job". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Ed. by Regan Mandryk et al. New York, NY, USA: ACM, pp. 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3174046.

Grafsgaard, Joseph F. et al. (2013). "Automatically Recognizing Facial Indicators of Frustration: A Learning-centric Analysis". In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, pp. 159–165. ISBN: 978-0-7695-5048-0. DOI: 10.1109/ACII.2013.33.

Grippenkoven, Jan et al. (2018). "Perceived safety: A necessary precondition for successful autonomous mobility services". In: *Proceedings of the human factors and ergonomics society Europe*.

Gross, James J. (1998). "The emerging field of emotion regulation: An integrative review". In: *Review of general psychology* 2.3, pp. 271–299. ISSN: 1089-2680.

— (2008). "Emotion and emotion regulation: Personality processes and individual differences". In: *15938583*. ISSN: 15938583.

Hancock, G. M., P. A. Hancock, and C. M. Janelle (2012). "The impact of emotions and predominant emotion regulation technique on driving performance". In: *Work* 41.Supplement 1, pp. 3608–3611. ISSN: 1051-9815.

Harris, Helen and Clifford Nass (2011). "Emotion regulation for frustrating driving contexts". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Ed. by Desney Tan et al. New York, NY, USA: ACM, pp. 749–752. ISBN: 9781450302289. DOI: 10.1145/1978942.1979050.

Herrenkind, Bernd et al. (2019). "Investigating end-user acceptance of autonomous electric buses to accelerate diffusion". In: *Transportation Research Part D: Transport and Environment* 74, pp. 255–276. ISSN: 13619209. DOI: 10.1016/j.trd.2019.08.003.

Hoemann, Katie et al. (2020). “Context-aware experience sampling reveals the scale of variation in affective experience”. In: *Scientific reports* 10.1, pp. 1–16. ISSN: 2045-2322.

Hoque, Mohammed and Rosalind W. Picard (2011). “Acted vs. natural frustration and delight: Many people smile in natural frustration”. In.

Huang, Dong et al. (2012). “Asymmetric Spatial Pattern for EEG-based emotion detection”. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–7. ISBN: 978-1-4673-1490-9. DOI: 10.1109/IJCNN.2012.6252390.

Ihme, Klas, Stefan Bohmann, et al. (2021). “Development and Evaluation of a Data Privacy Concept for a Frustration-Aware In-Vehicle System”. In: *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY, USA: ACM, pp. 201–208. ISBN: 9781450380638. DOI: 10.1145/3409118.3475147.

Ihme, Klas, Christina Dömeland, et al. (2018). “Frustration in the face of the driver: A simulator study on facial muscle activity during frustrated driving”. In: *Interaction Studies* 19.3, pp. 487–498. ISSN: 1572-0373.

Ihme, Klas, Anirudh Unni, et al. (2018). “Recognizing frustration of drivers from face video recordings and brain activation measurements with functional near-infrared spectroscopy”. In: *Frontiers in human neuroscience* 12, p. 327. ISSN: 1662-5161.

Jeon, M. and B. N. Walker (2011). “What to detect?: Analyzing Factor Structures of Affect in Driving Contexts for an Emotion Detection and Regulation System”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55.1, pp. 1889–1893. ISSN: 1541-9312. DOI: 10.1177/1071181311551393.

Jeon, Myounghoon (2012). “A systematic approach to using music for mitigating affective effects on driving performance and safety”. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp ’12*. New York, New York, USA: ACM Press. DOI: 10.1145/2370216.2370455.

— (2015). “Towards affect-integrated driving behaviour research”. In: *Theoretical Issues in Ergonomics Science* 16.6, pp. 553–585. ISSN: 1463-922X.

— (2016). “Don’t cry while you’re driving: Sad driving is as bad as angry driving”. In: *International Journal of Human–Computer Interaction* 32.10, pp. 777–790. ISSN: 1044-7318.

Jeon, Myounghoon and Wei Zhang (2013). “Sadder but Wiser? Effects of Negative Emotions on Risk Perception, Driving Performance, and Perceived Workload”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57.1, pp. 1849–1853. ISSN: 1541-9312. DOI: 10.1177/1541931213571413.

Kapoor, Ashish, Winslow Burleson, and Rosalind W. Picard (2007). “Automatic prediction of frustration”. In: *International journal of human-computer studies* 65.8, pp. 724–736. ISSN: 1071-5819.

Karaduman, O. et al. (2013). “An effective variable selection algorithm for Aggressive/Calm Driving detection via CAN bus”. In: *2013 International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, pp. 586–591. ISBN: 978-1-4799-2491-2. DOI: 10.1109/ICCVE.2013.6799859.

Keller, Christine et al. (2019). “Understanding the Usefulness and Acceptance of Adaptivity in Smart Public Transport”. In: *HCI in Mobility, Transport, and Automotive Systems*. Ed. by Heidi Kromker. Cham: Springer, pp. 307–326. ISBN: 978-3-030-22666-4.

Klein, Jonathan, Youngme Moon, and Rosalind W. Picard (1999). "This computer responds to user frustration". In: *CHI '99 extended abstracts on Human factors in computing systems - CHI '99*. Ed. by Michael E. Atwood. New York, New York, USA: ACM Press, p. 242. ISBN: 1581131585. DOI: 10.1145/632716.632866.

Kleinsmith, Andrea and Nadia Bianchi-Berthouze (2012). "Affective body expression perception and recognition: A survey". In: *IEEE transactions on affective computing* 4.1, pp. 15–33. ISSN: 1949-3045.

Kosch, Thomas et al. (2020). "Emotions on the Go". In: *Proceedings of the International Conference on Advanced Visual Interfaces*. Ed. by Genny Tortora, Giuliana Vitiello, and Marco Winckler. New York, NY, USA: ACM, pp. 1–9. ISBN: 9781450375351. DOI: 10.1145/3399715.3399928.

Krüger, Sandra et al. (2021). "In-Vehicle Frustration Mitigation via Voice-User Interfaces – A Simulator Study". In: *HCI International 2021 - Posters*. Ed. by Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa. Vol. 1421. Communications in Computer and Information Science. Cham: Springer International Publishing, pp. 241–248. ISBN: 978-3-030-78644-1. DOI: 10.1007/978-3-030-78645-8_30.

Kukkala, Vipin Kumar et al. (2018). "Advanced Driver-Assistance Systems: A Path Toward Autonomous Vehicles". In: *IEEE Consumer Electronics Magazine* 7.5, pp. 18–25. ISSN: 2162-2248. DOI: 10.1109/MCE.2018.2828440.

Landis, J. Richard and Gary G. Koch (1977). "The measurement of observer agreement for categorical data". In: *biometrics*, pp. 159–174. ISSN: 0006-341X.

Lazarus, Richard S. (1991). "Progress on a cognitive-motivational-relational theory of emotion". In: *American psychologist* 46.8, p. 819. ISSN: 1935-990X.

Lee, Y. C., ed. (2010). *Measuring drivers' frustration in a driving simulator*. Vol. 19. Sage Publications Sage CA: Los Angeles, CA. ISBN: 1541-9312.

Li, Wenbo et al. (2021). "A Spontaneous Driver Emotion Facial Expression (DEFE) Dataset for Intelligent Vehicles: Emotions Triggered by Video-audio Clips in Driving Scenarios". In: *IEEE transactions on affective computing*, p. 1. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2021.3063387.

Linxen, Sebastian et al. (2021). "How WEIRD is CHI?" In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Ed. by Yoshifumi Kitamura et al. New York, NY, USA: ACM, pp. 1–14. ISBN: 9781450380966. DOI: 10.1145/3411764.3445488.

Liu, Shu et al. (2021). "The empathetic car: Exploring emotion inference via driver behaviour and traffic context". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5.3, pp. 1–34. ISSN: 2474-9567.

Löcken, Ihme, and Unni (2017). "Towards Designing Affect-Aware Systems for Mitigating the Effects of In-Vehicle Frustration". In: *Boll, Löcken et al. (Hg.) 2017 – Proceedings of the 9th International*, pp. 88–93. DOI: 10.1145/3131726.3131744.

Malta, Lucas et al. (2010). "Analysis of real-world driver's frustration". In: *IEEE Transactions on Intelligent Transportation Systems* 12.1, pp. 109–118. ISSN: 1524-9050.

Maryam Fakhrhosseini, Seyedeh and Myounghoon Jeon (2017). "Affect/Emotion Induction Methods". In: *Emotions and Affect in Human Factors and Human-Computer Interaction*.

Elsevier, pp. 235–253. ISBN: 9780128018514. DOI: 10.1016/B978-0-12-801851-4.00010-0.

McCuaig, Judi, Mike Pearlstein, and Andrew Judd (2010). "Detecting Learner Frustration: Towards Mainstream Use Cases". In: *Intelligent Tutoring Systems*. Ed. by David Hutchison et al. Vol. 6095. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 21–30. ISBN: 978-3-642-13436-4. DOI: 10.1007/978-3-642-13437-1_3.

McDuff, Daniel and Mary Czerwinski (2018). "Designing emotionally sentient agents". In: *Communications of the ACM* 61.12, pp. 74–83. ISSN: 0001-0782.

Mehrabian, Albert and James A. Russell (1974). *An approach to environmental psychology*. the MIT Press. ISBN: 0262130904.

Mesken, Jolieke et al. (2007). "Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports,(observed) behaviour, and physiology". In: *Transportation research part F: traffic psychology and behaviour* 10.6, pp. 458–475. ISSN: 1369-8478.

Millonig, Alexandra and Peter Fröhlich (2018). "Where Autonomous Buses Might and Might Not Bridge the Gaps in the 4 A's of Public Transport Passenger Needs". In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY, USA: ACM, pp. 291–297. ISBN: 9781450359467. DOI: 10.1145/3239060.3239079.

Nass, Clifford et al. (2005). "Improving automotive safety by pairing driver emotion and car voice emotion". In: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. Ed. by Gerrit van der Veer and Carolyn Gale. New York, NY, USA: ACM, pp. 1973–1976. ISBN: 1595930027. DOI: 10.1145/1056808.1057070.

Nesbit, Sundé M., Judith C. Conger, and Anthony J. Conger (2007). "A quantitative review of the relationship between anger and aggressive driving". In: *Aggression and Violent Behavior* 12.2, pp. 156–176. ISSN: 1359-1789.

Noroozi, Fatemeh et al. (2018). "Survey on emotional body gesture recognition". In: *IEEE transactions on affective computing*. ISSN: 1949-3045.

Oehl, Michael et al. (2019). "Towards a frustration-aware assistant for increased in-vehicle UX". In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*. Ed. by Christian P. Janssen et al. New York, NY, USA: ACM, pp. 260–264. ISBN: 9781450369206. DOI: 10.1145/3349263.3351518.

Ong, Desmond C. et al. (2021). "Modeling emotion in complex stories: the Stanford Emotional Narratives Dataset". In: *IEEE transactions on affective computing* 12.3, pp. 579–594. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2019.2955949.

Orwell, George (1954). *Nineteen eighty-four : a novel*. Penguin in association with Secker & Warburg Harmondsworth. ISBN: 9780241453513.

Palmiero, Massimiliano and Laura Piccardi (2017). "Frontal EEG asymmetry of mood: A mini-review". In: *Frontiers in Behavioral Neuroscience* 11, p. 224. ISSN: 1662-5153.

Paschalidis, Evangelos, Charisma F. Choudhury, and Stephane Hess (2019). "Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour". In: *Analytic methods in accident research* 22, p. 100089. ISSN: 2213-6657.

Picard, Rosalind W. and Jonathan Klein (2002). "Computers that recognise and respond to user emotion: theoretical and practical implications". In: *Interacting with Computers* 14.2, pp. 141–169. ISSN: 0953-5438. DOI: 10.1016/s0953-5438(01)00055-8.

Precht, Lisa, Andreas Keinath, and Josef F. Krems (2017). "Effects of driving anger on driver behavior–Results from naturalistic driving data". In: *Transportation research part F: traffic psychology and behaviour* 45, pp. 75–92. ISSN: 1369-8478.

Reuderink, Boris, Christian Mühl, and Mannes Poel (2013). "Valence, arousal and dominance in the EEG during game play". In: *International journal of autonomous and adaptive communications systems* 6.1, pp. 45–62. ISSN: 1754-8632.

Sangineto, Enver et al. (2014). "We are not All Equal". In: *Proceedings of the 22nd ACM international conference on Multimedia*. Ed. by Kien A. Hua. ACM Digital Library. New York, NY: ACM, pp. 357–366. ISBN: 9781450330633. DOI: 10.1145/2647868.2654916.

Scherer, Klaus R. (2009). "Emotions are emergent processes: they require a dynamic computational architecture". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535, pp. 3459–3474. ISSN: 0962-8436.

Scherer, Klaus R. and Agnes Moors (2019). "The emotion process: Event appraisal and component differentiation". In: *Annual review of psychology* 70.1, pp. 719–745.

Schmidt, Philip et al. (2019). "Wearable-based affect recognition—A review". In: *Sensors* 19.19, p. 4079.

Schömig, Nadja et al. (2018). "Experimental Induction and Measurement of Negative Affect Induced by Interacting with In-vehicle Information Systems". In: *Human-Computer Interaction. Theories, Methods, and Human Issues*. Ed. by Masaaki Kurosu. Vol. 10901. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 441–452. ISBN: 978-3-319-91237-0. DOI: 10.1007/978-3-319-91238-7_36.

Schuster, Timo Jens (2014). "Elektroenzephalographische Erfassung des emotionalen Nutzerzustands in der simulierten Mensch-Maschine Interaktion". PhD thesis. Universität Ulm.

Shinar, David (1998). "Aggressive driving: the contribution of the drivers and the situation". In: *Transportation research part F: traffic psychology and behaviour* 1.2, pp. 137–160. ISSN: 1369-8478. DOI: 10.1016/s1369-8478(99)00002-9.

Sidney, K. D. et al., eds. (2005). *Integrating affect sensors in an intelligent tutoring system*.

Smith, Ezra E. et al. (2017). "Assessing and conceptualizing frontal EEG asymmetry: An updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry". In: *International Journal of Psychophysiology* 111, pp. 98–114. ISSN: 0167-8760.

Snowden, Edward (2019). *Permanent record*. Pan Macmillan. ISBN: 9788793772137.

Song, Meishu, Adria Mallol-Ragolta, et al. (2021). "Frustration recognition from speech during game interaction using wide residual networks". In: *Virtual Reality & Intelligent Hardware* 3.1, pp. 76–86. ISSN: 20965796. DOI: 10.1016/j.vrih.2020.10.004.

Song, Meishu, Zijiang Yang, et al. (2019). "Audiovisual Analysis for Recognising Frustration during Game-Play: Introducing the Multimodal Game Frustration Database". In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, pp. 517–523. ISBN: 978-1-7281-3888-6. DOI: 10.1109/ACII.2019.8925464.

Steck, Felix et al. (2018). "How Autonomous Driving May Affect the Value of Travel Time Savings for Commuting". In: *Transportation Research Record: Journal of the Transportation Research Board* 2672.46, pp. 11–20. ISSN: 0361-1981. DOI: 10.1177/0361198118757980.

Stephan, Achim (2015). "Empathy for Artificial Agents". In: *International Journal of Social Robotics* 7.1, pp. 111–116. ISSN: 1875-4791. DOI: 10.1007/s12369-014-0260-0.

Tingvall, Claes (1997). "The Zero Vision". In: *Transportation, Traffic Safety and Health.* Ed. by Hans von Holst, Åke Nygren, and Roland Thord. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 37–57. ISBN: 978-3-662-03411-8. DOI: 10.1007/978-3-662-03409-5_4.

Wallbott, Harald G. (1998). "Bodily expression of emotion". In: *European journal of social psychology* 28.6, pp. 879–896. ISSN: 0046-2772.

Wilms, R., R. Lanwehr, and A. Kastenmüller (2021). "Always look on the bright side of life? Exploring the between-variance and within-variance of emotion regulation goals". In: *Motivation and Emotion* 45.2, pp. 235–247. ISSN: 1573-6644. DOI: 10.1007/s11031-021-09867-5. URL: https://link.springer.com/article/10.1007/s11031-021-09867-5.

Yang, Wenlu et al. (2018). "Physiological-Based Emotion Detection and Recognition in a Video Game Context". In: *2018 International Joint Conference on Neural Networks (IJCNN).* IEEE, pp. 1–8. ISBN: 978-1-5090-6014-6. DOI: 10.1109/IJCNN.2018.8489125.

Zepf, Sebastian, Monique Dittrich, et al. (2019). "Towards Empathetic Car Interfaces". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.* Ed. by Stephen Brewster et al. New York, NY, USA: ACM, pp. 1–6. ISBN: 9781450359719. DOI: 10.1145/3290607.3312883.

Zepf, Sebastian, Javier Hernandez, et al. (2020). "Driver Emotion Recognition for Intelligent Vehicles". In: *ACM Computing Surveys* 53.3, pp. 1–30. ISSN: 0360-0300. DOI: 10.1145/3388790.

Ziefle, Martina, Julian Halbey, and Sylvia Kowalewski (2016). "Users' Willingness to Share Data on the Internet: Perceived Benefits and Caveats". In: *Proceedings of the International Conference on Internet of Things and Big Data.* SCITEPRESS - Science and and Technology Publications, pp. 255–265. ISBN: 978-989-758-183-0. DOI: 10.5220/0005897402550265.