

Discriminating Drivers' Emotions through the Dimension of Power: Evidence from Facial Infrared Thermography and Peripheral Physiological Measurements

Meng Zhang^{*}, meng.zhang@dlr.de, Klas Ihme^{*}, Uwe Drewitz^{*}

^{*} Institute of Transportation Systems, German Aerospace Center, Braunschweig, Germany,

ABSTRACT

The goal of this study was to investigate changes in body temperature as indicators for the emotional dimension of power during driving. Therefore, a driving simulator experiment with 18 participants was conducted, in which two emotions (Fear and noFear) with different characteristics in the dimension of power (low power and high power), which is described as power or weakness feelings of control, were induced using threat and challenge events in the driving scenarios. Infrared thermography video and automatic facial feature recognition were implemented to assess participants' facial temperature. It was revealed that forehead temperature, which is supposed to represent emotional dimension of power, decreased significantly more after threat than after challenge events ($t(17) = -1.83$, $p = .04$, Cohen's $d = 0.54$). These results suggest that forehead temperature as an indicator for the emotional dimension of power can help to measure drivers' fear and thus aid reliable in-vehicle emotion recognition.

Keywords: driver state, infrared thermography, dimension of power, fear, forehead temperature.

1. Introduction

Different negative emotions have different effects on drivers' cognitive capabilities necessary for driving, risk perception and experienced comfort (Jeon, Walker, & Yim, 2014; Jeon, 2015). For example, recent research has shown that fear and anger oppositely affect risk perception of drivers and driving performance (Jeon, Yim, & Walker, 2011; Lu, Xie, & Zhang, 2013; Roidl, Frehse, & Höger, 2014). In addition, fear was found to raise the experienced workload of driving (Jeon, Yim, & Walker, 2011). Moreover, increased fear during driving leads to slower or more cautious driving (Schmidt-Daffy, 2013), while other negative emotions, such as anger, could promote higher risk taking (e.g. frequently crossing yellow lights, (Abdu, Shinar, & Meiran, 2012). Notably, negative emotions do not only affect the human as driver. Even with increasing levels of automation as currently foreseen for the automotive world when the role of the human may not be driving anymore, but could change to monitoring, being the fall back or just a passenger, emotions will still be important for human-vehicle interaction. Emotions that occur when users interact with new technologies influence the usage and acceptance of these systems (Wood & Moreau, 2006). As Hudlicka (2003) points out considering user affect in system design could reduce the frequency of accidents and incidents attributed to human error, but also enhance the user's experience (Hudlicka, 2003). Hence, it is important to reduce negative emotions occurring in traffic and during human-vehicle automation in order to increase safety and acceptance of future (automated) vehicles. Because it is very hard to reduce all potential sources of negative emotions in vehicles and the traffic system as such, one idea is to design affect-aware systems that are able to interpret the user's emotional state

and provide assistance to support humans to balance their emotional state or reduce the negative consequences of certain emotional states (Klein, Moon, & Picard, 2002; Löcken, Ihme, & Unni, 2017; Ihme, Unni, Zhang, Rieger, & Jipp, 2018; Jeon, 2015; Tews, Oehl, Siebert, Höger, & Faasch, 2011). As a prerequisite, future affect-aware vehicles would need a reliable recognition of the current driver state using unobtrusive sensors.

Although some approaches to measure drivers' emotions exist, so far reliable systems for in-vehicle emotion recognition cannot be found on the market. One reason for this is that most studies on automated emotion recognition assume an arousal-valence dimensional model (Russell & Barrett, 1999), which considers emotion as a combination of valence (pleasure-displeasure) and arousal (calm-activity) as theoretical background for assessment of drivers' state (e.g. Schmidt, Decke, & Rasshofer, 2016) and drivers' emotions (e.g. Grimm et al., 2007; Minhadd, Hamid, & Reaz, 2017). For example, Grimm et al. (2007) systematically divided drivers' emotional states based on the valence-arousal model into three types, namely happiness (positive valence and moderate arousal), anger (negative valence and high arousal) and drowsiness (negative valence and low arousal). However, distinguishing different negative emotions in the valence-arousal space is challenging, because some emotions exist that are considered as unpleasant and active, such as fear and anger. However, both fear and anger occur relatively frequent during driving (Mesken, Hagenzieker, Rothengatter, & de Waard, 2007) with differential effects on driving performance (Jeon et al., 2011; Lu et al., 2013; Roidl et al., 2014) and therefore need to be distinguished by affect-aware vehicles in order to provide specific assistance. This suggests that the valence-arousal model may not be sufficiently detailed to serve as basis for emotion recognition of affect-aware systems. Recently, empirical evidence for a more comprehensive dimensional model of emotions was presented (Fontaine, Scherer, Roesch, & Ellsworth, 2007; Gillioz, Fontaine, Soriano, & Scherer, 2016), which proposes that, in addition to valence and arousal, the dimensions of power and novelty are necessary to describe the full space of emotional experience. In this model, the dimension of power represents appraisals of power or weakness with respect to feelings of control. High power refers to the feeling of potency-control or dominance, while low power represents the feeling of submission, as also introduced in the Pleasure-Arousal-Dominance (PAD) emotion model developed by Mehrabian (Mehrabian, 1997). According to these higher dimensional models of emotion, fear, unlike other negative emotions such as anger, is characterized by experience of low control of the situation (see also Lu et al., 2013), so that indicators for the dimension of power could help to differentiate between fear and other negative emotions. Both, Fontaine et al. (2007) and Gillioz et al. (2016) conducted a component analysis on semantic phrases describing emotional experience. From the results of these analyses, it is possible to extract potential indicators, which can be measured with sensors to tap this very aspect of experience. For the dimension of power both papers report that participants attribute "feeling cold" to low power and "feeling hot" to high power. From this, it can be concluded that one potential indicator of low power could be the body temperature of drivers.

Measuring body temperature of participants can be accomplished via different sensors. On the one hand, it is possible to use temperature sensors that are directly applied to the skin of drivers, for example on the finger tips (Kistler, Mariauzouls, & von Berlepsch, 1998). On the other hand, it is possible to assess drivers' temperature with infrared cameras directed on the face, as successfully applied in recent research (Cardone, Pinti, & Merla, 2015). Interestingly, a recent lab study by Cruz-Albarran, Benitez-Rangel, Osornio-Rios and

Morales-Hernandez (2017) revealed that camera-based infrared imaging of the face is promising for emotion recognition in a controlled setting. Clearly, infrared cameras have the advantage of being unobtrusive without the need for applying sensors to the drivers' skin (though admittedly future wearable fitness trackers may also include sensors for finger temperature) and therefore may be easily integrated into driver monitoring systems. Furthermore, in contrast to finger sensors for temperature assessment which are susceptible to motion artefacts (e.g. during steering), a camera as contactless sensor can avoid artefacts caused by movements while driving. However, to the best of our knowledge, no study has so far investigated whether changes in the dimension of power can be assessed using infrared thermography cameras in a realistic driving simulation.

Therefore, the aim of the current study is to investigate whether body temperature is suitable as an indicator for the in-vehicle measurement of the dimension of power and thus eventually aid the differentiation of different negative emotions in a higher dimensional emotional space. For this, we induced different emotional states accompanied by variations in power (a threat event with low power and a challenge event with high power) in a realistic driving simulation and measured participants' facial temperature using infrared thermography. Based on the aforementioned considerations, we assume that a decrease in emotional power following threatening events comes along with a reduction of facial temperature compared to challenging events with high emotional power.

2. Method

2.1. Participant

Eighteen volunteers (four females) with an age range from 22 to 40 years (mean [M] = 27.5 years, standard deviation [SD] = 4.5 years) participated in the study. All of them possessed a valid driving license and had at least two years of driving experience. Participants provided written informed consent to take part in the study and received 15 € plus bonus depending on performance (see below) as financial reimbursement for their participation.

2.2. Set-up, Design and Procedure

The study was accomplished in a driving simulator consisting of three screens and steering wheel as well as gas and brake pedals to control a virtual car in a driving simulation (Virtual Test Drive, Vires, Germany). The driving scenarios took place in an urban setting with one lane per direction and a speed limit of 50 km/h. The two target emotional states (Fear and noFear) with differences in the dimension of power (low power vs. high power) were induced through threat events and challenge events in a within-participants design. As a cover story, participants were told that they had the task to deliver a parcel within seven minutes in order to gain a bonus of 1 € (threat scenario) to 2 € (challenge scenario) per drive (details, see below). Participants had to drive two scenarios per target emotion (presented in random order) always starting with a one-minute drive without any events. After each driving scenario, participants completed self-report questionnaires on their emotional experience during the drive. The scenarios inducing the target emotional states were as follows:

Threat: The two scenarios had a length of ~5 km each with three threat events per scenario. Each threat event involved a crash or almost-crash produced by a vehicle swerving abruptly from the opposite lane (see **Fig.**

1, top), in order to induce fear and the accompanying feeling of low power.

Challenge: The two scenarios had a length of ~6 km creating more challenge in the delivering task to increase participants' motivation. The feeling of high power was supposed to be induced by blocking the road (three events per scenario), for instance through a slow lead vehicle or a traffic jam on both lanes (see **Fig. 1**, bottom).



Fig. 1. Screenshots of the driving simulation during the Threat (top) and Challenge scenario (bottom). In the Threat scenario participants were faced with a car swerving into the ego-lane during an overtaking maneuver. In the Challenge scenario participants were stuck by certain traffic events, such as the truck blocking the road as in this example.

2.3. Self-Report Questionnaires

After each driving scenarios, participants had to complete the Positive and Negative Affect Schedule (PANAS) and the Self-Assessment Manikin (SAM) in order to assess their emotional experience during the drives. Significance of comparisons between the conditions threat and challenge was assessed using a series of paired *t*-tests.

The PANAS (Original: Watson, Clark, & Tellegen, 1988; German version: Krohne, Egloff, Kohlmann, & Tausch, 1996) is composed of 20 adjectives describing ten positive and ten negative emotions, on a Likert scale (1 - very slightly, 2 - a little, 3 - moderately, 4 - quite a bit, 5 - extremely). We focused our analysis on the negative items, especially "scared", which was treated as an indicator, whether or not fear was induced. Furthermore, we also specifically focused on two lower-order factors as described in a previous study

(Mehrabian, 1997): “afraid” (average score of “scared”, “nervous”, “afraid”, “guilty”, “ashamed” and “jittery”) and “upset” (average score of “distressed”, “irritable”, “hostile”, “upset”). These were used because “afraid” was evidenced to be negatively related to dominance/power, while “upset” was solely related to valence and arousal (Mehrabian, 1997).

The SAM (Bradley & Lang, 1994) uses pictures to represent emotional responses on the three dimensions valence (pleasure-displeasure), arousal (calm-activity) and dominance/power (control-out of control). Each dimension is represented by a Likert scale from one to nine.

2.4. Measurements of Emotional Arousal

The measurement of emotional arousal was based on physiological measures: Electrocardiography (ECG) and electrodermal activity. We employed a finger sensor (Heally, SpaceBit, Germany) on the forefinger of the non-dominant hand to assess skin conductance at a sampling rate of 25 Hz during the entire experiment. Skin conductance was downsampled to 10 Hz. Our analysis was focused on skin conductance level (SCL), which represents tonic changes of skin conductance as measure of arousal (Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006). In addition, ECG was measured with three electrodes placed on participants' chest at a sampling rate of 500 Hz. The cardiac interbeat interval (IBI) was extracted using the manufacturer's software HealthLab (SpaceBit, Germany). For comparing the arousal of the threat and challenge events, average SCL and IBI at five seconds before the onset and at ten seconds after the onset of events were considered and compared between the conditions by paired *t*-tests. Because IBI is a parameter that can only be calculated when a new heart beat occurs, we took the IBI values closest to these two time points for the comparisons. The ten seconds after the onset of events were considered, because this duration was one of the most common durations of changing of physiological parameters described in a recent review on the measurement of emotions using physiological parameters (Kreibig, 2010). Hence, this time window of analysis was used for all parameters in this study.

2.5. Measurements of Emotional Power

Measuring of emotional power was accomplished using finger temperature, forehead temperature and nasal tip temperature. The finger temperature (FT) signal was recorded from the same finger sensor as the electrodermal activity with a sampling rate of 25 Hz and also downsampled to 10 Hz. The changing of FT at five seconds before onset of events and at ten seconds after onset of events served as one indicator for power and was compared between the conditions by a *t*-test.

Additionally we employed an infrared thermography camera (Optris PI640, resolution 640*480 pixels, 10 frames per second) in combination with the manufacturer's software (Optris, Germany) to record participants' facial temperature. The thermography camera was placed on the dashboard roughly 1 to 1.2 meter away from participants' faces. In order to acquire average temperature in the areas of interest (AOIs) forehead and nose tip, the following data analysis pipeline was used:

Definition of AOIs: As AOIs we used the forehead (5 x 5 pixels between eyebrows) and the nasal tip (5 x 5 pixels of nasal tip) (cf. Ioannou, Gallese, & Merla, 2014; Abdelrahman, Velloso, Dingler, Schmidt, & Vetere, 2017).

Extraction of screenshots: For each event, screenshots (see **Fig. 2**, left) from the thermography videos were extracted at three time points: after the first 10 seconds of a scenario (to serve as a baseline, see below), five seconds before the events and ten seconds after the events. In total, 503 screenshots of thermography videos (1 lost) were extracted.

Labeling of training data set: 71 baseline screenshots were used to train a Histograms of Oriented Gradients (HOG) based detector for automated extraction of the AOIs. For this, middle face (the area from the outer corners of the left eye to the right eye and from eyebrows to the nasal tip) was manually marked and labelled in these screenshots.

Train and implement face detector: The Dlib library (King, 2009) for the programming language Python was used to learn the labeled middle face. Dlib is a machine learning toolkit, which has been used for the recognition human faces and landmarks on faces in earlier research (Day, 2015; Amos, Ludwiczuk, & Satyanarayanan, 2016). In this study, dlib (version: 19.4.0) was compiled in Python (version: 3.6.6). As face detector classifier, we used a support vector machine (default in Dlib) trained on the 71 labelled screenshots to label the middle face in the remaining 421 screenshots of the thermal videos (11 of these failed and were discarded accordingly). The middle face was recognised and marked in these screenshots (see **Fig. 2**, center).

Extraction of temperature values in the AOIs: Based on the middle faces, the average temperature values in the AOIs forehead (5x5 pixels matrix in the top center of the middle face) and nasal tip (5x5 pixels matrix in the lower center of the middle face) were extracted from the temperature matrix logged by Optris (see **Fig. 2**, right). Based on these we could compare the temperature before the event with the temperature after the events in the respective AOIs using *t*-test.

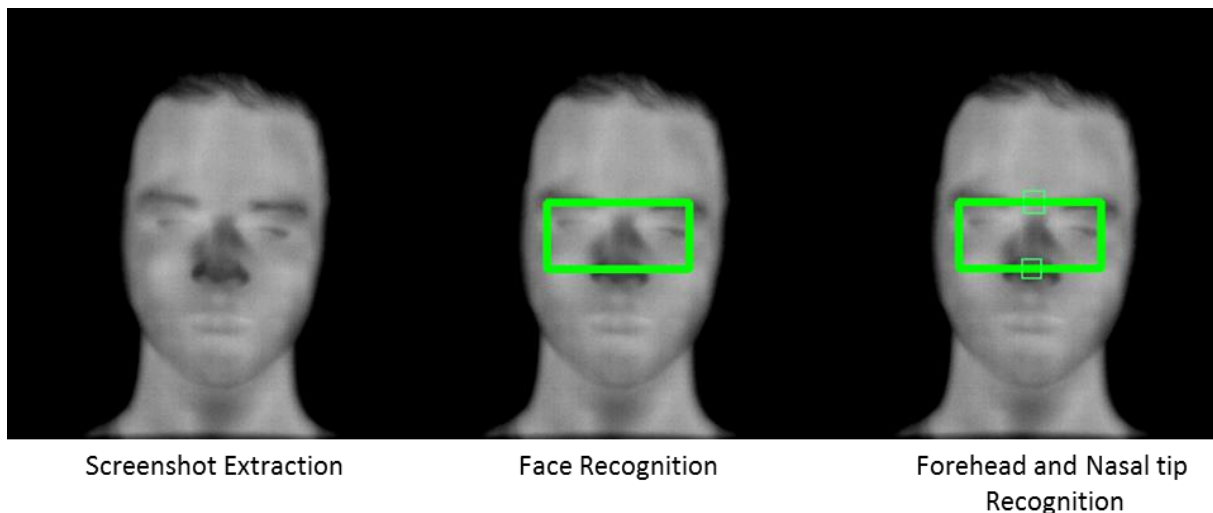


Fig. 2. Procedure for AOIs extraction from infrared thermography screenshots. Original screenshots of infrared thermography of a participant's front face (left) were used for middle face detection (middle) and the detection of forehead and nasal tip (right).

3. Results

3.1 Manipulation Check: Self-Report

Participants' average ratings on the PANAS item "scared" was significantly higher in the threat scenarios

than in the challenge scenarios (for M , SD and results of the t -test, see **Table 1**). In addition, it was revealed that the average score of the lower-order factor “afraid” (negatively associated with dominance, Mehrabian, 1997) was significantly higher in the threat ($M = 1.7$, $SD = 0.5$) than challenge ($M = 1.4$, $SD = 0.4$) ($t(17) = 4.5$, $p < .001$, Cohen’s $d = 0.8$) condition. In contrast, the lower-order factor, “upset”, which is solely associated to valence and arousal (Mehrabian, 1997), did not differ significantly between the two conditions (see **Table 1**).

Table 1

PANAS (partial) Scores in Threat and Challenge Scenarios

	Threat		Challenge		t-test		
	M	SD	M	SD	$t(17)$	p	Cohen’s d
scared	2.7	1.0	1.6	0.5	5.7	<.001***	1.5
ashamed	1.3	0.5	1.1	0.2	2.1	<.05*	0.5
jittery	1.6	0.7	1.3	0.5	2.1	<.05*	0.4
Factor: upset	1.8	0.6	2.1	0.9	-1.8	.09	0.4
Factor: afraid	1.7	0.5	1.4	0.4	4.5	<.001***	0.8

* $p < .05$ significant

*** $p < .001$ significant.

In addition, we exploratorily investigated the difference between all negative affect items in the PANAS (see **Fig. 3**). Except “scared”, significant differences were only found in participants’ ratings for “ashamed” and “jittery”. Participants reported that they felt more “ashamed” and “jittery” in threat scenarios. No significant differences between threat and challenge scenarios were found for all dimensions of the SAM (see **Table 2**).

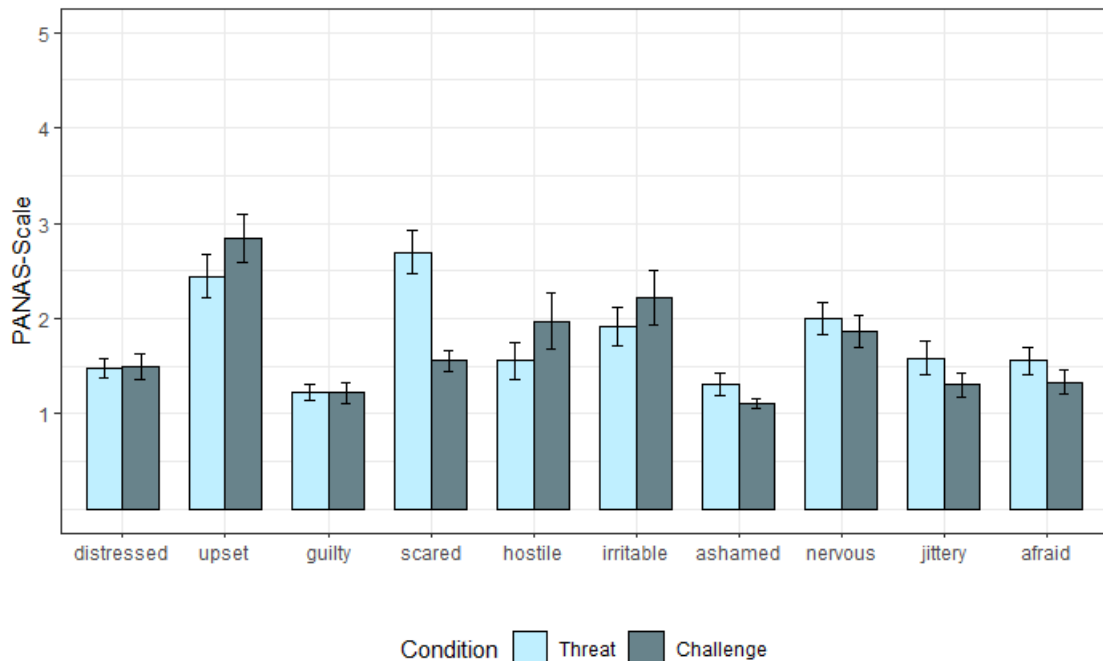


Fig. 3. Participants’ ratings of ten negative affect items of the PANAS scale in the Threat (lightblue) and Challenge (darkblue) scenarios.

Table 2

SAM scores in Threat and Challenge scenarios. Always mean values (*M*), standard deviations (*SD*) as well as the results of the *t*-test are presented.

	Threat		Challenge		t-test		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> (17)	<i>p</i>	Cohen's <i>d</i>
Valence	5.8	1.7	5.3	1.7	1.8	.09	0.3
Arousal	4.2	2.1	4.0	2.2	0.7	.48	0.1
Dominance	5.5	1.3	5.5	1.2	-0.1	.91	0.0

3.2 Manipulation Check: Arousal

Emotional arousal as assessed with SCL and IBI differed between the two conditions:

SCL: **Fig. 4** (left) presents the average of the changing in skin conductance level from five seconds before to ten seconds after events. The skin conductance level increased after threat events ($M = 0.49 \mu S$, $SD = 0.45 \mu S$), while there was barely a changing after challenge events ($M = -0.03 \mu S$, $SD = 0.33 \mu S$). The difference between the changing of SCL between the two condition was significant according to a *t*-test ($t(17) = 4.7$, $p = < .001$, Cohen's $d = 1.32$).

IBI: **Fig. 4** (right) shows the results from the IBI. IBI decreased after threat events ($M = -63.7$ ms, $SD = 48.1$ ms), (which means an increase of heart rate) and decreased after challenge events ($M = 2.8$ ms, $SD = 29.8$ ms). The difference between the changing of IBI was significant according to a *t*-test ($t(17) = 4.7$, $p = < .001$, Cohen's $d = 1.66$).

In addition, there was a significant correlation between SCL and IBI ($r = -.23$, $p < .01$).

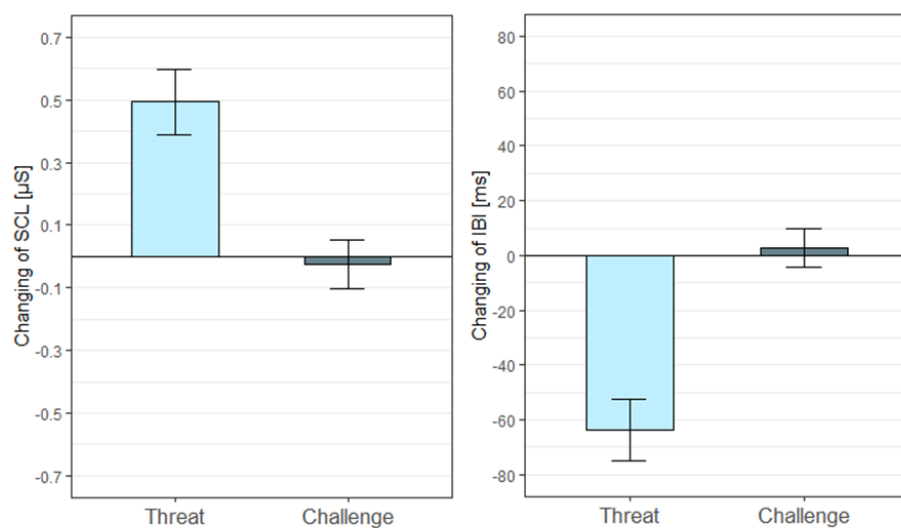


Fig. 4. Changing of skin conductance level (SCL, left) and changing of interbeat interval (IBI, right) after

Threat (lightblue) and Challenge (darkblue) events, where the difference between five seconds before the onset and ten seconds after the onset of events was considered.

3.3 Emotional Power

For the temperature indicators for emotional power, the following results were obtained:

Finger: The average of the changing of finger temperature from five seconds before event onset and ten seconds after event onset was calculated (see **Fig. 5** left). The *t*-test indicates that the changing of finger temperature did not differ between the threat ($M = 0.02\text{ }^{\circ}\text{C}$, $SD = 0.02\text{ }^{\circ}\text{C}$) and challenge scenarios ($M = 0.02\text{ }^{\circ}\text{C}$, $SD = 0.05\text{ }^{\circ}\text{C}$) ($t(17) = 0.22$, $p = .41$, Cohen's $d = 0.08$).

Forehead: **Fig. 5** (middle) presents the average of the changing in forehead temperature from five seconds before to ten seconds after events. The results showed a reduction of forehead temperature after threat events ($M = -0.02\text{ }^{\circ}\text{C}$, $SD = 0.08\text{ }^{\circ}\text{C}$) and an increase after challenge events ($M = 0.02\text{ }^{\circ}\text{C}$, $SD = 0.09\text{ }^{\circ}\text{C}$). The difference between these two changes was significant according to a *t*-test ($t(17) = -1.83$, $p = .04$, Cohen's $d = 0.54$). Additionally, forehead temperature did not correlate significantly with the indicators of arousal, namely SCL ($r = .03$, $p = .66$) and IBI ($r = .05$, $p = .48$).

Nasal tip: **Fig. 5** (right) presents the average of the changing in nasal tip temperature from five seconds before to ten seconds after events. According to the *t*-test ($t(17) = 0.09$, $p = .54$, Cohen's $d = 0.03$), there was no significant difference between the changing after threat events ($M = -0.06\text{ }^{\circ}\text{C}$, $SD = 0.25\text{ }^{\circ}\text{C}$) and challenge events ($M = -0.06\text{ }^{\circ}\text{C}$, $SD = 0.28\text{ }^{\circ}\text{C}$).

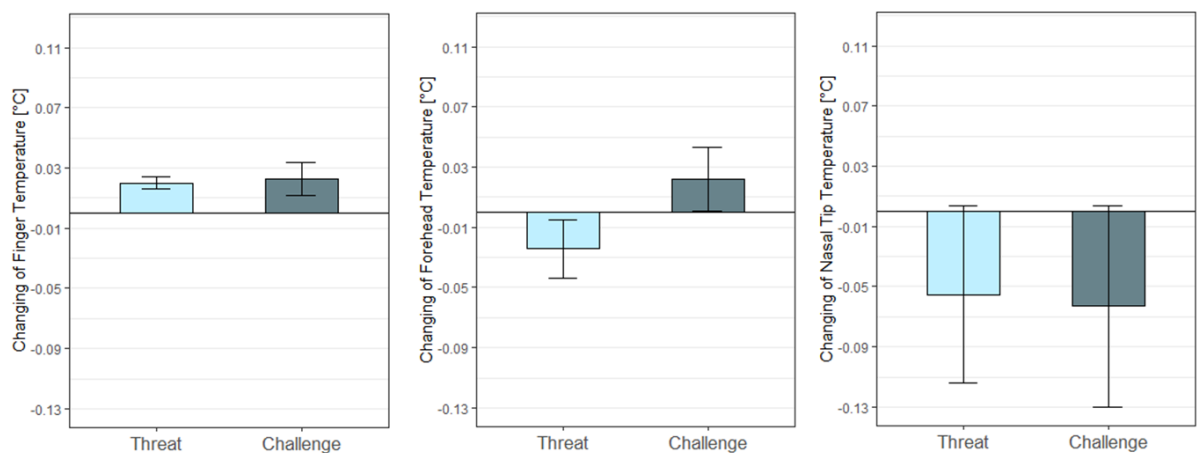


Fig. 5. Changing of temperatures of finger (left), forehead (middle) and nasal tip (right) after Threat (lightblue) and Challenge (darkblue) events, where the difference between five seconds before the onset and ten seconds after the onset of events was considered.

4. Discussion

The goal of this study was to investigate whether body temperature is suitable as an indicator for the in-vehicle measurement of the dimension of power. It was revealed that body temperature can be seen as an

indicator discriminating between emotional driver states with different characteristics in the dimension of power. Specifically, we could show that when experiencing an emotional state with low power, the forehead temperature of drivers measured with an infrared thermography camera is reduced as compared to when experiencing a high power emotional state.

According to self-report and measures of emotional arousal the experimental manipulation was successful. On one hand, the PANAS item "scared" had a higher value in threat scenarios. Besides, the ratings of "ashamed" and "jittery" were significantly higher in the threat scenarios. These two emotional states are located lower on the power dimension in Fontaine's four-dimensional emotional space (Fontaine et al., 2007). Furthermore, the lower-order factor, "afraid", which has been shown to be negatively related to power, was significantly higher in the threat scenario, while the lower-order factor, "upset", which is thought to be solely related to valence and arousal, did not differ between the two conditions. Together, these results provide evidence that not only the induction of fear, but also the experimental manipulation of emotional power was successful. However, evidence for a successful manipulation of the experiment was not found in the SAM, i.e. the subjective ratings in the dominance scale between the threat and challenge scenarios did not differ significantly. This may be due to the fact that the SAM's representativeness of emotional dimensions is still a question at issue (Schmidtke, Schröder, Jacobs, & Conrad, 2014). Specifically, a lack of verbal anchor may hinder the comprehensibility of the pictures of the SAM. On the other hand, SCL and IBI as measures of drivers' emotional arousal differed between the two emotional states. This is in line with previous research showing decreasing IBI (Sinha, 2001) and increasing SCL in fear (Rhudy & Meagher, 2000). Furthermore, the changing of both parameters was negatively correlated. In sum, these results suggest that we successfully induced the emotional state of fear accompanied by the experience of low power.

Our main hypothesis was that body temperature can be used as an indicator for the emotional dimension of power. We could show that forehead temperature significantly decreased after threat events compared with challenge events. However, no significant differences between threat and challenge were revealed for nasal tip and finger temperature. This pattern for fear (no difference for nasal tip and decrease of forehead) is in line with a recent study by Cruz-Albarran, Benitez-Rangel, Osornio-Rios and Morales-Hernandez (2017) who found that nasal temperature does not differentiate between emotional states, while decreases in forehead temperature are discriminative for the experience of fear. Furthermore, the changing of forehead temperature was not correlated with the indicators of emotional arousal (SCL and IBI). Together with assumptions in the semantic studies, in which low power was mostly attributed to a feeling of cold (Fontaine et al., 2007; Gillioz et al., 2016), we can conclude that forehead temperature is a potential indicator for emotional power.

For the measurement of facial temperature, we employed infrared thermography in combination with automated facial landmark detection, which was similarly already employed in a previous study (Abdelrahman et al., 2017). Due to the unobtrusiveness and automated data processing, this greatly enhances the possibility of using facial temperature tracking for emotion recognition in vehicle future automated vehicles. However, the accuracy of recognition is currently limited by the thermographic properties of the face itself. Obviously, the greater the difference in facial feature temperature, the easier the thermographic face could be discriminated

from the background by computer vision. Thus, individual differences in the facial temperature range may affect the face recognition and therefore the usage of this technique. Hence, future work using this technique should consider controlling the experimental set-up, e.g. by keeping the room temperature constant or restraining participants' movements to avoid touching of their face. Still, in our study, we even revealed differences in facial temperature in relation to emotional events in a realistic driving simulator experiment. For that reason, this study extends the results of Cruz-Albarran, Benitez-Rangel, Osornio-Rios and Morales-Hernandez (2017), who showed that assessing emotional states using facial thermography is feasible in a very controlled lab setting, to a very realistic environment. Thus, the current study showed that facial thermography is a promising approach to measure emotions even in rather uncontrolled settings and therefore demonstrates its potential for in-vehicles usage.

As a limitation, it has to be mentioned that our experimental procedure did not only manipulate emotional power between the experimental conditions, but also the level of arousal. Still, the results of the correlation analyses suggest that the effects on forehead temperature revealed in our study are not tapping arousal, but power, because there was no correlation between forehead temperature and the well accepted physiological measures of arousal, SCL and IBI. At this point, it has to be noted though that the number of participants ($N = 18$) was relatively low for a correlational analysis, especially given the generally huge inter-individual differences in physiological measures, which limits the explanatory power of the non-significant correlations. In spite of this, an optimal experimental design for studying emotional power should only induce differences in power, not in valence and arousal. Two negative emotions that suffice these criteria are, as mentioned above, fear and anger, which are both considered as unpleasant and active. Hence, future studies could compare body temperature in relation to fear and anger for extending our results on potential indicators of the emotional dimension of power. A further limitation results from the gender distribution in the sample that included more males than females. Prior research has shown that gender differences in the processing (e.g. Donges, Kersting, & Suslow, 2012)), experience (e.g. Deng, Chang, Yang, Huo, & Zhou, 2016) and expressivity of emotions (e.g. Kring & Gordon, 1998) exist, so that there may also be a difference in emotion-related body temperature changes between the two sexes. Thus, future studies could investigate this effect and its impact for driver monitoring systems in larger participant samples with balanced distribution of males and females. The results of this work suggest that discriminating different negative emotions of drivers in (automated) vehicles could be supported by recording their facial temperature as an indicator of emotional power, which is very important because different negative emotions may have differential effects on behavior and cognition (Jeon et al., 2011). In addition, the results propose that infrared thermography is feasible for acquiring indicators for facial temperature in an unobtrusive way in the vehicle cockpit. This paves the way towards reliable and unobtrusive measurement of drivers' emotions that can be used for the creation of future affect-aware vehicles supporting the drivers in balancing their emotional state or mitigating the negative effects of these. Hence, eventually, this could increase safety as well as acceptance of automated vehicles to promote their usage.

To sum up, drivers' body temperature could indicate variations in the dimension of power and thus support the automated in-vehicle recognition of emotions enabling the parametrization of human-automation interaction according to the current emotional needs of the driver. In the future, a systematic analysis of facial

temperature preferably using automated measures with emotions representing the entire valence-arousal-dominance space could be valuable to validate the results of the current study.

Acknowledgments

This work was supported by the funding initiative Niedersächsisches Vorab of the Volkswagen Foundation and the Niedersächsische Ministerium für Wissenschaft und Kultur (Ministry of Science and Culture of Lower Saxony) as a part of the Interdisciplinary Research Centre on Critical Systems Engineering for Socio-Technical Systems. No other relationships or activities that could appear to have influenced the submitted work.

REFERENCES

- Abdelrahman, Y., Velloso, E., Dingler, T., Schmidt, A., & Vetere, F. (2017). Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Vol. 1, p. 33). <https://doi.org/10.1145/3130898>
- Abdu, R., Shinar, D., & Meiran, N. (2012). Situational (state) anger and driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(5), 575–580. <https://doi.org/10.1016/j.trf.2012.05.007>
- Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). OpenFace: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Cardone, D., Pinti, P., & Merla, A. (2015). Thermal Infrared Imaging-Based Computational Psychophysiology for Psychometrics. *Computational and Mathematical Methods in Medicine*, 2015, 1–8. <https://doi.org/10.1155/2015/984353>
- Cruz-Albarran, I. A., Benitez-Rangel, J. P., Osornio-Rios, R. A., & Morales-Hernandez, L. A. (2017). Human emotions detection based on a smart-thermal system of thermographic images. *Infrared Physics & Technology*, 81, 250–261. <https://doi.org/10.1016/j.infrared.2017.01.002>
- Day, M. (2015). Exploiting Facial Landmarks for Emotion Recognition in the Wild. Presented at the ACM International Conference on Multimodal Interaction (ICMI 2015).
- Deng, Y., Chang, L., Yang, M., Huo, M., & Zhou, R. (2016). Gender Differences in Emotional Response: Inconsistency between Experience and Expressivity. *PLOS ONE*, 11(6), e0158666. <https://doi.org/10.1371/journal.pone.0158666>
- Donges, U.-S., Kersting, A., & Suslow, T. (2012). Women's Greater Ability to Perceive Happy Facial Emotion Automatically: Gender Differences in Affective Priming. *PLoS ONE*, 7(7), e41745. <https://doi.org/10.1371/journal.pone.0041745>
- Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The World of Emotions is not Two-Dimensional. *Psychological Science*, 18(12), 1050–1057. <https://doi.org/10.1111/j.1467-9280.2007.02024.x>

- Gillioz, C., Fontaine, J. R. J., Soriano, C., & Scherer, K. R. (2016). Mapping Emotion Terms into Affective Space: Further Evidence for a Four-Dimensional Structure. *Swiss Journal of Psychology*, 75(3), 141–148. <https://doi.org/10.1024/1421-0185/a000180>
- Grimm, M., Kroschel, K., Harris, H., Nass, C., Schuller, B., Rigoll, G., & Moosmayr, T. (2007). On the Necessity and Feasibility of Detecting a Driver's Emotional State While Driving. In A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction* (Vol. 4738, pp. 126–138). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74889-2_12
- Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction. *International Journal of Human-Computer Studies*, 59(1–2), 1–32. [https://doi.org/10.1016/S1071-5819\(03\)00047-8](https://doi.org/10.1016/S1071-5819(03)00047-8)
- 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. <https://doi.org/10.3389/fnhum.2018.00327>
- Ioannou, S., Gallese, V., & Merla, A. (2014). Thermal infrared imaging in psychophysiology: Potentialities and limits: Thermal infrared imaging in psychophysiology. *Psychophysiology*, 51(10), 951–963. <https://doi.org/10.1111/psyp.12243>
- Jeon, M. (2015). Towards affect-integrated driving behaviour research. *Theoretical Issues in Ergonomics Science*, 16(6), 553–585. <https://doi.org/10.1080/1463922X.2015.1067934>
- 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. <https://doi.org/10.1016/j.trf.2014.04.003>
- Jeon, M., Yim, J.-B., & Walker, B. N. (2011). An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload (pp. 137–142). Presented at the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Salzburg, Austria.
- King, D. E. (2009). Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, 10, 1755–1758.
- Kistler, A., Mariauzouls, C., & von Berlepsch, K. (1998). Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology*, 29(1), 35–41. [https://doi.org/10.1016/S0167-8760\(97\)00087-1](https://doi.org/10.1016/S0167-8760(97)00087-1)
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(2), 119–140.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kring, A. M., & Gordon, A. H. (1998). Sex Differences in Emotion: Expression, Experience, and Physiology. *Journal of Personality and Social Psychology*, 74, 686–70.
- Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). *Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule"(PANAS)* [Data set]. *Diagnostica*. <https://doi.org/10.1037/t49650-000>
- Löcken, A., Ihme, K., & Unni, A. (2017). Towards Designing Affect-Aware Systems for Mitigating the Effects of In-Vehicle Frustration. In *Proceedings of the 9th International Conference on Automotive User Interfaces*

- and Interactive Vehicular Applications Adjunct - AutomotiveUI '17* (pp. 88–93). Oldenburg, Germany: ACM Press. <https://doi.org/10.1145/3131726.3131744>
- Lu, J., Xie, X., & Zhang, R. (2013). Focusing on appraisals: How and why anger and fear influence driving risk perception. *Journal of Safety Research*, 45, 65–73. <https://doi.org/10.1016/j.jsr.2013.01.009>
- Mehrabian, A. (1997). Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression. *Journal of Psychopathology and Behavioral Assessment*, 19(4), 331–357. <https://doi.org/10.1007/BF02229025>
- Mesken, J., Hagenzieker, M. P., Rothengatter, T., & de Waard, D. (2007). Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports, (observed) behaviour, and physiology. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(6), 458–475. <https://doi.org/10.1016/j.trf.2007.05.001>
- Minhad, K. N., Hamid, S., & Reaz, M. B. I. (2017). A Design Framework for Human Emotion Recognition using Electrocardiogram and Skin Conductance Response Signals. *Journal of Engineering Science and Technology*, 12(11), 3102–3119.
- Ravaja, N., Saari, T., Salminen, M., Laarni, J., & Kallinen, K. (2006). Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation. *Media Psychology*, 8(4), 343–367. https://doi.org/10.1207/s1532785xmep0804_2
- Rhudy, J. L., & Meagher, M. W. (2000). Fear and anxiety: divergent effects on human pain thresholds: *Pain*, 84(1), 65–75. [https://doi.org/10.1016/S0304-3959\(99\)00183-9](https://doi.org/10.1016/S0304-3959(99)00183-9)
- Roidl, E., Frehse, B., & Höger, R. (2014). Emotional states of drivers and the impact on speed, acceleration and traffic violations—A simulator study. *Accident Analysis & Prevention*, 70, 282–292. <https://doi.org/10.1016/j.aap.2014.04.010>
- Russell, J. A., & Barrett, L. F. (1999). Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. *Journal of Personality and Social Psychology*, 76(5), 805.
- Schmidt, E., Decke, R., & Raschofer, R. (2016). Correlation between subjective driver state measures and psychophysiological and vehicular data in simulated driving. In *2016 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1380–1385). <https://doi.org/10.1109/IVS.2016.7535570>
- Schmidt-Daffy, M. (2013). Fear and anxiety while driving: Differential impact of task demands, speed and motivation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 16, 14–28. <https://doi.org/10.1016/j.trf.2012.07.002>
- Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods*, 46(4), 1108–1118. <https://doi.org/10.3758/s13428-013-0426-y>
- Sinha, R. (2001). How does stress increase risk of drug abuse and relapse? *Psychopharmacology*, 158(4), 343–359. <https://doi.org/10.1007/s002130100917>
- Tews, T.-K., Oehl, M., Siebert, F. W., Höger, R., & Faasch, H. (2011). Emotional Human-Machine Interaction: Cues from Facial Expressions. In M. J. Smith & G. Salvendy (Eds.), *Human Interface and the Management of Information. Interacting with Information* (pp. 641–650). Springer Berlin Heidelberg.

- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063.
- Wood, S. L., & Moreau, C. P. (2006). From Fear to Loathing? How Emotion Influences the Evaluation and Early Use of Innovations. *Journal of Marketing*, 70(3), 44–57. <https://doi.org/10.1509/jmkg.70.3.44>