Recognizing Behavioral Factors while Driving: A Real-World Multimodal Corpus to Monitor the Driver's Affective State

Alicia Lotz[1], **Klas Ihme**[2], **Audrey Charnoz**[3], **Pantelis Maroudis**[4], **Ivan Dmitriev**[5], **Andreas Wendemuth**[1]

[1] Otto-von-Guericke-Universität Magdeburg, D-39106 Magdeburg, {firstname}.{surname}@ovgu.de

[2] German Aerospace Center, D-38108 Braunschweig, klas.ihme@dlr.de

[3] École polytechnique fédérale de Lausanne, CH-1015 Lausanne, audrey.charnoz@epfl.ch

[4] VALEO Comfort & Driving Assistance Systems, F-93012 Bobigny, pantelis.maroudis@valeo.com

[5] Institut VEDECOM Du Véhicle Décarboné et Communicant et de sa Mobilité, F-78000 Versailles, ivan.dmitriev@vedecom.fr

Abstract

The presented study concentrates on the collection of emotional multimodal real-world in-car audio, video and physiological signal recordings while driving. To do so, three sensor systems were integrated in the car and four relevant emotional states of the driver were defined: neutral, positive, frustrated and anxious. To gather as natural as possible emotional data of the driver, the subjects needed to be unbiased and were therefore kept unaware of the detailed research objective. The emotions were induced using so-called Wizard-of-Oz experiments, where the drivers believed to be interacting with an automated technical system, which in fact was controlled by a human. Additionally, on board interviews while driving were conducted by an instructed psychologist. To evaluate the collected data, questionnaires were filled out by the subjects before, during and after the data collection. These included monitoring of the drivers perceived state of emotion, stress, sleepiness and thermal sensation but also detailed questionnaires on their driving experience, attitude towards technology and big five OCEAN personality traits. Afterwards, the data was annotated by expert labelers. Exemplary results of the evaluation of the experiments are given in the result section of this paper. They indicate that the emotional states were successfully induced and the annotation results are consistent for both performed annotation approaches.

Keywords: in-car emotions, multimodal corpus, multimodal interaction, affective computing, natural emotions

1. Introduction

While driving in a car, the driver can be affected by various emotionally challenging situations. They can either be triggered by the current driving situation, e.g. being cut off by another driver, or caused by a personal event, e.g. receiving good news. On the one hand, emotions can effect the driving behavior in positive and negative ways. By sensing fear, the driver is able to perceive a situation as a possible risk and adapt his driving towards the situation, while anger may lead to an underestimation of the risk level and therefore may increase the risk of causing an accident (Lu et al., 2013). On the other hand, positive as well as negative emotions can influence the driving performance in a negative way (Pêcher et al., 2009; Rhodes and Pivik, 2011; Taubman-Ben-Ari, 2012). Pêcher et al. show that positive emotions (by listening to music) with high intensity lead the participants to drive with risky behavior (e.g. degraded lateral control, sudden speed decreases). In a similar way, Taubman-Ben-Ari shows that positive emotions (with raw intensity) tempt people to drive in a reckless manner. However, positive emotions are rarely considered in terms of road safety, as their occurrence is less common compared to negative emotions (Lewis et al., 2007).

Especially negative emotions such as anger can seriously influence the driving behavior. Already slight provocation can lead to aggressive, violent and hostile driving and can result in road rage. Road rage or aggressive driving is a syndrome of frustration-driven behaviors, enabled by the driver's environment (Shinar, 1998). Frustrating situations, as traffic congestion or delays, can lead to anger emotions in a driving context that conclude in aggressive driving (Shinar, 1998; Zhang et al., 2015). Road rage can be expressed in extenuated ways such as verbal abuse or headlight flashing, but also in more dangerous ways such as traffic weaving, tailgating or aggressive braking (Garase, 2006).

Another negative emotion which should be addressed is fear. This includes anxious drivers who are afraid of driving itself, for example caused by stressful situations while driving, but also anxiety induced by driving in a self-driving/autonomous car. In this study we will focus on the second cause. A survey from the American Automobile Association revealed that 3/4 of U.S. drivers reported to feel afraid to drive in a self-driving car (American Automotive Association, 2017). A reason of this new form of driving anxiety is the fact that recent car functionalities (e.g. adaptive cruise control, lane keeping, self-parking system) take over more and more control from the driver. This transmits a feeling of not being in control of the situation (Koo et al., 2015).

To mitigate this negative safety impact of emotionally affected drivers, we aim to develop a driver monitoring system which detects emotional states of the driver and is able to take over control in critical situations. End users' surveys have shown, that potential users are willing to hand over the control to the car in safety critical situation (e.g. bad weather or road conditions), as well as for comfort reasons (e.g. traffic jam or road works) (Willstrand et al., 2017). 64% of the users were positive to get informed and warned by the car when being in a critical driver state, 16% would even consider a full handover of the control to the car. In the study presented here a special focus will be drawn on the impact of negative emotions. Therefore, a suitable database

is needed, aligned to the needs of this research objective. The presented study will focus on four emotional states of the driver: neutral, positive, frustration and anxiety. For each of these emotional states a specific experiment is designed. The developed system should be able to distinguish between these four types of emotional drivers and draw a conclusion on the capability of the driver to safely interact in road traffic.

2. Related work

There are various studies concentrating on the appearance of affects while driving (e.g. (Grimm et al., 2007a) and (Eyben et al., 2010)). Available datasets are mostly limited to the application of in-car speech recognition but not designed to evaluate the driver's emotional state. These are for example the AV@CAR Spanish multichannel multimodal corpus for in-vehicle automatic audio-visual speech recognition (Ortega et al., 2004) and AVICAR audio-visual speech corpus in car environment (Lee et al., 2004). The number of corpora consisting of natural real-world in-car emotional speech data is still limited, a corpus of multimodal audio-visual and physiological data is yet unknown. Available corpora concentrate either on already existing well known emotional datasets like the Berlin Database of Emotional Speech (Burkhardt et al., 2005), the Danish Emotional Speech corpus (Engbert and Hansen, 1996) or the eNTERFACE'05 Audio-Visual Emotion Database (Martin et al., 2006) and additive real car noises in different conditions as presented in (Grimm et al., 2007b), or focus only on basic modalities such as separate video and audio (Tawari and Trivedi, 2010) or physiological signals (Katsis et al., 2008). Other authors like Abdić et al. and Malta et al. focus on the evaluation of frustration but leave other car-related emotional states disregarded (Abdić et al., 2016; Malta et al., 2011; Ihme et al., in press).

The main disadvantage of most of the presented works is the disregard of in-car acoustics. By superimposing noise, the acoustic characteristic of the car is being left unconsidered. This cannot be compensated by additively overlaying real in-car noise recordings as presented in (Jones and Jonsson, 2007), as the replaying of car noises through stereo speakers differs significantly from real in-car acoustics. In (Lotz et al., 2018) it was shown that it is also not sufficient to replay standard emotional corpora in a real car environment, as the SNR and classification performance can differ significantly for different recording setups of the original dataset. Botinhao et al. investigate the effect of different speaking styles, noise levels and listener age on speech intelligibility. Further, not only different in-car acoustics influence the speech signals' quality significantly but also different types of route taken, weather conditions, background noise in the car and whether the windows are open or closed (Botinhao and Yamagishi, 2017). Furthermore, also the movement of the car needs to be considered. This effects the mounting of all considered sensors and will lead to random noise in the recorded signals.

A big advantage of the developed data set is that all recordings were done in a real-world in-car environment while driving. By not only letting the subject perform small interaction tasks with car assistant systems, which will not lead to pure natural speech, but by also conducting conversation-like interviews, we obtain more natural information considering the driver's speech and facial expressions. By inducing natural emotions, we are able to also consider physiological signals for further analysis.

3. Test Environment

3.1. Test Vehicle and Environment

The test vehicle of the data collection was the research vehicle FASCar (Fischer et al., 2014). The FASCar is a test vehicle for testing driver assistance systems and automated driving functions. It is equipped with a unique steer-by-wire system to support innovative haptic feedback and intervention strategies. For the safety driver/co-driver an additional brake pedal is available (DLR, 2017).

All experiments were carried out on the DLR test ground at the DLR compound in Braunschweig, Germany, which is a designated test ground for driving experiments. As the test vehicle is not permitted to drive on open road, this ensured the most natural driving experience and driving environment for the participants, comparable to road traffic in quiet residential areas. On the site, driving is allowed with a maximum speed of 30 km/h on a fixed driving course (see Figure 1). One round course of roughly 900 meters on the available streets in the site took approx. 2.5 min. To ensure comparability of all recordings, the data was collected during day light and under similar and constant weather conditions. Termination criteria for in-car audio recordings were strong rain and/or thunderstorm. In addition to the participant, two further persons were in the car, one investigator sitting on the passenger seat, and one technician for the supervision of the sensor data recording sitting on the rear bench behind the passenger seat.

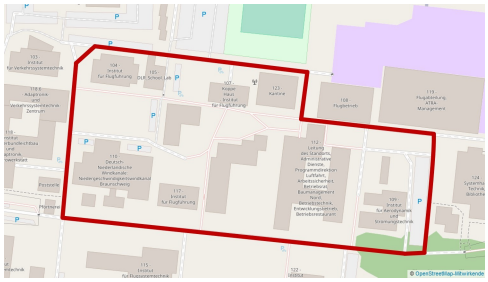

Figure 1: Driving course at DLR compound in Braunschweig, Germany (map taken from `https://www.openstreetmap.de/`).

3.2. Sensor integration

The test vehicle was equipped with a microphone, video and physiological sensing system.

The audio speech stream was recorded using two Shure VP 82 shotgun microphones attached to the dashboard above the steering wheel and close to the right A-pillar using elastic mounting to dampen the car's movement. Additionally, to collect high quality reference recordings, a Sennheiser HSP-4 EW-3 headset microphone was worn by the driver. The microphone tracks were synchronized using a Steinberg UR44 audio interface stored in the trunk of the car. Video images were captured using a Smart Eye Pro

Figure 2: Setup of the microphone and SEP camera system (Smart Eye AB, Gothenburg, Sweden, www.smarteye.se) on the dashboard of the FASCar research vehicle.

(SEP) Multi Camera System (Smart Eye AB, Gothenburg, Sweden, www.smarteye.se) including two high resolution cameras with infrared (IR) filters and active IR illumination attached to the dashboard on both sides of the steering wheel. Peripheral physiological data was recorded using the wireless sensor system Spacebit Heally to measure the electrocardiogram (ECG) and galvanic skin response (GSR). It consists of a finger sensor and a standard 3-lead ECG wearable. The signals are transmitted via Bluetooth to a computer. The sensors integrated onto the dashboard of the test vehicle are depicted in figure 2.
All sensor systems were triggered by a time synchronous signal coming from SEP, to ensure the synchronicity of all systems.

4. Data Characteristics

4.1. Involved Participants

Data was gathered from 30 participants of one age group (25 - 40 years). All of them were native standard German speakers without speaking disorders. For safety reasons, only drivers with a valid driver's license and an annual mileage of at least 5000 km were considered. Further exclusion criteria were: Pregnancy, physical impairment, heart and/or neurological problems, partial or total deafness and medical or alcohol consumption. All participants provided written informed consent to participate in the study and received 30 €as reimbursement for participation. After the study, the participants were fully informed about the goals of the study. The procedure of the study was reviewed and approved by the ethics committee of the Otto-von-Guericke University of Magdeburg, Germany (ref.-number: 153/17).

4.2. Considered Emotional States

This study concentrates only on the driver-relevant states. Therefore, four emotional states occurring frequently while driving a car were defined. These include neutral, positive, frustrated and anxious drivers. It can be assumed that the most commonly occurring emotional state is neutral. To be able to consider a broad range of emotional states and at the same time distinguish between emotions having a highly risky impact on the driving performance, all positive emotions were summarized into the driver state "positive", while the relevant negative emotions were subdivided into "frustrated/angry" and "anxious/fearful". As the collected data will be used to train machine learning algorithms, it is not reasonable to distinguish between separate states of frustration and anger, and anxiety and fear, respectively (Devillers et al., 2005). The characteristic features of the data collected for these states is assumed to be very similar. Therefore, the algorithms will not be able to differentiate between them. The authors of this paper are aware of the fact that from a psychological point of view there is a difference between frustration and anger, and anxiety and fear, but this issue will not be addressed in this paper.
To define these states, the circumplex model of emotions concepts (Russell and Lemay, 2000) was used. This defines the neutral emotional state in a region around the origin of the valence/arousal-axis with moderate arousal and neutral valence, the positive emotional state as all positive expressions with a positive valence, and frustration and anxiety both with high arousal and negative valence (cf. Figure 3). Therefore, to be able to distinguish between frustration and anxiety, additional definitions were concluded. Frustration/anger is defined as the unpleasant feeling occurring in situations in which a person is detained from reaching a desired outcome/goal and anxiety as the unpleasant feeling of dread over anticipated negative events (Lazarus, 1991; Schmidt-Daffy, 2013). According to these definitions, scenarios for each emotional state were designed.

4.3. Emotional Scenarios

The scenarios were designed such that the driving itself would only minimally influence the emotional state of the driver. Participants drove five rounds per scenario. Each scenario started with one round as baseline. The four following rounds depended on the scenario and are described in the remainder of this section. The order of the scenarios was kept constant starting with the neutral scenario followed by positive, frustrated, and anxious. The emotions were induced by conducting experimental studies and interview-like conversations. The experimental studies were assisted by the interviewer reinforcing the situation. In the presented scenarios the co-driver was a trained psychologist who took the role of the interviewer. He did not hesitate to react to the driver's answers and ask follow up questions to keep the conversation alive. Afterwards, the driver was asked to narrate a situation where he felt the considered emotion. By memorizing/narrating an emotional experience, this emotion can be recalled by the driver

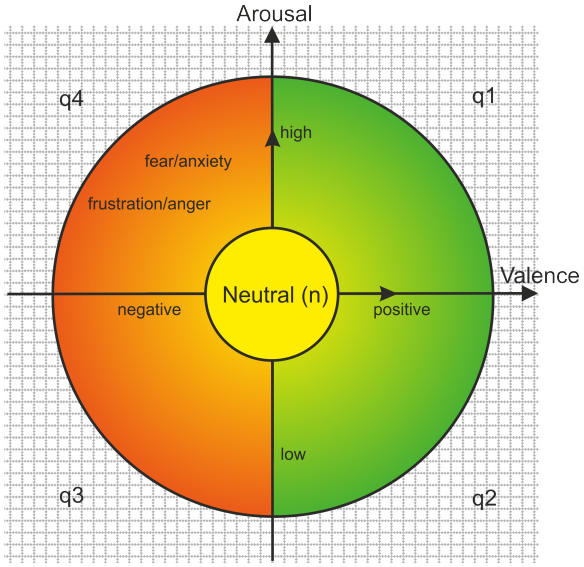

Figure 3: Defined emotional states as areas in the circumplex model of emotions concept by Russell and Lemay (2000).

and is reflected in his facial expressions, speech and physiological signals. This is a commonly used method to induce emotions also used in (Martin et al., 2006), (Amir et al., 2000) and (Forgas, 2002).

neutral: After the baseline round, the investigator initiated a conversation with the participant on a neutral topic (e.g. on educational background, basic personal information, weather etc.) which lasted roughly two rounds. In the last two rounds, the participant drove without conversation to gather baseline information on the facial expression of the driver.

positive: After the baseline round, to not reveal the research objective of the study, the participants were told that a check of the sound quality is necessary and a sound file needed to be played during the following two rounds. This sound file comprised two episodes of a funny radio podcast "Wir sind die Freeses." of the radio station NDR2 (Altenburg, 2017), which is well-known in this region of Germany. In the last two rounds, the investigator started a conversation with the participant on the show, followed up with positive topics (happy situations, holidays) and repeatedly asked the participant to narrate situations in which s/he felt positive/happy.

frustrated/angry: For this scenario, participants were told that the goal of this drive is to evaluate a speech-based navigation system, which was briefly introduced by the investigator before the drive. The participants had the task to enter a certain address and start the navigation. The participants were also told that one of the main innovations of the navigation system was its capability to recognize whether the user is talking to the system or to other people in the car. The navigation system was a mock-up created with MS Power Point, which was controlled by the technician on the rear bench in a Wizard-of-Oz (WOZ)-like scenario. Participants were asked to "think aloud" while interacting with the system, which means that they should utter any thoughts about and experiences with the systems.

The drive started again with one baseline round. Following this, the participant interacted with the navigation system for two rounds. In order to induce frustration, the technician regularly "misunderstood" certain commands of the participant and elicited wrong selections, so that the participant could not finalize the selection for a couple of minutes. After that the investigator initiated a conversation on similarly frustrating experiences, e.g. in interaction with technological systems or while driving. During the course of the conversation (for another two rounds), the investigator encouraged the participant to narrate frustrating situations.

anxious: Again the scenario started with a baseline round. Then, the participant was told that a usability evaluation of an automated brake assistant would follow, in which the participant was asked to "think aloud" again. The brake assistant would only brake at certain locations, namely when the car is about to pass two traffic cones on the side of the road. Before braking, the system would present three audio warning tones. However, similar to the previous scenario, the system was controlled by a WOZ (which was the investigator on the passenger seat who had a second brake pedal). In total, three locations on the track were marked with traffic cones. Initially, the system worked just fine and braked at the marking preceded by the warning tone. Then, however, the WOZ started to brake or play the tone without braking at random locations on the track.
The goal of this procedure was that the participant would pair the tone with the sudden brake. In the following this happened over and over again, so that the participants anticipated the negative event of the brake when the tone started, which created the anxiety of the upcoming abrupt brake. The brake assistant was "active" for two rounds. After this, the investigator initiated a conversation on similar situations or experiences in the car, in which other road participants or assistance systems behaved in a way that created anxiety or uncertainty. The conversation lasted for two rounds.

5. Assessment Methodology

5.1. Questionnaires

Before the data collection, the participants filled out a basic demographic questionnaire including questions about their driving experience. In addition, the participants had to complete the ATI-scale measuring their attitude towards technology (Franke et al., 2017) and the Big Five Inventory (BFI-10) (Rammstedt and John, 2007) to assess the big five OCEAN personality traits: openness on experience (O), conscientiousness (C), extraversion (E), agreeableness (A), neuroticism (N). Moreover, the participants' emotional baseline was assessed using the Self-Assessment Manikin (SAM) (Bradley and Lang, 1994) on the dimensions valence, arousal and dominance as well as the Geneva Emotion Wheel (GEW) (Scherer et al., 2013). For the GEW, the participants were asked to rate all emotions in the wheel (alternative 3). In addition, one item assessing the novelty of the situation was filled in by the participants. After each drive, the participants filled in the GEW plus three dummy questionnaires to camouflage the actual purpose of

the study (Karolinska Sleepiness Scale, (Akerstedt and Gillberg, 1990); Stress Scale, (Dahlgren et al., 2005); Scale of Thermal Sensation, (Gagge et al., 1967)). After all drives, the participants filled out a questionnaire asking detailed questions on their emotional experiences during the drives. These included among others the SAM scales and novelty item as well as free input on their experienced emotions during each drive.

5.2. Ground Truth

The considered emotional states were defined using the circumplex model of emotions concepts (see section 4.2.). By using the valence and arousal level to define the emotions, we could overcome the use of a garbage class in the annotation of the data. We opted for an annotation in two stages, this made the annotation process more controllable for the labelers. The annotation was conducted as followed:

1. Annotation of valence and arousal using the 5-point SAM scale (Bradley and Lang, 1994). The outcome of the labeling was averaged over all labelers.

2. Categorical annotation of emotions into *neutral*, *positive*, *frustrated/angry*, *anxious/fearful* and free space to insert a *different emotional state*.

Additionally to the perceived emotional state, the labelers should also give feedback on how satisfied they were with their decision. From the satisfaction level the reliance of their decision could be concluded, e.g. if they were very dissatisfied with their emotional assessment. This implied that they were very indecisive in their decision and vice versa. To do so, a 5-point Likert-scale (1-Very dissatisfied, 2-Dissatisfied, 3-Unsure, 4-Satisfied, 5-Very satisfied) was used.
For all considered modalities, the data was rated by expert labelers. To receive feasible results by the labelers, at least 3 labelers were employed and their inter-rate reliability of the assessed emotional states was determined using Krippendorf's alpha. Using the inter-rater reliability, we could exclude outliers before averaging the result over the remaining labelers (Siegert et al., 2014).
For acoustic annotation, the ikannotate labeling tool (Böck et al., 2011) was used. The collected audio data was divided into short audio snippets of the same length and annotated afterwards. For the video annotation, the CAPTIV-L2100 software was used. The video signal was annotated in sequences of the same emotional state. All tools were adapted to the above mentioned requirement of the annotation.
The outcome of the audio and video annotation were used as ground truth for the physiological data. Depending on the reliance of the decision of the labelers, either the annotation results of the audio or video data were used.
Additionally to the expert rating, the results of the subjective self-reported feedback forms, filled out by each driver after every emotion scenario, were used as reference value of success for the emotion induction.

6. Study Realization

Before the data collection the participant filled out a self-reported personality questionnaire on the big five OCEAN personality traits and a general questionnaire on personal information. Then the driver was equipped with the physiological sensors (wristband and ECG electrodes on chest), a headset microphone, and the SEP camera system was calibrated. The outfitting of the driver took approx. 20 min. Afterwards, a short period of acclimatization (5 min) was given to the driver to get used to the equipment. During this time, the driver was introduced to the co-driver of the experiment and became acquainted with him. This was important as the co-driver took the role of the investigator in the different emotion scenarios. Therefore, the co-driver needed to be a trained psychologist.
After the acclimatization, the driver got in the car. In total, three persons were present in the car while conducting the experiments: the participant himself, the investigator sitting on the passenger seat, and one technician for the supervision of the sensor data recording sitting on the rear bench behind the passenger seat. The driving scenarios were fixed for all participants as depicted in figure 1.
For each emotional state, the experimental setup was as followed: 2.5 min of baseline driving without distraction of the driver to gather baseline data of the physiological data of the participant; 5 min of driving while conducting a task; 5 min of conversation with the investigator; 2 min filling out questionnaires.
The completion of the different emotional scenarios took in total approx. 10 to 15 min. In between every emotion scenario, the driver had at least 5 minutes of recess to get back to baseline. Depending on the intensity of the driver's emotional state, this time span was extended. At the beginning of each break, the driver was asked to fill out a short subjective self-reported feedback form on his emotional, sleepiness, stress and thermal state.
At the end of the study, the driver was debriefed by the test leader and a special debriefing information sheet was handed out to the participant. This debriefing also included discovering the detailed research objective of the data collection, which was detained from the participant previously. During an interview-like conversation, the driver was asked to give detailed subjective feedback on the recently experienced situations/moments. This was assisted by the interviewer who filled out a specially designed feedback form containing the answers given by the participant. This took approx. 15 min.
The data collection of one participant took approx. 180 min, in total.

7. Exemplary Results

This section will give a broad overview on the results by presenting the audio annotation of one exemplary participant. This participant was a male driver of 29 years. The evaluation of the demographic questionnaire shows a frequent usage of motorized vehicles with little experience in using advanced driver assistance systems. Form the ATI-scale a value of 4.8 was evaluated, which indicates an above-average positive attitude towards technology. From the BFI-10, it was drawn that the participant is strongly open to experience, shows high conscientiousness and high agreeableness. The results of the audio annotation will be presented in the remainder of this section, including a com-

parison with the self-reported feedback gathered from the GEW and detailed end-questionnaire.

7.1. Audio Annotation

For the audio annotation, three female labelers were employed. The inter-rater reliability of all three labelers can be seen in the first row of table 1. It shows an accordance for the labels of 0.27 for the annotation of the valence, 0.21 for annotation of arousal and 0.27 for the categorical annotation. By considering all labelers separately it was noticed, that one of the three labelers showed significant differences in the annotation result. This labeler was also not satisfied with some of the annotation results, while the other labelers never chose "dissatisfied" or "very dissatisfied" as satisfaction level. Therefore, this labeler was excluded for the further analysis of the results. The inter-rater reliability for the remaining two labelers could be increased to 0.39, 0.33 and 0.38, respectively. This is a reasonable result for the annotation of highly natural emotional audio data samples (Siegert et al., 2014).

	IRR		
Labelers	Valence	Arousal	Categories
All	0.27	0.21	0.27
Best	0.39	0.33	0.38

Table 1: Inter-rater reliability of all labelers vs. best two labelers.

In total, the labelers annotated 14 min of speech resulting in 556 speech samples. The labeling process took on average 4.3 hrs. In this time, the samples were annotated into the dimensions valence-arousal, into the categories neural, positive, frustrated/angry and anxious/fearful, and the labelers rated the satisfactory-level of the annotation. By considering the labelers' satisfactory-level, outliers were excluded from the sample-set before averaging the results over the remaining labelers. For the presented participant, the remaining labelers were never "dissatisfied" or "very dissatisfied" with their decision, this is why no samples were excluded from the sample-set.

For the valence-arousal annotation, the annotation results are presented as mapping onto the four quadrants of the circumplex model as pictured in figure 3. To transform the valence-arousal annotation results into these dimensional categories, the annotated SAM scale values (1-5) were averaged over the remaining two labelers. The origin of the two dimensions, in which the region of the neutral emotional state is located, is allocated to the valence and arousal value of 3. The neutral emotional state is defined as the region around the origin, with valence and arousal values lying within the scale (2-4). All values outside of this region are located in "q1", "q2", "q3" and "q4", or on the x/y-axes of the two dimensions. This approach resulted in 297 samples in region "n", 0 in "q1", 43 in "q2", 145 in "q3" and 3 in "q4". From those speech samples lying on the x/y-axes of the two dimensions, only samples with low arousal and neutral valence were recognized (68).

For the annotation of the emotional categories, only those samples annotated consistently by both remaining labelers were evaluated. In case of a high inter-rater reliability for all three labelers, a majority voting would have been carried out. This is not possible in case of two labelers. The results of the labelers was consistent for 263 samples. The corresponding annotation results are: 68 samples labeled as neutral, 48 as positive, 60 as frustrated/angry, 87 as anxious/fearful and none of the speech samples was annotated as different emotional states. The majority of the confusion between samples where the labelers did not find a consistent label was distributed equally between neutral and any other emotional state (202). This is expectable because of the high naturalness of the recorded samples.

		Dimensional					
		n	q1	q2	q3	q4	low
Categorical	Neutral	17	0	3	7	0	41
	Positive	35	0	13	0	0	0
	Anxious/fearful	27	0	1	55	0	4
	Frustrated/angry	48	0	0	9	3	0

Table 2: Confusion matrix of the categorical and dimensional annotation results. The entry "low" for the dimensional annotation denotes those samples annotated with low arousal and neutral valence.

Table 2 shows the confusion matrix of the categorical and dimensional annotation results. The entries marked in green indicate a suitable assignment between the categorical and dimensional annotation. Red entries indicate an unsuitable assignment. This implies a high consistency of the dimensional and categorical annotation in case of high values for green entries and low values for red values, which is given in this case. Entries marked in yellow are confusions of the categorical results with the neutral region of the dimensional approach. This is reasonable, as highly natural emotional samples of low emotional content and expressivity were examined. By optimizing the transformation of the dimensional annotation results, the accuracy of the two annotation approaches can be improved. Because of the high naturalness of the recorded audio samples, the neutral region around the origin of the dimensional approach needs to be reduced, to also be able to track small changes in the valence and arousal level. It is also noticed that those samples categorically annotated as neutral were of low arousal and neutral valence. This is in line with the emotional models presented in (Holzapfel et al., 2002) and (Almeida et al., 2016).

7.2. Evaluation of Questionnaires

The subjective self-reported questionnaires, conducted before and after the experiment, as well as the GEW, filled out by the participant in between each emotional scenario, were used to confirm a successful inducement of emotions. Evaluating the questionnaires and scales, the participant stated to be in a neutral emotional state while conducting the *neutral experiment*, with neutral valence, low arousal and moderate dominance. While conducting the *positive experiment*, he stated to be in a positive mood with moderate arousal and moderate dominance. While conducting the *frustrated experiment*, he stated to be in a negative state of valence, moderate arousal and very low dominance. He verified this statement by mentioning that he felt frustrated

and a bit ashamed. For the *anxious experiment*, the participant stated a negative valence, high arousal and moderate dominance, which was confirmed by his statement of being insecure while conducting the experiment leading to an anxious and confused feeling. From these statements we can assume that the inducement of the emotional states was realized successfully.

8. Conclusion and Outlook

This paper presents the collection of emotional multimodal real-world in-car audio, video and physiological signal recordings. The aim of this study was to be able to recognize behavioral factors of drivers while driving in a car, focusing on the emotional state of the driver. By conducting the presented experiments, it is possible to obtain multimodal, essentially natural emotional data, which enables a monitoring of the driver's emotional state. This could also be confirmed by the annotation results and participants' subjective feedback on the conducted experiments. This was exemplarily presented by evaluating the results of one randomly chosen participant of the study. It could be shown that the outcome of the two presented annotation approaches are consistent in their results and indicate a high naturalness of the annotated speech samples.

Description of the full data set, and corresponding annotation statistics, will be reported in a forthcoming paper. This will also include results of usability of the data for developing monitoring systems for mitigating the negative safety impact of emotionally affected drivers.

9. Acknowledgment

This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 68890.

10. Bibliographical References

Abdić, I., Fridman, L., McDuff, D., Marchi, E., Reimer, B., and Schuller, B. (2016). Driver Frustration Detection From Audio and Video in the Wild. In *Proc. of the IJCAI-2016*, pages 1354–1360, New York, NY, USA.

Akerstedt, T. and Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1–2):29–37.

Almeida, P. R., Ferreria-Santos, F., Chaves, P. L., Paiva, T. O., Barbosa, F., and Marques-Teixeira, J. (2016). Perceived arousal of facial expressions of emotion modulates the N170, regardless of emotional category: Time domain and time-frequency dynamics. *International Journal of Psychophysiology*, 99:48–56.

Altenburg, A. (2017). Wir sind die Freeses. http://www.ndr.de/ndr2/wir_sind_die_freeses/podcast4250.html.

American Automotive Association. (2017). Americans Feel Unsafe Sharing the Road with Fully Self-Driving Cars. http://newsroom.aaa.com/2017/03/americans-feel-unsafe-sharing-road-fully-self-driving-cars/.

Amir, N., Ron, S., and Laor, N. (2000). Analysis of an emotional speech corpus in Hebrew based on objective criteria. In *ITRW on Speech and Emotion*, pages 29–33, Newcastle, Northern Ireland, UK.

Böck, R., Siegert, I., Haase, M., Lange, J., and Wendemuth, A. (2011). ikannotate - A Tool for Labelling, Transcription, and Annotation of Emotionally Coloured Speech. In Sidney D'Mello, et al., editors, *Proc. of the ACII-2011*, volume 6974 of *LNCS*, pages 25–34, Memphis, TN, USA. Springer Verlag Berlin, Germany.

Botinhao, C. V. and Yamagishi, J. (2017). Speech intelligibility in cars: the effect of speaking style, noise and listener age. In *Proc. of the INTERSPEECH-2017*, pages 2944–2948, Stockholm, Sweden.

Bradley, M. M. and 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.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., and Weiss, B. (2005). A Database of German Emotional Speech. In *Proc. of the INTERSPEECH-2005*, pages 1517–1520, Lisbon, Portugal.

Dahlgren, A., Kecklund, G., and Akerstedt, T. (2005). Different levels of work-related stress and the effects on sleep, fatigue and cortisol. *Scandinavian Journal of Work, Environment and Health*, 31(4):277–285.

Devillers, L., Vidrascu, L., and Lamel, L. (2005). Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18(4):407–422.

DLR. (2017). FASCar - test vehicle for assistance and automation. http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-11367/19950_read-46557/.

Engbert, I. S. and Hansen, A. V. (1996). Documentation of the Danish Emotional Speech Database DES. Technical report, Center for Person Kommunikation, Aalborg University, Denmark.

Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B., and Nguyen-Thien, N. (2010). Emotion on the Road - Necessity, Acceptance, and Feasibility of Affective Computing in the Car. *Advances in Human-Computer Ineraction*, 2010:s.p.

Fischer, M., Richter, A., Schindler, J., Plättner, J., Temme, G., Kelsch, J., Assmann, D., and Köster, F. (2014). Modular and Scalable Driving Simulator Hardware and Software for the Development of Future Driver Assistence and Automation Systems. In *New Developments in Driving Simulation Design and Experiments*, Driving Simulation Conference, pages 223–229. Paris, France.

Forgas, J. P. (2002). Feeling and Doing: Affective Influences on Interpersonal Behavior. *Psychological Inquiry - An International Journal for the Advancement of Psychological Theory*, 13(1):1–28.

Franke, T., Attig, C., and Wessel, D. (2017). Affinity for technology interaction - a personal-resource perspective. In *Proc. of HFES Europe-2017*, Rome, Italy.

Gagge, A. P., Stolwijk, J. A. J., and Hardy, J. D. (1967). Comfort and thermal sensations and associated physiological responses at various ambient temperatures. *Environmental Research*, 1(1):1–20.

Garase, M. L. (2006). *Road Rage*. LFB Scholarly Publishing LLC.

Grimm, M., Kroschel, K., Harris, H., Nass, C., Schuller,

B., Rigoll, G., and Moosmayr, T. (2007a). On the Necessity and Feasibility of Detecting a Drivers Emotional State While Driving. In *Proc. of the ACII-2007*, volume 4738 of *LNCS*, pages 126–138, Lisbon, Protugal. Springer, Berlin, Heidelberg.

Grimm, M., Kroschel, K., Schuller, B., Rigoll, G., and Moosmayr, T. (2007b). Acoustic Emotion Recognition in Car Environment Using a 3D Emotion Space Approach. In *Proc. of the DAGA-2007*, Stuttgart, Germany.

Holzapfel, H., Fuegen, C., Denecke, M., and Waibel, A. (2002). Integrating Emotional Cues into a Framework for Dialogue Management. In *Proc. of the ICMI-2002*, pages 141–148, Pittsburgh, PA, USA.

Ihme, K., Dömeland, C., Freese, M., and Jipp, M. (in press). Frustration in the face of the driver: A driving simulator study on facial muscle activity during frustrated driving. *Interaction Studies*.

Jones, C. M. and Jonsson, I.-M. (2007). Performance Analysis of Acoustic Emotion Recognition for In-Car Conversational Interfaces. In C. Stephanidis, editor, *Proc. of the UAHCI-2007*, volume 4555 of *LNCS*, pages 411–420, Beijing, China. Springer, Berlin, Heidelberg.

Katsis, C. D., Katertsidis, N., Ganiatsas, G., and Fotiadis, D. I. (2008). Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach. *IEEE Transactions on Systems, Man, and Cyberenetics - Part A: Systems and Humans*, 38(3):502–512.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., and Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing*, 9(4):269–275.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, 46(8):819–834.

Lee, B., Hasegawa-Johnson, M., Goudeseune, C., Kamdar, S., Borys, S., Liu, M., and Huang, T. (2004). AVICAR: Audio-Visual Speech Corpus in a Car Environment. In *Proc. of the INTERSPEECH-2004*, pages 2489–2492, Jeju, Jeju Island, South Korea.

Lewis, I. M., Watson, B. C., Tay, R. S., and White, K. M. (2007). The Role of Fear Appeals in Improving Driver Safety: A Review of the Effectiveness of Fear-arousing (threat) Appeals in Road Safety Advertising. *International Journal of Behavioral Consultation and Therapy*, 3(2):203–222.

Lotz, A. F., Faller, F., Siegert, I., and Wendemuth, A. (2018). Emotion Recognition from Disturbed Speech - Towards Affective Computing in Real-World In-Car Environments. In André Berton, et al., editors, *Elektronische Sprachsignalverarbeitung 2018*, Studientexte zur Sprachkommunikation, page s.p. TUDpress, Ulm, Germany.

Lu, J., Xie, X., and Zhang, R. (2013). Focusing on appraisals: how and why anger and fear influence driving risk perception. *Journal of Safety Research*, 45:65–73.

Malta, L., Miyajima, C., Kitaoka, N., and Takeda, K. (2011). Analysis of Real-World Driver's Frustration. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):109–118.

Martin, O., Kotsia, I., Macq, B., and Pitas, I. (2006). The eNTERFACE'05 Audio-Visual Emotion Database. In *Proc. of the ICDEW-2006*, Atlanta, GA, USA.

Ortega, A., Sukno, F., Lleida, E., Frangi, A., Miguel, A., Buera, L., and Zacur, E. (2004). AV@CAR: A Spanish Multichannel Multimodal Corpus for In-Vehicle Automatic Audio-Visual Speech Recognition. In *Proc. of the LREC-2004*, pages 763–767, Lisbon, Protugal.

Pêcher, C., Lemercier, C., and Cellier, J.-M. (2009). Emotions drive attention: Effects on driver's behaviour. *Safety Science*, 47(9):1254–1259.

Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1):203–212.

Rhodes, N. and Pivik, K. (2011). Age and gender differences in risky driving: the roles of positive affect and risk perception. *Accident Analysis and Prevention*, 43(3):923–931.

Russell, J. and Lemay, G. (2000). Emotion Concepts. In Michael Lewis et al., editors, *Handbook of Emotions*, pages 491–503. Guilford Press, 2 edition.

Scherer, K. R., Shuman, V., Fontaine, J. J. R., and Soriano, C. (2013). The GRID meets the Wheel: Assessing emotional feeling via self-report. In Johnny J. R. Fontaine, et al., editors, *Components of Emotional Meaning: A sourcebook*, pages 281–298. Oxford University Press.

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.

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.

Siegert, I., Böck, R., and Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human-computer interaction: comparison and methodological improvements. *Journal on Multimodal User Interfaces*, 8(1):17–28.

Taubman-Ben-Ari, O. (2012). The effects of positive emotion priming on self-reported reckless driving. *Accident Analysis and Prevention*, 45:718–725.

Tawari, A. and Trivedi, M. (2010). Speech Based Emotion Classification Framework for Driver Assistance System. In *Proc. of 2010 IEEE Intelligent Vehicle Symposium*, pages 174–178, San Diego, CA, USA.

Willstrand, T. D., Anund, A., Strand, N., Nikolaou, S., Touliou, K., Gemou, M., and Faller, F. (2017). Deliverable 1.2 - Driver/Rider models, Use Cases and implementation scenarios. http://www.adasandme.com/wp-content/uploads/2017/06/ADASME-20170619-D1.2-Driver-Rider-models-Use-Cases-and-implementation-scenarios.pdf.

Zhang, T., Chan, A. H. S., and Zhang, W. (2015). Dimensions of driving anger and their relationships with aberrant driving. *Accident Analysis & Prevention*, 81:124–133.