Driver Emotion Recognition for Intelligent Vehicles: A Survey

SEBASTIAN ZEPF, Mercedes-Benz AG
JAVIER HERNANDEZ, Massachusetts Institute of Technology
ALEXANDER SCHMITT, Mercedes-Benz AG
WOLFGANG MINKER, Ulm University
ROSALIND W. PICARD, Massachusetts Institute of Technology

Driving can occupy a large portion of daily life and often can elicit negative emotional states like anger or stress, which can significantly impact road safety and long-term human health. In recent decades, the arrival of new tools to help recognize human affect has inspired increasing interest in how to develop emotion-aware systems for cars. To help researchers make needed advances in this area, this article provides a comprehensive literature survey of work addressing the problem of human emotion recognition in an automotive context. We systematically review the literature back to 2002 and identify 63 peer-review published articles on this topic. We overview each study's methodology to measure and recognize emotions in the context of driving. Across the literature, we find a strong preference toward studying emotional states associated with high arousal and negative valence, monitoring the different states with cardiac, electrodermal activity, and speech signals, and using supervised machine learning to automatically infer the underlying human affective states. This article summarizes the existing work together with publicly available resources (e.g., datasets and tools) to help new researchers get started in this field. We also identify new research opportunities to help advance progress for improving driver emotion recognition.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → *Consumer health*;

Additional Key Words and Phrases: Affective computing, intelligent user sensing, emotion measurement, machine learning, literature survey, road safety

ACM Reference format:
Sebastian Zepf, Javier Hernandez, Alexander Schmitt, Wolfgang Minker, and Rosalind W. Picard. 2020. Driver Emotion Recognition for Intelligent Vehicles: A Survey. *ACM Comput. Surv.* 53, 3, Article 64 (June 2020), 30 pages.
https://doi.org/10.1145/3388790

1 INTRODUCTION

Emotions have been shown to be critical for most of our daily functioning, such as decision making, motivation, and interpersonal communication [31], and driving is no exception [49, 50, 127]. The

Authors' addresses: S. Zepf and A. Schmitt, Mercedes-Benz AG, Benz-Straße Tor 18, 71063 Sindelfingen, Germany; emails: {sebastian.zepf, alexander.as.schmitt}@daimler.com; J. Hernandez and R. W. Pickard, Massachusetts Institute of Technology, 75 Amherst Street, Cambridge, MA 02319; emails: {javierhr, pickard}@media.mit.edu; W. Minker, Ulm University, Albert-Einstein-Allee 43, 89081 Ulm, Germany; email: wolfgang.minker@uni-ulm.de.

0360-0300/2020/06-ART64 $15.00
https://doi.org/10.1145/3388790

daily commute can occupy a significant part of our day and is often associated with negative emotional states like anger [127] or anxiety [32]. Some of the main emotional triggers are the lack of control, travel delays, potential accidents, and the high cognitive load that is required. These triggers may be even more frequent for drivers who heavily rely on commuting as part of their professional activity (e.g., taxi drivers, package delivery). Although certain amounts of stress can help people achieve their goals, such as arriving at their destination safely, too much or too little may negatively impact driving performance and overall well-being [23]. Therefore, future vehicles that can sense and react to the emotional state of drivers and their passengers offer the opportunity of not only improving road safety but also potentially promoting greater mental health.

Recent technological innovations like wearable devices have enabled the study of emotions in real-life settings, leading to a growing number of papers examining the negative impact of certain emotions while driving (e.g., anger [127], sadness [49], or anxiety [32]). For a more detailed overview on the relevance of specific emotional states during driving and the necessity of emotion recognition in the automotive context, we refer readers to Eyben et al. [31]. In a seminal study by Jonsson et al. [55] and Nass et al. [81], for instance, it was shown that the emotional quality of the voice of the navigation system could interact with the driver's emotions to improve or worsen safety. In particular, using a cheerful navigation voice with happy or sad drivers led to their best and worst performance, respectively. To effectively select the "safest" tone of voice, a system that understands the affective state of the driver would need to be developed.

To enable automated affect recognition, researchers usually draw on the research area of affective computing [89]. This research draws from a variety of disciplines, including electrical engineering, psychology, psychophysiology, and computer science, and has been widely adopted in several fields, such as in education to increase student engagement levels [106], in market research to better understand customers [38], and in entertainment to provide personalized content [118]. However, the best methodology to collect, analyze, and use the emotional information is heavily dependent on the specifics of each setting or context (e.g., ambulatory vs laboratory, quality requirements of the measurements, affordances in the environment). To help stimulate such research in the context of driving, this article surveys the literature across the context of driver emotion recognition. In contrast to other published surveys (e.g., [30, 91]), this work is the first to systematically analyze the methods to sense and recognize emotions in the context of driving. Furthermore, we provide an overview of all relevant methodological steps, such as the representation, elicitation, and annotation of emotions, as well as potential emotion-enabled interactions with drivers.

The article is organized as follows. First, we describe the scope of the survey and the selection criteria for the papers. Second, we survey the different techniques used to study emotions, which include their representation, their elicitation, and their annotation. Third, we review the sensing, which includes the different signal modalities, how to measure them, and the main pre-processing steps. Fourth, we survey the methods used to perform the analysis and automated recognition of emotional states from the different modalities. Fifth, we highlight some relevant studies in the context of affective interaction that help motivate many of the potential applications. Sixth, we summarize publicly available resources such as datasets and tools that were used by the surveyed papers. Finally, we overview some of the main challenges and promising areas to better equip researchers who want to work in this growing area.

2 SCOPE AND METHODOLOGY

To find relevant papers to include while reducing familiarity bias, we first defined a selection criteria for identifying studies addressing the problem of emotion measurement and/or the analysis of emotions in the context of driving. We searched the IEEE, ACM, Springer, and Google Scholar research databases using the following combinations of keywords: affective computing automotive,

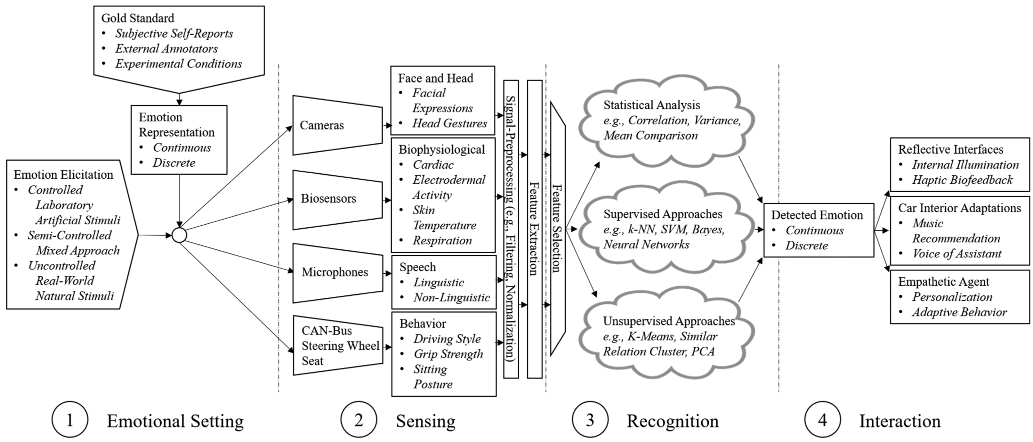

Fig. 1. Overview of the research methodology for investigation of emotion measurement in cars.

affective computing car, emotion recognition automotive, emotion recognition car, driver affective state, mood detection driver, emotion car passenger, emotion sensing passengers, emotion recognition co-driver, car occupants emotion recognition, emotions traffic occurrence, and emotions car frequency. To make the search more inclusive, we expanded the search to relevant references of the qualifying papers. In addition, we excluded studies mainly considering mental and physical states of drivers such as drowsiness/fatigue [10, 100, 109], inattention/distraction [25, 109], and mental workload [10], as those have been extensively covered by prior work. In contrast, we decided to include stress as part of this review due to its high impact on emotional well-being and the increasing interest in measuring it while driving. Using the stated criteria, we found a total of 63 papers that directly address the problem of automotive emotion recognition (Tables 1, 2, 3, and 4). Figure 1 highlights the main stages of automated driver emotion analysis, which are explained in more detail in the following.

One decision we made when deciding what to include or exclude relates to performance or accuracy. We deliberately decided to not list performance numbers in our tables summarizing the methods' attributes. When two rates are listed side by side in a table, the common human tendency is to assume that the higher rate means a better method, which within a single paper making careful comparisons is usually true. However, across the different papers we survey, they report rates for things that sound similar but are often fundamentally very different when you examine them carefully—in how they defined affective states, contexts, and the difficulty of the data (e.g., real open road with different lighting changing vs lab with constant lighting). As the experimental conditions of the surveyed papers are high dimensional and differ significantly, we deliberately excluded listing their numeric performance metrics to reduce the likelihood of readers making wrong conclusions. Instead, this article provides a comprehensive review of the methodologies, approaches, and contexts used, which we hope will help readers locate the papers of greatest interest to read for understanding specific performance rates. We also discuss the need for more shared datasets to facilitate meaningful quantitative comparisons.

3 EMOTIONS

This section reviews relevant information about the representation, elicitation, and measurement of emotions in the selected papers.

Table 1. Studies Focused on Automotive Emotion Recognition Part 1

Reference	Emotions	Setting	Emotion Origin	Labeling	Signals	Methods	Data Composition
Agrawal et al. 2013 [2]	Happiness, Anger, Sadness, Surprise	—	—	—	Face	Fuzzy Rules Based System	500 image frames
Akbas 2011 [3]	Stress	Real	Natural	Exp., Ann., Self.	EDA, ECG, EMG	Statistical Analysis	10 subjects, 50 min to 1.5 h each
Alvarez et al. 2012 [4]	Anger, Annoyance, Confusion, Boredom, Neural, Happiness, Joy	—	—	—	Speech	Logistic Model Trees, Multi-Layer Perceptron, Logistic Regression, Naive Bayes	<500 utterances
Begum et al. 2012 [8]	Stress	Real	Natural	Ann., Self.	ECG	k-NN	18 operators, 2.5 h each
Boril et al. 2010 [12]	Negative, Non-Negative	Real	Induced	Exp., Ann.	Speech	GMMs + SVMs	68 subjects
Boril et al. 2011 [13]	Negative, Non-Negative	Real	Induced	Exp., Ann.	Speech	GMMs + SVMs	68 subjects
Boril et al. 2012 [11]	Stress	Real	Induced	Exp., Ann.	Speech, CAN-Signals	Speech: GMMs; CAN: Multiple Interval Thresholds	15 subjects
Conjeti et al. 2012 [20]	Stress	Real	Natural	Experiment	PPG, EDA	RNN	20 subjects
Cruz and Rinaldi 2017 [21]	Stress	Real	Natural	Annotators	Face	CNN	308,202 frames
Deng et al. 2013 [22]	Stress	Real	Natural	Exp., Ann., Self.	EMG, EDA, ECG, RESP	Combinatorial Fusion	10 subjects, 50 min to 1.5 h each
Fernandez and Pickard 2003 [33]	Stress	Simulator	Induced	Experiment	Speech	Variations of HMMs, SVMs, NNs	598 utterances
Gao et al. 2014 [34]	Anger, Disgust	Non-Car/ Real(static)	Acted	Acted	Face	SVMs	21 subjects, 42 videos @ 30 s; 12 subjects, 10 videos
Grimm et al. 2007 [37]	Valence, Activation, Dominance	Non-Car	Natural	Annotators	Speech	Support Vector Regression	47 speakers, 947 utterances
Haouji et al. 2018 [28]	Stress	Real	Natural	Exp., Ann., Self.	EMG, ECG, EDA, RESP	Random Forest	10 subjects, 50 min to 1.5 h each
Healey and Pickard 2005 [43]	Stress	Real	Natural	Exp., Ann., Self.	ECG, EDA, RESP	Fisher Projection Matrix + Linear Discriminant Analysis	16 subjects, 50 min to 1.5 h each
Hoch et al. 2005 [47]	Positive, Negative, Neutral	Non-Car	Acted	Acted	Speech, Face	Speech: NN, SVMs; Facial: SVMs; Fusion: Linear Function Coefficient	840 audiovisual seq. from 7 speakers
Ihme et al. 2018 [48]	Frustration	Simulator	Induced	Exp., Self.	Face	Correlation Analysis	28 subjects, 40 min each
Jeong et al. 2007 [51]	Stress	Real	Induced	Self-Reports	ECG	Qualitative	6 subjects, 410 km each
Jones and Jonsson 2005 [53]	Boredom, Sadness/Grief, Frustration/Anger, Happiness, Surprise	Simulator	Natural	Annotators	Speech	Comparison Automatic Speech Recognition and Human Transcript	60 subjects, 20 min each
Jones and Jonsson 2007 [54]	Boredom, Sadness/Grief, Frustration/Anger, Happiness, Surprise	Simulator	Natural	Annotators	Speech	Statistical Analysis and NN	41 subjects, 20 min each
Jones and Jonsson 2008 [52]	Boredom, Sadness/Grief, Frustration/Anger, Happiness, Surprise	Simulator	Natural	Annotators	Speech	Statistical Analysis and NN	18 subjects, 45 min each

Table 2. Studies Focused on Automotive Emotion Recognition Part 2

Reference	Emotions	Setting	Emotion Origin	Annotation	Signals	Methods	Data Composition
Karaduman et al. 2013 [56]	Aggression, Calmness	Real	Natural	–	CAN	Similar Relation Cluster	5 tours @ 5 km
Karimi and Sedaaghi 2013 [57]	Anger, Neutral, Happiness, Sadness, Surprise, Fear, Boredom, Disgust	Non-Car	Acted	Act., Ann.; Act.	Speech	Multi: Bayes, k-NN, GMMs; Binary: SVMs, Bayes, k-NN, NN	4 actors, 30 min; 10 actors, 800 utterances; 10 students, 1,200 utterances
Kato et al. 2011 [58]	Positive, Negative	Simulator	Induced	Self-Reports	ECG, PPG	Discriminant Function Analysis	3 subjects, 120 min each
Katsis et al. 2008 [59]	Stress, Disappointment, Euphoria	Simulator	Natural	Annotators	EMG, ECG, EDA, RESP	SVMs, Adaptive Neuro-Fuzzy Inference System	1 subject
Keshan et al. 2015 [61]	Stress	Real	Natural	Exp., Ann., Self.	ECG	Naive Bayes, Logistic Regression, MLP, SVMs, k-NN, Decision Tree, Random Forest, Random Tree	10 subjects, 50 min to 1.5 h each
Kolli et al. 2011 [62]	Anger, Disgust, Fear, Happiness, Sadness, Surprise	Non-Car	–	Acted	Face	Modified Hausdorff Distance	35 subjects
Leng et al. 2007 [64]	Fear, Amusement	Non-Car	Induced	–	PPG, EDA, ST	ANOVA	5 subjects
Lisetti and Nasoz 2005 [67]	Frustration/Anger, Panic/Fear, Boredom/Sleepiness	Simulator	Induced	Exp., Self.	EDA, ECG, ST	k-NN, Discriminant Function Analysis, Marquardt Backpropagation, Resilient Backpropagation	41 subjects, 12 to 16 min each
Ma et al. 2017 [69]	Happiness, Bother, Concentrated, Confusion	Real	Natural	Annotators	Face	SVMs	10 subjects, 30 videos, 8 to 20 min each
Malta et al. 2008 [72]	Irritation	Real	Induced	Annotators	EDA, CAN	Bayesian Network	30 subjects, 1 h each
Malta et al. 2011 [73]	Frustration	Real	Induced	Annotators	Pedals, EDA, Face, Events	Bayesian Network	30 subjects
Minhad et al. 2016 [82]	Happiness, Sadness, Anger, Disgust, Fear	Non-Car	Induced	Self-Reports	EDA	SVMs	23 subjects
Moriyama 2012 [76]	Aggression (Tension, Irritation)	Non-Car	Acted	Acted	Face	Mutual Subspace Method + Principal Component Analysis	10 subjects, 5 min each
Munla et al. 2015 [77]	Stress	Real	Natural	Exp., Ann., Self.	ECG	SVMs, k-NN, RBF	16 subjects, 50 min to 1.5 h each
Nasoz et al. 2002 [80]	Neutral, Anger, Fear, Sadness, Frustration	Non-Car	Induced	Exp., Self.	EDA, ECG, ST	k-NN, Discriminant Function Analysis	10 subjects, 45 min each
Nasoz et al. 2010 [79]	Panic/Fear, Frustration/Anger, Boredom/Fatigue	Simulator	Induced	Exp., Self.	EDA, ECG, RESP, EMG, Finger Pressure	k-NN, Marquardt Backpropagation, Resilient Backpropagation	41 subjects, 12 to 16 min each
Oehl et al. 2011 [83]	Happiness, Anger, Neutral	Simulator	Induced	Exp., Self.	Grip Strength	Mean and Standard Deviation Comparison	59 subjects
Ooi and Ahmad 2016 [84]	Neutral, Anger, Stress	Simulator	Induced	Exp., Self.	EDA	SVMs	20 subjects @ 15 min
Paredes et al. 2018 [86]	Stress	Simulator	Induced	Exp., Self.	Steering Angle	Mean and Standard Deviation Comparison	25 subjects @ 112 turns
Parsons and Courtney 2016 [87]	Stress	Simulator	Induced	Experiment	ECG, EDA, RESP	ANOVA	50 subjects

Table 3. Studies Focused on Automotive Emotion Recognition Part 3

Reference	Emotions	Setting	Emotion Origin	Annotation	Signals	Methods	Data Composition
Paschero et al. 2012 [88]	Drowsiness, Alert; Happiness, Anger, Sadness, Fear, Disgust, Surprise	Non-Car	Acted	Acted	Face	Multi-Layer Perceptron	5 different datasets
Rebolledo-Mendez et al. 2014 [93]	Concentration, Tension, Tiredness, Relaxation	Real	Natural	Self-Reports	EEG, EDA	Principal Component Analysis, Logistic Regression Models, *k*-Means	24 subjects, 8 min each
Riener et al. 2009 [94]	Arousal	Real	Natural	–	ECG, GPS	Qualitative	1 subjects, 22 trips, >500 km
Rigas et al. 2012 [95]	Stress	Real	Natural	Self-Reports	EDA, ECG, RESP, CAN, GPS	Bayesian Network	13 subjects, 50 min each
Saeed and Trajanovski 2017 [99]	Stress	Real; Sim.	Nat.; Ind.	Exp., Ann., Self.; Exp.	EDA, ECG, PPG	Multi-Task NN	10 subjects, 50 min to 1.5 h each; 19 subjects @ 25 min
Schuller 2004 et al. [103]	Happiness, Anger, Sadness, Fear, Disgust, Surprise, Neutral	Non-car; –	Act.; Ind.	Act.; Self.	Speech	Non-Linguistic: *k*-Means, k-NN, GMM, MLP, SVM; Linguistic: Belief Network; Fusion: Means, MLP	13 subjects, 2,829 samples; 700 utterances
Schuller et al. 2006 [102]	Anger, Confusion, Neutrality	Simulator	Natural	Annotators	Speech	SVMs	10 subjects, 2,022 phrases
Schuller 2008 [105]	Anger, Boredom, Disgust, Fear, Happiness, Sadness, Neutral	Non-Car	Act., Ind.	Act., Ann.; Ann.	Speech	SVMs	10 actors, 494 samples; 44 subjects, 1,170 samples
Schuller et al. 2008 [104]	Fear, Stress, Screaming, Neutrality, Anger, Boredom, Disgust, Happiness, Sadness, Surprise, Aggression, Intoxication, Cheerful, Nervousness, Tiredness	Non-Car	Act.; Ind.; Nat.	Ann.; Act., Ann.; –	Speech, Face	SVMs	8 subjects, 396 samples; 10 actors, 800 sentences; 44 subjects, 1,170 samples; 7 subjects, 3,663 samples
Siebert et al. 2010 [108]	Happiness, Anger, Neutral	Simulator	Induced	Experiment	Grip Strength	Mean and Standard Deviation Comparison	22 subjects
Singh et al. 2010 [110]	Stress	Real	Natural	Experiment	PPG, EDA, RESP, ECG	ANOVA	14 subjects, 31 to 39 min each
Singh et al. 2011 [111]	Stress	Real	Natural	Experiment	EDA, PPG	Triggs Tracking Variable and Desirability Function Approach	9 subjects, 20 min each
Singh et al. 2012 [112]	Stress	Real	Natural	Experiment	EDA, PPG	KSOM Cluster Analysis	22 subjects, 34 to 37 min each
Singh et al. 2013 [113]	Stress	Real	Natural	Experiment	EDA, PPG	NNs	19 subjects, 24 min each
Taib et al. 2014 [119]	Frustration	Simulator	Induced	Exp., Self.	Seat Posture Distance and Pressure	Bayesian NN, SVMs, GMMs, MNR, GMM+SVM	19 subjects, 24 min each
Tawari and Trivedi 2010 [122]	Positive, Negative, Neutral, Anger, Boredom, Disgust, Anxiety, Happiness, Sadness	Non-Car; Real(static)	Act.; Act., Nat.	Acted; –	Speech	SVMs	10 actors, 494 samples; 4 subjects, 224 samples
Tawari and Trivedi 2010 [120]	Positive, Negative, Neutral, Anger, Boredom, Disgust, Anxiety, Happiness, Sadness	Non-Car; Real(static)	Act.; Act., Nat.	Acted; –	Speech	SVMs	10 actors, 535 samples; 4 subjects, 224 samples

Table 4. Studies Focused on Automotive Emotion Recognition Part 4

Reference	Emotions	Setting	Emotion Origin	Annotation	Signals	Methods	Data Composition
Tawari and Trivedi 2010 [121]	Positive, Negative, Neutral	Non-car; Real(static)	Act., Nat.	–	Speech	SVMs	4 subjects, 224 samples
Tews et al. 2011 [125]	Anger, Happiness, Neutral	Simulator	Induced	Exp., Self.	Face	Statistical Variance	10 subjects
Tischler et al. 2007 [126]	Speech: Arousal and Valence; Face: Anger, Fear, Disgust, Happiness, Surprise, Sadness	Real	Natural	Self-Reports	Speech, Face	Qualitative	8 subjects
Wang and Gong 2008 [129]	Happiness, Anger, Sadness, Fatigue, Neutral	Simulator	Induced	Exp., Ann., Self.	RESP, EDA, ST, PPG	Temporal Transition Model with Latent Variable	13 subjects, 5 sessions @ 15 to 20 min
Wang et al. 2013 [130]	Stress	Real	Natural	Exp., Ann., Self.	ECG	k-NN, PCA, LDA	–

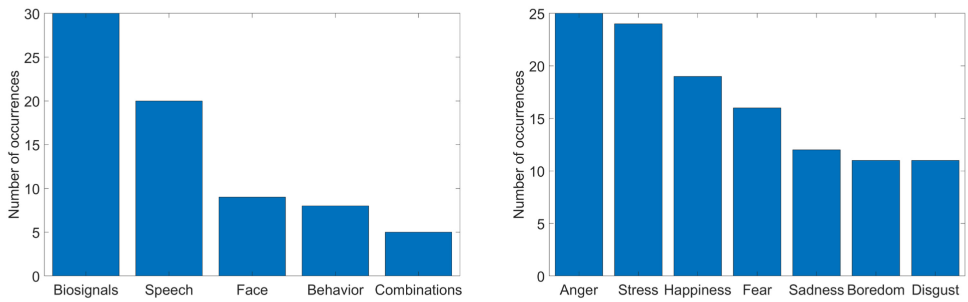

Fig. 2. Barplots showing the number of occurrences for the four different signal types (left) and the most frequently considered affective states (right) within the surveyed articles.

3.1 Representation

The first step toward studying emotions is to select a representation model that allows the definition and comparison of different emotional states. Although there are many methods to represent emotions, the selected papers mostly relied on discrete and continuous representations. However, discrete representations of emotions are based on the hypothesis that emotions can be discretized into several categories. For instance, Ekman et al. [26] argued that there are six basic emotions (e.g., anger, disgust, happiness, sadness, surprise, fear). On another note, continuous representations of emotions [98] emphasize that emotional states can be expressed as having several continuously changing components. Two of the most important components are emotional valence, which ranges from negative to positive, and emotional arousal, which ranges from low to high, and is sometimes termed *energy* or *activation*. Discrete emotions can then be mapped into the 2D model (or one can add higher dimensions as well). For instance, happiness would tend to be associated with positive valence and increased arousal, and sadness would tend to be associated with negative valence and lowered arousal [98].

Both kinds of representations have advantages and disadvantages. Although the continuous dimensional approach is beneficial for mathematical analysis, the interpretation of dimensions such as "arousal" tends to vary, which can lead to ambiguous emotional ratings [24]. The discrete approach can be easier for people to rate; however, considering only a pre-defined set of discrete emotions can result in significant biases and priming effects that can limit what gets reported while leading to noisy labels when the labels do not fit what is experienced [7, 24].

A few studies have investigated which emotional states occur most frequently while driving. For instance, Mesken et al. [75] found that anxiety occurred most frequently, followed by anger and happiness. In a separate study, Dittrich and Zepft [24] requested drivers to self-report their emotional states while driving and found that the most frequent terms were "good/okay/alright," "anger," "annoyance," "joy," and "relaxation/serenity." These labels, which arise from asking drivers what they feel, only overlap in part with the labels most frequently described by emotion theorists.

The majority of the surveyed papers focused on high arousal emotional states (83%), especially with negative valence (56%). Some of the most commonly considered emotional states for drivers were anger (25 papers), stress (24 papers), happiness (19 papers), and sadness (16 papers) (see Figure 2, right). Certain emotions (e.g., anger, frustration) are studied because they can negatively impact the driving performance, yet other states (e.g., stress, happiness) are studied as regulatory, as it has been shown that arousal and valence were associated with driving performance in an inverted U-shape, indicating that performance can reach its peak with certain levels of arousal and valence [18].

Fig. 3. Example of a driving simulator.

3.2 Elicitation

A critical factor to consider when comparing studies is the selection and design of the experimental setting. Among the selected papers, 56% of the studies were performed under laboratory conditions, in which external influencing factors can be easily controlled, and 44% were performed in a real-life setting, in which the results may be more representative but the data are usually more difficult to analyze.

In the context of laboratory conditions, researchers have explored a wide variety of emotion elicitation techniques, such as giving presentations, watching videos, and recalling past emotional events (noted as "Non-Car" in Tables 1 through 4). A popular method in the context of driving involves the use of car simulators like the one shown in Figure 3 to recreate a driving experience in a safe and repeatable manner. For instance, Ihme et al. [48] challenged the participants with difficult driving tasks (e.g., work for a delivery service under high time pressure with impeding road conditions). In a separate study, Ooi and Ahmad [84] used the simulator to induce stress and anger with challenging routes (e.g., snowy mountains) and by adding aggressive drivers, respectively. To help make the simulator experience more natural, some researchers have explored incorporating additional naturalistic stimuli. For instance, Jones and Jonsson [53] added sound effects to simulate real-life driving, Gutmann et al. [40] added motion to simulate the external forces, and Lin et al. [66] networked multiple simulators to share the driving experience with other people. In addition, some recent studies have explored the use of virtual reality technologies to help elicit a more intense physiological response than traditional displays [29]. Despite the many benefits, controlled experiments still suffer from undesired experimental factors (e.g., novelty factor, white-coat hypertension, and the knowledge that a mistake or a crash causes no real harm); hence, the findings may not be easily generalizable to real-life settings. For instance, Ruscio et al. [97] compared driving behavior and physiological signals between simulated and real-life driving. In particular, they observed similar driving behavior but significantly different average speed and significantly different physiological responses. In addition, simulated driving experiences have also been shown to induce motion sickness in the participants [16].

In the context of real-life conditions, five studies took drivers out on real roads and partially controlled the experiment by pre-defining a driving route and artificially inducing certain emotional states. For instance, Boril et al. [12, 13] asked participants to drive a specific route while making phone calls to elicit cognitive load and a negative emotional experience. Finally, 20 studies took drivers onto real roads with a pre-determined route and allowed participants to experience natural emotions. For instance, Healey and Picard [43] and Singh et al. [113] set pre-defined routes to cover different driving environments, such as quiet areas on a university campus, challenging city scenarios, and highway driving. To the best of our knowledge, no studies have considered completely uncontrolled settings to study emotions in driving scenarios.

3.3 Annotation

To effectively recognize the emotional state of drivers, it is important to obtain reliable annotations of emotions that can be used as a gold standard to train and evaluate the models. The surveyed papers considered three main approaches: self-reports, external annotators, and experimental context. The first approach involves leveraging emotional self-reports, which requires participants to be able to verbalize and/or quantify how they feel at a particular moment. For instance, Taib et al. [119] collected a self-report after several driving subtasks using a 9-point Likert scale for frustration. In a separate study, Ihme et al. [48] also investigated frustration by asking participants to complete the Self-Assessment Manikin [15]. In the work of Kato et al. [58], the researchers used the Positive and Negative Affect Scale [131], and introduced the Multiple Mood Scale [124] and the Profile of Mood Status [74]. Some of the challenges associated with self-reports, however, are that they require the cognitive attention of participants, they are subjective, and they may reflect strong biases (e.g., false memories, desire to impress the experimenter) [65, 107]. In the literature we surveyed, 10% of the papers used self-reports to annotate emotions.

The second approach involves the use of external annotators that can recognize certain emotional states based on different signals (e.g., behaviors, facial expressions) of the participants [52, 69]. For instance, Jones and Jonsson [52] used a human listener to transcribe and annotate emotional voice recordings. Similarly, Ma et al. [69] used this method to collect six independent annotations from external observers for each video segment and then analyzed their consistency (a.k.a., inter-rater agreement). However, this approach is very time and labor intensive and requires the use of experienced and trained observers, which may be difficult to find and/or would be expensive, especially at scale. In our survey, 23% of the papers used external annotators to capture the perceived emotional state of the driver.

The third and most popular approach involves using different experimental conditions to label the emotional experience (e.g., [33, 112]). In the case of simulator studies, for instance, the researchers have several opportunities to modify the experimental conditions and elicit certain emotional states. For instance, adding a manipulated secondary task leading to either successful or failed completion can be used to push the driver's emotional state into positive or negative states, respectively [33, 48, 87]. Alternatively, manipulating the driving conditions such as by increasing the amount of traffic or changing the behavior of other road users can be used to elicit negative states such as stress, frustration, or annoyance [48]. As these factors may be more difficult to modify during real-world driving tasks, researchers have explored differentiating road segments and different times of day of driving to make it more likely to elicit different states, such as driving through congested city intersections where lots of pedestrians disobey crosswalk rules for eliciting a high level of stress, driving at non-rush hour on a straight highway under good weather conditions for a low-to-average level of stress, and being stationary, resting in a garage with eyes closed, for a low level of stress [43, 112]. Although this approach minimizes the burden of participants, it makes some strong assumptions that may not generalize well to all participants and road conditions. In our survey, 31% of the papers used the experimental conditions as a reference.

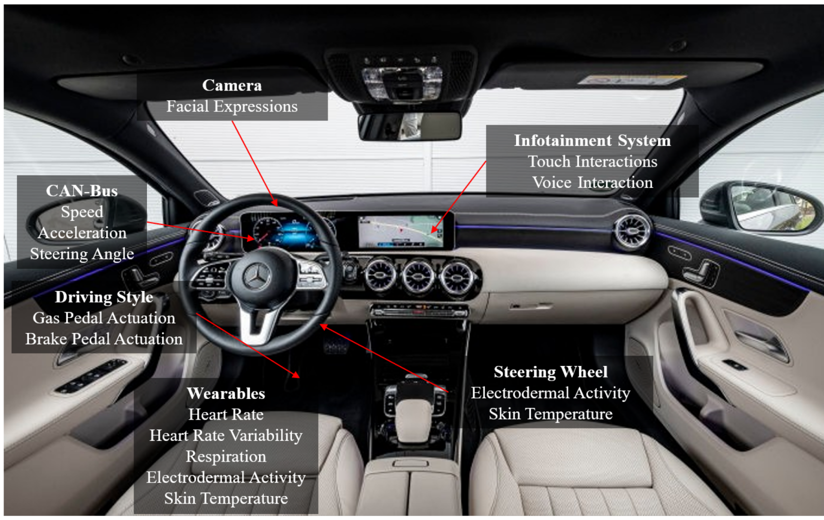

Fig. 4. Relevant signals and their potential measurement location for emotion recognition.

Some studies also explored multiple approaches to overcome the weakness of any one approach. For example, Healey and Picard [43] combined the three methods to develop ground truth for driver stress level: they (i) asked drivers to rate different parts of a drive (e.g., exiting parking garage, merging onto highway) from 1 = "no stress" to 5 = "high stress," as well as to rank order all of the driving events from "most" to "least" stressful; (ii) they asked coders to watch a video replaying the driver's experience (looking at the driver and the environment around them) and count the complexity of the events second by second (e.g., avoiding a pothole, turning the head, seeing a pedestrian walking toward car); and (iii) they monitored drivers in different road conditions (e.g., rest, highway without traffic, and congested city) associated with stress levels (e.g., low, medium, high). Although no one method is perfect, finding multiple methods that converge to the same labels can boost confidence in the annotations. In our survey, 36% of the papers explored a combination of multiple approaches to obtain emotion annotations.

4 SENSING AND PRE-PROCESSING

This section provides an overview of the signals used to capture different aspects of emotions in the context of driving (Figure 4), together with the acquisition methods, pre-processing steps, and features used to characterize relevant changes. The surveyed papers used four main groups of signals. On the left side of Figure 2, the frequencies of use of the four main signal groups among the considered papers are illustrated.

4.1 Face and Head

A total of nine studies considered facial and head gestures for inferring driver emotion, typically by examining facial expressions (e.g., smiling, frowning) and head gestures (e.g., nods, tilts) in the context of the drive.

To capture face and head signals, the studies mostly employed traditional RGB cameras (e.g., [69, 88]. Some less frequently explored approaches included the use of thermal cameras [62] and infrared cameras [34], which may be more robust to certain types of illumination changes. To accurately detect the dynamic range of facial expressions, a frontal view of the driver is usually preferred. In controlled laboratory studies, researchers usually place the camera on top of a display or the simulator to capture a frontal view (e.g., [2, 76]). In less controlled environments, researchers have placed the camera on the car windshield, although it may partially obstruct the driver's view [21]. Another considered location is the car dashboard, but the camera view may be partially occluded by the steering wheel during turns [34].

Once videos or images have been captured, several pre-processing steps may be needed. As a first step, areas of interest or regions of interest such as the face or the head need to be detected. For instance, Agrawal et al. [2] compared different approaches based on colors, growing regions, morphological operations, and their combination, although no significant differences were found. In a separate study, Gao et al. [34] used a supervised descent method for face detection. Finally, Paschero et al. [88] used the open source Viola-Jones face detector [128] from the OpenCV library. After face/head detection, relevant points or smaller regions on the face and/or the body are usually detected to appropriately capture certain motions (e.g., smiles). Note that although it may be tempting to think that smiling means the driver is happy, drivers may also smile when frustrated or when bright sun is in their eyes. Context is important when interpreting information sensed from the driver.

The number of points/areas and their locations varied significantly across studies depending on the specific focus. In our survey, four studies [48, 69, 76, 125] detected facial points and/or areas of facial muscle movements to capture the facial action units (AUs), identified within the Facial Action Coding System (FACS) [27]. Other detection approaches included eye and mouth tracking with selected areas [2] and vertical lines [88] that are influenced during facial expressions (e.g., from neutral to laughter). Finally, different aspects of the images may require normalization to make the analysis more robust to different changing factors (e.g., driver and car movements, illumination changes). For instance, Gao et al. [34] explored a pose normalization method using a 3D cylindrical head model to reduce the negative effects of pose mismatch.

As a final processing step, several features are typically extracted from the different facial/body points to facilitate analysis. Two of the most commonly used groups of features were shape-based (e.g., angles and distances between facial landmarks) and appearance-based features (e.g., color, texture). For instance, Cruz and Rinaldi [21] used local binary pattern features from three orthogonal planes to capture the texture, and Kolli et al. [62] used a histogram of oriented gradients to capture the appearance. When combining different types of features, additional normalization steps at the feature level may be needed to help correct the different ranges (e.g., angles and distances between facial points). For instance, Paschero et al. [88] performed a two-step normalization in which they first corrected the range of different variables to be between 0 and 1, then subtracted the mean and divided it by its standard deviation. It is important to note, however, that not all studies follow the same steps, as they are very dependent on the learning approach. For instance, deep learning approaches can automatically find relevant areas of interest and extract features while also providing recognition [21].

4.2 Biophysiological Signals

A total of 30 studies considered the measurement of biosignals that are related to the regulation of the body and influenced and/or affected by the experience of emotions. Although there are many potential signals, the four most popular considered subgroups are cardiac (CAR) considered in 26 studies, electrodermal activity (EDA) considered in 24, respiratory (RESP) considered in 9, and skin temperature (ST) considered in 4 studies.

To measure each of the biosignals, the studies leveraged a wide variety of methods. For instance, 19 studies measured CAR signals from electrocardiographic (ECG) signals (e.g., [94]), which usually capture the electrical activity of the heart with electrodes attached to the chest, and 9 studies measured CAR signals from photoplethysmographic (PPG) signals, which usually use LEDs or cameras to capture color changes at the surface of the skin due to the underlying blood movement. By detecting and/or counting the specific heartbeats in each of the signals, new temporal signals have been derived, such as heart rate (HR), which indicates the number of beats per minute, and heart rate variability (HRV), which indicates the variability of inter-beat intervals. For a thorough discussion on HRV analysis, we refer readers to the review by Marek et al. [71]. RESP signals capture respiratory activity and are usually measured with a chest-worn strap (e.g., [43, 59, 87]). In addition, some studies extract RESP signals by analyzing different frequencies of HRV (e.g., [92, 111]). Similar to CAR signals, new temporal information such as breathing rate (BR) can be derived by counting the number of oscillations. EDA (often *termed galvanic skin response* in older literature) is the phenomenon whereby the skin changes electrically with changes in the sympathetic nervous system; it is usually measured by placing two electrodes on the surface of the skin and measuring skin conductance. The electrodes are classically used with gel and placed where the eccrine sweat glands can be found in high density (e.g., palms of the hand [43] or soles of the feet [70]). However, recent studies have also obtained data using dry electrodes in more practical locations less likely to suffer from motion artifacts (e.g., wrist, ankle) [40]. For a thorough discussion on EDA, we refer readers to Boucsein [14]. Finally, ST reflects the temperature of the skin. Both EDA and ST have been monitored with different types of wearables on different body positions (e.g., BodyMedia SenseWear Armband on the upper arm [67, 80], Empatica E4 on the wrist [40]).

When using these signals for emotion analysis, several pre-processing steps are usually performed. Unwanted motions from the car and/or the person can corrupt the quality of the measurements and introduce sensor artifacts (e.g., sudden drops of the EDA signal when the electrode is pulled away from the skin, or corruption of the PPG signal with underlying muscle movement). Thus, different approaches have been proposed to detect and exclude these segments from the analysis. For instance, Singh et al. [113] applied a 1D median filter to remove signal spikes, and Munla et al. [77] used a bandpass filter to remove noise. However, biophysiological signals need to be appropriately normalized to account for different baselines and physiological ranges. These differences can be caused due to different factors, such as demographics (e.g., age, gender) and placement of the sensors. Therefore, several papers address this challenge by correcting the range of values to be between 0 and 1 (e.g., [113]) or by z-scoring them so they have zero mean and unit variance (e.g., [67]).

Finally, researchers have explored a wide variety of features to characterize biophysiological signals, which can be grouped into time domain and frequency domain features depending on the domain from which they were computed. Among the 30 studies, 16 studies focused on only time domain features, 1 study focused on only frequency domain features [51], and 13 studies considered a combination of both. Although the studies considering EDA or ST mostly relied on time domain features due to their limited high-frequency components (e.g., [64, 67, 84]), studies considering CAR and RESP signals usually combined both types of features. The most frequently used time domain features across all of the signals were the mean, the standard deviation, and the root mean square error over a specific time window (see more details in the following section). The most frequently used frequency domain features were the low-frequency/high-frequency ratio from HRV, and the amount of energy of HRV and RESP (e.g., [22, 43]) at different frequency bands.

4.3 Speech

A total of 20 studies considered speech to perform driver emotion recognition. In these studies, the challenge is to process, in a noisy and changing automotive environment, the signals produced by the vocal cords that may be related to the emotional state of the driver, such as the pitch and volume of the voice.

The reviewed studies considered a wide variety of devices to record sound, including directional microphones [122], condenser microphones [102], and microphone arrays [52]. To help capture relevant information while minimizing car and environmental noise, the placement of the microphones is critical. For instance, Jones and Jonsson [52] explored three different setups and found that a four-microphone directional beam located 1.5 m in front of the driver delivered the best recordings. Other studies considered placing the microphone in the middle of the instrument panel [37] or above the windshield [12].

Once streams of audio have been collected, different filters are usually applied to help amplify relevant acoustic signals and remove other overlapping noise (e.g., engine of the car). For instance, Grimm et al. [37] used finite impulse response filters to help amplify the emotional content of sounds. However, Schuller [105] showed that incorporating certain amounts of noise benefited the task of emotion recognition in naturalistic settings. Finally, as individuals have different speech signatures that need to be normalized, researchers have explored different normalization methods. For instance, Schuller [105] and Tawari and Trivedi [120] showed that using a speaker adaption step with a z-scored normalization helped further improve generalization performance.

Most studies considering speech signals relied on non-linguistic (a.k.a., paralinguistic) features, which focus on the way in which things are being said. The only exception was the study by Schuller et al. [103], who investigated a combination of linguistic and non-linguistic features. The most common speech characteristics were the pitch, the loudness, the length of sounds, and the spectral features such as mel-frequency cepstral coefficients (MFCCs). The majority of studies (15) combined features from both the time and frequency domains.

4.4 Behavior

A total of eight studies considered behavioral characteristics that focus on signals relating to driver behavior that may be influenced by the emotional state of the driver, especially interactions that are directed toward the car, such as changes in steering wheel and pedal activations.

One of the most commonly used sources of driver and car information can be found at the controller area network (CAN). However, CAN bus signals are mostly accessible for internal developers and are usually kept confidential. A smaller subset of these signals may be more readily available through the on-board diagnostics (e.g., OBD II), which offers a standardized data interface. Four studies [11, 56, 72, 95] used this approach to capture signals such as acceleration, braking, and steering. In addition, the Advanced Driver Assistance Systems capture information about the road conditions and the driving style of the vehicle, such as the distance to the car in front [72]. To capture additional sources of behavioral information, researchers have also instrumented cars with a wide variety of sensing mechanisms. For instance, Oehl et al. [83] and Seibert et al. [108] measured grip strength by integrating an optical fiber into the steering wheel, Lin et al. [66] captured the same information by integrating piezoresistive resistances, Malta et al. [73] instrumented the gas and brake pedals with force sensors to capture leg motion, and Taib et al. [119] added pressure sensors on the seat to track changes in body posture.

Similar to the other categories, different pre-processing steps are usually applied to minimize individual differences. For instance, Rigas et al. [95] calculated the driver-specific mean values of the features and detected significant deviations at different temporal resolutions. In a separate study, Taib et al. [119] applied a min-max normalization to the signals obtained from the seat distance and pressure sensors. In addition, they recommended capturing behavioral baselines for each individual when exploring real-life applications.

5 ANALYSIS AND RECOGNITION

This section describes how the different signals and features have been used to study the different emotional states, some of the strongest associations between signals and emotions, and the most popular methods when analyzing and recognizing emotion.

5.1 Face and Head

Studies relying on face and head signals considered a total of 20 different emotional states. Several considered mainly Ekman's six basic emotions or a smaller subset (e.g., [62, 88]), with the states of happiness and anger being the most frequently studied.

One of the most commonly used approaches to study the relationship between face and head changes and emotions involves the use of statistical approaches such as correlation analysis. For instance, Ihme et al. [48] compared self-assessments of frustrated drivers with the annotations of certified FACS coders and concluded that frustration showed positive correlations with AU10 (upper lip raiser), AU12 (lip corner puller), AU17 (chin raiser), AU20 (lip stretcher), AU23 (lip tightener), and AU24 (lip pressor). In a separate study, Moriyama [76] used a mutual subspace method to capture the changes of pre-defined facial regions (e.g., forehead, right cheek, and left cheek). In particular, they concluded that driver tension was associated with increased activity in AU12, AU24, AU14 (dimpler), and AU28 (lip suck), and that driver irritation was associated with AU4 (brow lowerer) and AU9 (nose wrinkler).

Another commonly used approach involved the use of supervised machine learning or pattern recognition techniques, which require an annotated dataset (e.g., facial expressions with emotional annotations) to train the emotion recognition models. Although there are a wide variety of methods to learn the modes, the papers using face and head gestures considered three main methods. In particular, three papers used k-nearest neighbor (k-NN) [62, 76, 125], two papers used support vector machines (SVMs) with different kernels [34, 69], and two papers used variations of neural networks (NNs) [21, 88]. A critical component when developing such models was the temporal resolution of the predictions, which can usually be provided at a frame level (a.k.a., static approach) or at a window level (a.k.a., temporal approach). For instance, Ma et al. [69] and Kolli et al. [62] both considered predictions at a frame level. In addition, the study by Ma et al. [69] showed that considering the previous frame improved the recognition performance. To provide recognition at a window level, several studies have considered different voting schemes in which several frame-level predictions are aggregated to provide a single estimate. For instance, Gao et al. [34] and Paschero et al. [88] used a majority voting approach, and Moriyama [76] used an average voting approach. In a separate study, Cruz and Rinaldi [21] dynamically changed the number of frames, which was adjusted based on the rate of change of visual information. Overall, the considered studies show a tendency toward temporal approaches to better recognize emotions. All of the reviewed studies considered only a single classification method, limiting the potential performance comparison across methods.

5.2 Biophysiological Signals

Studies relying on biophysiological signals considered a total of 21 affective states. Further, 80% of these studies included the recognition of different stress levels in their analysis e.g., [20, 28, 43, 61, 99].

Due to the potentially large number of features when considering biophysiological signals, several studies considered feature selection techniques. For example, several used information gain to select features (e.g., [22, 28, 58]). According to the conclusions of such studies, EDA and CAR signals provided the most emotional information, followed by RESP and ST. In Begum et al. [8], for instance, a mixture of 20% features from the time domain and 80% features from the frequency domain based on HRV yielded the best results for the classification of stress. In a separate study, Deng et al. [22] used a feature selection method and showed that a combination of EDA features were the most representative to capture stress. Similarly, Healey and Picard [43] found higher correlation between stress and EDA features than between stress and HRV measures. In addition, some studies have explored the use of dimensionality reduction approaches, which automatically find the most relevant components. In particular, two commonly used methods were principal component analysis (PCA) (e.g., [93, 112, 130]), which finds a set of uncorrelated features that explain the variance in the original data, and linear discriminant analysis (LDA) [43, 130], which similarly fits the data with a linear combination of features while finding a linear function that discriminates classes (e.g. high stress vs low stress).

To perform emotion recognition, 22 papers followed a supervised learning approach. In particular, the most frequently applied methods were k-NN (seven times), followed by SVMs (five times) and naive Bayes (four times). In addition, some studies compared the performance of several methods within the same dataset. For instance, Singh et al. [113] evaluated the performance of seven different configurations of NNs for the recognition of stress, achieving the highest performance with recurrent neural networks (RNNs). In a separate study, Keshan et al. [61] compared 10 different supervised learning algorithms and showed that a Bayesian approach outperformed other methods when discriminating between two different stress levels, and that a decision tree approach (random tree) outperformed other methods when discriminating between three different stress levels. When considering the temporal resolution of the predictions, some studies focused on making predictions for a whole driving segment (e.g., [3, 64, 110]), whereas others focused on a specific time window when participants self-reported their emotional state (e.g., [79, 80]). The duration of the windows varied significantly across studies. For instance, Kato et al. [58] considered a duration of 180 seconds, Healey and Picard [43] considered durations ranging from 1 second to 5 minutes, Wang and Gong [129] considered time windows with a duration of 60 seconds, and Minhad et al. [82] considered a duration of 5 seconds. Overall, the most frequent durations were 5 minutes (five studies) and 10 seconds (five studies).

5.3 Speech

Studies relying on the speech signal considered a total of 28 emotional states. The most frequently studied states when using speech were anger, boredom, happiness, and sadness.

As with biophysiological signals, several studies considered feature selection algorithms to reduce the dimensionality (e.g., [102, 105, 121]). For instance, Karimi and Sedaaghi [57] compared four different feature selection algorithms and showed that sequential floating forward selection yielded the best results when considering seven emotional states in the presence of babble noise. Using similar approaches, other studies systematically studied what acoustic features were the most relevant to perform emotion recognition (e.g., [47, 103, 126]). Overall, the results of these studies suggest that features related to pitch, energy, and intensity, as well as spectral features such as MFCCs, provided the highest information gain. Furthermore, Schuller [105] showed that spectral features (especially the ones related to energy) outperformed other time domain features under noisy conditions.

To perform emotion recognition, 16 studies leveraged supervised learning algorithms. The most commonly used methods were SVMs (used in 10 studies) and NNs (5 studies). Karimi and Sedaaghi [57] compared several different supervised algorithms and showed that SVMs yielded the best results when discriminating between two emotional states, such as anger vs no anger, and that a Bayes classifier yielded the best results for a multi-class discrimination. In a separate study, Alvarez et al. [4] compared six different algorithms and showed that logistic model trees provided the best performance when classifying seven emotional states. In terms of temporal granularity of the predictions, the studies considered windows ranging from 20 msec [103, 105, 120] to 2 seconds [54].

5.4 Behavior

Studies relying on behavioral signals considered a total of eight emotional states. Overall, stress and anger were the most frequently considered emotions (four times each). To analyze the relationship between different behavioral changes and emotions, several studies performed mean comparisons across different data segments (e.g., [83, 86, 108]). For instance, Siebert et al. [108] and Oehl et al. [83] showed that the average grip strength significantly varied for both anger and happiness. In addition, Paredes et al. [86] demonstrated that it is possible to measure stress by using the steering angle and a mass spring damper model. In addition, some studies performed correlation analysis between different emotion annotations and behavioral driving features. For instance, Karaduman et al. [56] evaluated and selected different features from the CAN bus to discriminate between calm and aggressive driving. However, in this and in many multiple-behavior models, the authors did not provide readers with intuition into how the features and behaviors were associated with the different emotions.

To perform emotion recognition, five studies leveraged supervised learning methods. In particular, Boril et al. [11] used a Bayes approach that added the probabilities for the considered emotional states based on the distribution of the testing data. In a separate work, Taib et al. [119] compared five different supervised learning algorithms and their combinations (Bayesian neural network (BNN), SVMs, Gaussian mixture models (GMMs), multinomial regression (MNR), and GMM+SVM), and concluded that SVMs and MNR were the most promising ones for the task of frustration recognition from a driver's sitting posture. To perform the final emotion prediction, different temporal windows were considered. For instance, Taib et al. [119] used 3-second windows, and Boril et al. [11] used windows of variable sizes that were determined by the duration of the driving maneuvers.

5.5 Combinations

To better capture the different components of emotions, five studies considered different groups of signals simultaneously. In particular, Malta et al. [72] combined EDA and CAN behavior signals to study irritation; Rigas et al. [95] combined several biophysiological signals (EDA, CAR, RESP), CAN bus, and the Global Positioning System (GPS) signal to study stress; Hoch et al. [47] and Schuller et al. [104] combined speech and face to study different sets of emotions; and Malta et al. [73] combined all of the signal groups to study frustration.

To aggregate the different types of modalities, different fusion approaches were explored, which varied in what phase of the analysis the modalities were combined. In particular, four papers used a fusion at the feature level [72, 73, 95, 104], in which features from different information sources were provided to the same classifier, and one paper used a fusion approach at the decision level [47], in which the output of separate classifiers (e.g., from face and speech) were aggregated to obtain a final decision. In terms of methods, Malta et al. [72, 73] and Rigas et al. [95] used BNNs to add the likelihood for a specific emotional state from several information nodes, Schuller et al. [104] used SVMs to combine features from different signals, and Hoch et al. [47] used a linear fusion coefficient to regulate the weighting of the different sources of information. Finally, only one study [95] evaluated different window sizes ranging from 2 to 30 seconds and found that 10 seconds outperformed the others.

6 INTERACTION

To help provide a more complete understanding about the possibilities of affect recognition in driving settings, this section briefly reviews some relevant studies that leveraged emotional information at different levels. In contrast to previous sections, the selection of these studies is not meant to be comprehensive but is intended to be illustrative of the space.

One of the simplest and most commonly explored forms of interaction involves providing the emotional information back to the driver so that he or she can use it in different ways. For instance, MacLean et al. [70] developed MoodWings, a wrist-worn wearable butterfly that varied the frequency of its wing flapping according to the physiological arousal of the driver. Researchers showed that the use of MoodWings increased stress awareness and potentially driving safety. However, participants reported that the device itself also acted as a stressor. In a separate study, Hernandez et al. [45] described different car interactions in the context of driver stress management. Two relevant interactions involved a reflective dashboard that changed the background color based on physiological arousal captured from EDA, and a communicative paint that similarly changed the external color of the car to share the state of the driver with other road users. More recently, Löcken et al. [68] interviewed several human factor experts and proposed different approaches to use ambient light patterns to help mitigate frustration. For instance, they suggested to use calming ambient light patterns when driving in packed cities and informative ambient light patterns when searching for parking spots.

A more complex form of interaction involves using emotional information to change some aspects of the car. For instance, the studies by Jonsson et al. [55] and Nass et al. [81], which were briefly covered in the introduction, fall into this category. In particular, Nass et al. [81] showed that modifying the navigation voice intelligently based on the emotion of the driver can enhance the driver's performance and safety. In their study, one of two emotional states (mildly happy and mildly upset) was elicited in the participants before spending 20 minutes in a driving simulator. During the driving, the participants were confronted with several questions and were invited to interact with a navigation system. The voice of the system had either an energetic or a subdued tone. Their results indicated that aligning the happy state (of the driver) with the energetic voice (of the system), and aligning the mildly upset state with the subdued voice, improved driving safety. In this study, driving safety was associated with fewer accidents and better attention on the road, as well as improved the driver's cognitive ability to answer questions posed by the system. Their results are further validated with the findings of Harris and Nass [42], who adapted the behavior of speech dialogue systems in the event of frustration events. In particular, they showed that voice prompts that emphasize or deflate the reason for negative reactions can impair or improve driving performance, respectively. Furthermore, the performance in the deflating condition was comparable to a mode without voice interaction. Researchers have also explored the automated selection of music due to its potential impact on emotion regulation. For instance, Krishnan et al. [63] developed a music-mood mapping for a real-time music recommendation in car settings. More recently, Paredes et al. [85] explored the feasibility of certain movements and breathing exercises in the car to help provide more relaxed driving. In addition, their study provided insights into the appropriate intervention and methods to effectively guide the relaxation with vibrotactile stimulation.

Finally, emotional information has been used to develop driver companions that assist and interact with the driver in more complex and empathetic ways. For instance, Williams et al. [133] developed AIDA (Affective Intelligent Driving Agent), a social robot that assists the driver to decrease cognitive load and promote road safety. In this case, AIDA used the emotional information to understand the driver and modulate the interaction with the robot so its communication became more natural and efficient. In a separate study, Gusikhin et al. [39] developed EDAS (Emotive Driver Advisor System), which similarly proposed an affective in-car communication system.

Table 5. Publicly Available Datasets for Driver Emotion Recognition Under Real-World Conditions

Dataset	Emotions	Annotation	Signals	Data Composition
CIAIR Corpus [60]	None	None	Audio, Video (Driver+Road), GPS	500+ subjects, 60 min each
DriveDB [43]	Stress (High, Medium, Low)	Experimental Conditions, Subjective Self-Ratings, External Annotators	ECG, EMG, EDA, RESP	17 subjects, 54 to 93 min each
Ma et al. 2017 [69]	Happy, Bothered, Concentrated, Confused	External Annotators	Facial AUs by OpenFace	10 subjects, 23.6 km each
UTDrive DB Classical [6]	Stress (High, Low)	Experimental Conditions (Highway, Urban)	Audio, Video (Driver), Pedal Pressure, Front-Car Distance, CAN, GPS	77 subjects, 4 countries
UTDrive DB Portable [6]	None	None	Audio, Video (Driver), Acceleration, GPS	7 subjects

The focus of EDAS was personalization and adaptive behavior, which enabled automatic and intelligent user interaction for several in-car entertainment functions, such as providing music recommendations.

7 PUBLIC RESOURCES

As mentioned previously, we deliberately left out performance rates of the automated systems since the results cannot be fairly compared across the hugely varying datasets and driving conditions. To help stimulate research in automotive affect recognition, more common datasets are needed that can be shared, with common tasks defined to facilitate better ability to make comparisons. This section overviews some of the publicly available resources, including those utilized and/or provided by the reviewed papers.

7.1 Databases

Table 5 summarizes the datasets that capture information of drivers in the surveyed papers, as well as some of their main characteristics, such as considered emotional states, annotation method, types of signals, and number of participants.

The UTDrive DB Classical [6] was collected in the context of stress and cognitive load, and involved 77 participants undergoing real-world driving in urban and highway scenarios. The emotional annotations were provided by the different conditions, and the collected signals included audio, video, and behavior (pedal pressure, distance with the preceding car, CAN, GPS). Similarly, the same authors collected a variation of this dataset called *UTDrive DB Portable* [6], which only relied on smartphone sensors to collect the data. In particular, they collected video of the face, audio, car acceleration, and GPS location.

The DriveDB dataset [43] was collected in the context of stress recognition and involved 17 participants undergoing from 54 to 93 minutes of real-life driving. Similarly, the emotional annotations were provided by the different driving conditions (e.g., highway, city), and the collected data included multiple biophysiological signals (ECG, electromyogram (EMG), EDA, and RESP).

Table 6. Public Tools Used by Surveyed Papers

Tool	Purpose
Annotation Tool [69]	Annotation
TORCS [116]	Car simulation
OpenCV [123]	Image analysis
OpenFace [5]	Image analysis
BioSig - Matlab [117]	Biosignal analysis
PhysioToolkit [35]	Biosignal analysis
OpenSmile [30]	Speech analysis
Praat [9]	Speech analysis
Snack Sound Toolkit [114]	Speech analysis
Wavesurfer [115]	Speech analysis
GPSBabel [36]	GPS-to-map conversion
Matlab Bayes Net [78]	Machine larning
Weka [41]	Machine learning

The database in Ma et al. [69] was collected in the context of four driving emotional states (happy, bothered, concentrated, confused) and involved 10 participants driving for around 24 km each. The emotional annotations were provided by several external observers (six annotations per segment), and the collected data were focused on face videos. However, the available dataset only contains facial features associated with facial expressions due to privacy reasons.

Finally, the CIAIR database [60] collected recordings from real-world driving and involved more than 500 subjects driving about 60 minutes each. No emotional annotations were provided, but the collected data included multi-channel video from three cameras, multi-channel audio from 16 microphones, and GPS signals.

Although these databases have helped advance research in driver affective state recognition, there is a significant need for larger (>500 people) datasets that also include annotations of driver state, and that take place under real-world (non-laboratory) driving conditions.

7.2 Research Tools

A lot of time can be saved by using tools recently developed and shared for pre-processing and analyzing the kinds of data often used in driver emotion recognition research and for eliciting and annotating states of interest. Table 6 summarizes the public research tools that were utilized by the reviewed papers for some part of their automotive emotion analysis.

In the context of emotion elicitation and annotation, researchers have used the Open Racing Car Simulator (TORCS), which provides a portable multi-platform car racing simulation with the possibility of adding customized content [116]. Ma et al. [69] developed a tool to help provide quick annotations of video segments. In particular, the tool offers two annotation modes depending on the targeted states: one for bothersome and happiness, and another one for concentration and confusion.

In terms of signal pre-processing, a wide variety of tools have been used that are usually focused on the analysis of a single signal modality. For instance, BioSig [117] and Physiotoolkit [35] have been used to extract features from biophysiological signals; OpenCV [123] and OpenFace [5] have been used to detect faces and identify regions of interest in face images; OpenSmile [30], Praat [9], Snack [114], and Wavesurfer [115] have been used to perform non-linguistic analysis of speech signals; and GPSBabel [36] has been used to connect GPS signals with other mapping programs.

In terms of emotion analysis and recognition, researchers have used different platforms that allow the analysis, training, and evaluation of the models. Some popular choices include Weka [41] and some MATLAB libraries (e.g., Bayes) [78].

8 DISCUSSION

Emotions while driving are a critical component of road safety and driving experience, and they are becoming increasingly important to understand with new systems being developed by automotive manufacturers that attempt to enhance or share the driving experience. Within automotive emotion recognition, this work found and overviewed 63 peer-reviewed studies that met the search criteria. This section discusses some of the main findings, current limitations, and opportunities for future research.

The emotional states that have been most frequently studied in the context of driving are associated with high arousal (83%) and negative valence (56%) such as anger and stress, states that can significantly impact road safety. To provide annotations, researchers mostly relied on one of three main methods: participant self-reports, annotations from external coders, and experimental conditions, with around 36% of the studies using a combination of these. When considering the location in which emotions were studied, around half of the studies were in the laboratory, usually in a driving simulator. The studies taking place under real-world driving conditions still have the limit of relying on controlled emotional events and/or partially controlled routes. Although studying emotions in controlled settings is convenient and valuable [17], the emotions are still not fully representative of those experienced under fully real-world conditions [132]. One main difference is that certain emotions, such as stress in the laboratory or in real life, can manifest quite differently, as potential mistakes while driving on real roads can have dramatically different consequences. In addition, real-life settings can vary significantly in different ways (e.g., different sources of artifacts [34, 62, 88], different display of emotions [69]), which may not be appropriately represented by controlled laboratory studies. Thus, there is still a need to perform completely uncontrolled studies to ensure the maximum generalization of the findings.

The signals that were more frequently measured can be grouped into four main categories based on their originating source: face and head, biosignals, speech, and behavior. Considering all of the signals, the most frequently used ones were biosignals (CAR used in 26 studies, EDA used in 24) closely followed by speech (20), which can be partly explained by their efficacy in capturing high arousal emotional states. The remaining signals appeared in fewer than 10 studies each. Although most of the reviewed studies (92%) considered a single group of signals, they indicated that the collected information was not sufficient to capture the whole complexity of emotions relevant to driver experience. Signals such as facial expressions tend to be better for capturing changes in valence, whereas biosignals and speech tend to be better suited for detecting changes in arousal. It is important to be clear that facial expressions do not always map exactly or via a simple fixed mapping onto the internal state [7]; in general, with machine learning, the combining of multiple signals with context will improve performance—for example, detecting bright sunlight in the eyes, loud speech from an adjacent passenger, or knowing when a driver is entering a busy intersection can help improve inference of the human affective state. Five studies investigated combinations of several signals and demonstrated that multi-modal approaches significantly enhanced emotion recognition performance [47, 72, 73, 95, 104].

Although considering new modalities usually requires adding new sensors, recent research advances suggest that the same type of sensor can be used to extract several types of information. For instance, all of the studies that considered cameras focused on the analysis of head gestures and facial expressions. However, recent advances in computer vision show that cameras can also be used to accurately track different body parts (e.g., [19]) and estimate different physiological parameters (e.g., [90]). Some of these have already been applied in the context of emotion recognition; however, applying them in the automotive domain could bring more insights about the state of drivers and passengers. Similarly, all of the studies that considered microphones with the exception of one [103] focused on the analysis of non-paralinguistic features. However, recent advances in text to speech, as well as signal processing, show that microphones can be used to accurately capture complementary paralinguistic features [103] and physiological parameters [101]. By expanding the amount of information that can be extracted from each signal sensor, future systems would not only increase the recognition power but also would provide sensing redundancy that can help account for potential noise and artifacts. We also see many opportunities likely to soon appear within automobiles including radar sensing that can measure BR and HR of remote people unobtrusively [1, 134].

Several of the reviewed studies use the experimental context to elicit specific emotions and/or assume the elicitation of specific emotional states due to certain contextual circumstances, which are subsequently used as emotion annotations. In simulator settings, for instance, it is possible to purposefully change specific contextual parameters like the amount of traffic and the behavior of other road users to elicit certain states, such as frustration (e.g., [48]). However, this is not readily possible in real-world driving conditions. To help address this, researchers have commonly considered the use of the overall contextual environment as an indicator of the emotional state (e.g., driving in the city or the highway for higher and lower stress levels [43]). Another approach involves considering the context as an additional input for the emotion recognition task. For instance, Harris and Nass [42] considered specific driving events such as turnings and overtaking to better identify the source of emotional changes and help improve their recognition performance. As current vehicles are being increasingly equipped with sensors providing rich contextual information, we believe that this type of approach will gain more attention in the future. Besides playing a key role in emotion understanding [17], context is also critical toward the development of meaningful car interventions. For instance, anger caused by other road users or by speech dialogue mistakes may require completely different types of interventions. Therefore, we believe that future studies will need to consider the addition of multiple sources of context.

To perform the analysis of emotions, the studies considered a wide variety of methods ranging from statistical and correlation analysis (27%) to supervised machine learning approaches (73%). Among the supervised machine learning methods, the most popular approach was SVMs (20 studies), followed by nearest neighbor (10 studies) and NNs (10 studies). Although recognition rates were sometimes obtained up to 97% from various studies, we chose to deliberately not report and review the recognition performance across the papers because side-by-side comparisons can be very misleading given that most studies focused on different datasets and addressed slightly different recognition tasks (even formulating the same emotion categories differently). Most of the reviewed studies for automating emotion recognition are initially conducted by processing the data offline so different methods can be compared on a fixed set of data. However, we believe that online emotion recognition systems will gain more relevance in the future to effectively leverage the information when most needed.

Across the 63 surveyed papers, we identified five public datasets that can be used to help establish benchmark comparisons. However, more labeled datasets that contain different driving conditions and measurement modalities would help accelerate research in the field. Finally, one of the main challenges when analyzing such datasets is the large individual differences associated with the expression and manifestations of emotion. To help attenuate this challenge, most studies applied some type of normalization while others added demographic data such as gender [122] to enhance performance. These and studies in different affect-sensing domains (e.g., stress in call centers [46], engagement in education [96]) demonstrate that person-specific adaptations can lead to significant improvements in performance. To help enable such research, longitudinal data collections that monitor the same drivers over long periods of time would be helpful. Furthermore, these datasets would also enable the systematic exploration of deep NNs, which have received renewed interest by their ability to effectively learn from a large number of samples.

Overall, the surveyed studies mainly considered the scenario of a person actively driving the vehicle. It is important to acknowledge that the interest of the automotive industry has been shifting toward semi- and fully autonomous vehicles in which drivers may play an increasingly passive role. Even with such a shift, the industry has expressed an interest in how to engage the driver in relaxation or other states the driver may desire while also enabling the passenger to switch back to being a driver if needed. As systems transition to being more autonomous, the state of alertness and readiness of the driver to efficiently retake control is of large concern, and these states are impacted by emotions. In future scenarios, a fully automated vehicle may capture the emotional state of all vehicle occupants for different purposes, such as helping facilitate their stress management on the way home or helping provide a more productive experience if they choose to work in the car. Alternatively, the car could provide personalized entertainment or news content to car passengers to increase their engagement levels (e.g., [44]) and help reduce their perceived commute time. Although many of the methods and approaches discussed in this survey would still be relevant in such scenarios, different research studies will be needed to address the growing number of questions, including the following: How can the activities of passengers be comfortably sensed inside the car, while respecting personalized needs for privacy? "How can cars change their driving behavior to minimize occupant stress? How can the car help passengers remain entertained or engaged in productive activities during their long commute? We believe that emotionally intelligent vehicles will be critical toward successfully answering these questions.

9 CONCLUSION

Automotive emotion recognition is a research area that is increasingly growing in importance and attention due to the continuous development of sensing technologies and their potential to deliver safer, more productive, and more engaging experiences. To help stimulate research in this area, this article surveys prior research efforts across the peer-reviewed literature that address the problem of automotive emotion recognition and summarizes how these studies address the main challenges, such as the measurement of emotions, sensing of relevant groups of signals, recognition of emotional states, and shaping of interaction to enhance driver experience. We are looking forward to a future in which intelligent emotion understanding of the driver and the passengers is used in meaningful ways to not only improve road safety but also support greater human well-being.

REFERENCES

[1] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, 837–846.

[2] Urvashi Agrawal, Shubhangi Giripunje, and Preeti Bajaj. 2013. Emotion and gesture recognition with soft computing tool for drivers assistance system in human centered transportation. In *Proceedingsof the 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC'13)*. 4612–4616. DOI: https://doi.org/10.1109/SMC.2013.785

[3] Ahmet Akbas. 2011. Evaluation of the physiological data indicating the dynamic stress level of drivers. *Scientific Research and Essays* 6, 2 (2011), 430–439. DOI: https://doi.org/10.5897/SRE10.943

[4] Ignacio Alvarez, Karmele Lopez de Ipiña, Shaundra B. Daily, and Juan E. Gilbert. 2012. Emotional adaptive vehicle user interfaces: Moderating negative effects of failed technology interactions while driving. In *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 57–60.

[5] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. *OpenFace: A General-Purpose Face Recognition Library with Mobile Applications. CMU School of Computer Science*, Pittsburgh, PA.

[6] Pongtep Angkititrakul, John H. L. Hansen, Sangjo Choi, Tyler Creek, Jeremy Hayes, Jeonghee Kim, Donggu Kwak, Levi T. Noecker, and Anhphuc Phan. 2009. UTDrive: The smart vehicle project. In *In-Vehicle Corpus and Signal Processing for Driver Behavior,* K. Takeda, J. H. L. Hangen, H. Erdogan, and H. Abut (Eds.). Springer, 55–67. DOI: https://doi.org/10.1007/978-0-387-79582-9

[7] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68. DOI: https://doi.org/10.1177/1529100619832930

[8] Ahahina Begum, Mobyen Uddin Ahmed, Ahmed Mobyen Uddin, Peter Funk, and Reno Filla. 2012. Mental state monitoring system for the professional drivers based on heart rate variability analysis and case-based reasoning. In *Proceedings of the Federal Conference on Computer Science and Information Systems (FedSIS'12).* 35–42. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6354476.

[9] Paul Boersma. 2014. The use of Praat in corpus research. In *The Oxford Handbook of Corpus Phonology,* J. Durand, U. Gut, and G. Kristoffersen (Eds.). Oxford Handbooks Online, 342–360. DOI: https://doi.org/10.1093/oxfordhb/9780199571932.013.016

[10] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews* 44 (2014), 58–75.

[11] Hynek Boril, Pinar Boyraz, and John H. L. Hansen. 2012. Towards multimodal driver's stress detection. In *Digital Signal Processing for In-Vehicle Systems and Safety,* J. H. L. Hansen, P. Boyraz, K. Takeda, and H. Abut (Eds.). Springer, 3–19. DOI: https://doi.org/10.1007/978-1-4419-9607-7

[12] Hynek Boril, Tristan Kleinschmidt, Pinar Boyraz, and John H. L. Hansen. 2010. Impact of cognitive load and frustration on drivers' speech.*Journal of the Acoustical Society of America* 127 (Sept. 2010), 1996. DOI: https://doi.org/10.1121/1.3385171 arxiv:33168

[13] H. Boril, S. O. Sadjadi, and J. H. L. Hansen. 2011. UTDrive: Emotion and cognitive load classification for in-vehicle scenarios. In *Proceedings of the 5th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems (DSP'11).* http://www.utd.edu/~hynek/pdfs/BorilSadjadiHansen_DSP11.pdf.

[14] Wolfram Boucsein. 2012. *Electrodermal Activity.* Springer Science & Business Media.

[15] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59. DOI: https://doi.org/10.1016/0005-7916(94)90063-9 arxiv:0005-7916(93)E0016-Z

[16] Johnell O. Brooks, Richard R. Goodenough, Matthew C. Crisler, Nathan D. Klein, Rebecca L. Alley, Beatrice L. Koon, William C. Logan, Jennifer H. Ogle, Richard A. Tyrrell, and Rebekkah F. Wills. 2010. Simulator sickness during driving simulation studies. *Accident Analysis and Prevention* 42, 3 (2010), 788–796. DOI: https://doi.org/10.1016/j.aap.2009.04.013

[17] John T. Cacioppo and Louis G. Tassinary. 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45, 1 (1990), 16–28. DOI: https://doi.org/10.1037/0003-066X.45.1.16

[18] Hua Cai and Yingzi Lin. 2011. Modeling of operators emotion and task performance in a virtual driving environment. *International Journal of Human Computer Studies* 69, 9 (2011), 571–586. DOI: https://doi.org/10.1016/j.ijhcs.2011.05.003

[19] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17).* DOI: https://doi.org/10.1109/CVPR.2017.143 arxiv:1611.08050

[20] Sailesh Conjeti, Rajiv Ranjan Singh, and Rahul Banerjee. 2012. Bio-inspired wearable computing architecture and physiological signal processing for on-road stress monitoring. *Biomedical and Health Informatics* 1, 0 (2012), 1–7.

[21] Albert C. Cruz and Alex Rinaldi. 2017. Video summarization for expression analysis of motor vehicle operators. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction.* 313–323. DOI: https://doi.org/10.1007/978-3-319-58706-6

[22] Yong Deng, Zhonghai Wu, Chao Hsien Chu, Qixun Zhang, and D. Frank Hsu. 2013. Sensor feature selection and combination for stress identification using combinatorial fusion. *International Journal of Advanced Robotic Systems* 10, 8 (2013), 306. DOI: https://doi.org/10.5772/56344

[23] Ding Ding, Klaus Gebel, Philayrath Phongsavan, Adrian E. Bauman, and Dafna Merom. 2014. Driving: A road to unhealthy lifestyles and poor health outcomes. *PloS One* 9, 6 (June 2014), 1–5. DOI: https://doi.org/10.1371/journal.pone.0094602

[24] Monique Dittrich and Sebastian Zepf. 2019. Exploring the validity of methods to track emotions behind the wheel. In *Proceedings of the International Conference on Persuasive Technology.* 115–127.

[25] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. 2011. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems* 12, 2 (2011), 596–614. DOI: https://doi.org/10.1109/TITS.2010.2092770

[26] Paul Ekman, Richard Davidson, Phoebe Ellsworth, Wallace V. Friesen, Robert Levenson, Harriet Oster, and Erika Rosenberg. 1992. Are there basic emotions? *Psychological Review* 99, 3 (1992), 550–553. DOI: https://doi.org/10.1037/0033-295X.99.3.550 arxiv:arXiv:1011.1669v3

[27] Paul Ekman and Wallace V. Friesen. 1978. *Facial Action Coding System: Investigator's Guide. Consulting Psychologists Press.*

[28] Neska El Haouij, Jean Michel Poggi, Raja Ghozi, Sylvie Sevestre-Ghalila, and Mériem Jaïdane. 2018. Random forest-based approach for physiological functional variable selection for driver's stress level classification. *Statistical Methods & Applications* 28 (2018), 157–185. DOI: https://doi.org/10.1007/s10260-018-0423-5

[29] Luis Eudave and Miguel Valencia. 2017. Physiological response while driving in an immersive virtual environment. In *Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN'17).* 145–148. DOI: https://doi.org/10.1109/BSN.2017.7936028

[30] Florian Eyben, Martin Wöllmer, Tony Poitschke, Björn Schuller, Christoph Blaschke, Berthold Färber, and Nhu Nguyen-Thien. 2010. Emotion on the road—Necessity, acceptance, and feasibility of affective computing in the car. *Advances in Human-Computer Interaction* 2010 (2010), Article 5.

[31] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. OpenSMILE: The Munich versatile and fast open-source audio feature extractor. In *Proceedings of ACM Multimedia.* 1459–1462. DOI: https://doi.org/10.1145/1873951.1874246

[32] Stephen H. Fairclough, Andrew J. Tattersall, and Kim Houston. 2006. Anxiety and performance in the British driving test. *Transportation Research Part F: Traffic Psychology and Behaviour* 9, 1 (2006), 43–52. DOI: https://doi.org/10.1016/j.trf.2005.08.004

[33] Raul Fernandez and Rosalind W. Picard. 2003. Modeling driver's speech under stress. *Speech Communication* 40 (2003), 145–149.

[34] H. Guo, A, Yüce, and J.-P. Thiran. 2014. Detecting emotional stress from facial expressions for driving safety. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'14)*, Vol. 1. 5961–5965.

[35] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. 2014. Physiobank, PhysioToolkit, and PhysioNet. Retrieved May 12, 2020 from https://www.ahajournals.org/doi/full/10.1161/01.cir.101.23.e215.

[36] GPSBabel. 2019. Home Page. Retrieved May 12, 2020 from https://www.gpsbabel.org/.

[37] Michael Grimm, Kristian Kroschel, Helen Harris, Clifford Nass, Bjorn Björn Schuller, Gerhard Rigoll, and Tobias Moosmayr. 2007. On the necessity and feasibility of detecting a driver's emotional state while driving. *Affective Computing and Intelligent Interaction* 4738 (2007), 126–138. DOI: https://doi.org/10.1007/978-3-540-74889-2_12

[38] Markus Groth, Thorsten Hennig-Thurau, and Gianfranco Walsh. 2009. Customer reactions to emotional labor: The roles of employee acting strategies and customer detection accuracy. *Academy of Management Journal* 52, 5 (2009), 958–974. DOI: https://doi.org/10.5465/AMJ.2009.44634116

[39] Oleg Gusikhin, Erica Klampfl, Dimitar Filev, and Yifan Chen. 2011. Emotive driver advisor system (EDAS). In *Informatics in Control, Automation and Robotics. Lecture Notes in Electrical Engineering, Vol.* Springer, 21–36. DOI: https://doi.org/10.1007/978-3-642-19539-6_2

[40] Markus Gutmann, Patrik Grausberg, and Kyandoghere Kyamakya. 2015. Detecting human driver's physiological stress and emotions using sophisticated one-person cockpit vehicle simulator. In *Proceedings of the 2015 Information Technologies in Innovation Business Conference (ITIB'15).* 15–18. DOI: https://doi.org/10.1109/ITIB.2015.7355064

[41] Mark A. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: An update. *SIGKDD Explorations* 11, 1 (2009), 10–18. DOI: https://doi.org/10.1145/1656274.1656278 arxiv:arXiv:1011.1669v3

[42] Helen Harris and Clifford Nass. 2011. Emotion regulation for frustrating driving contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 749–752.

[43] Jennifer A. Healey and Rosalind W. Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* 6, 2 (2005), 156–166. DOI: https://doi.org/10.1109/TITS.2005.848368

[44] Javier Hernandez, Zicheng Liu, Geoff Hulten, Dave DeBarr, Kyle Krum, and Zhengyou Zhang. 2013. Measuring the engagement level of TV viewers. In *Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG'13).* IEEE, Los Alamitos, CA, 1–7.

[45] Javier Hernandez, Daniel McDuff, Xavier Benavides, Judith Amores, Pattie Maes, and Rosalind Picard. 2014. AutoEmotive: Bringing empathy to the driving experience to manage stress. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems*. 53–56.

[46] Javier Hernandez, Rob R. Morris, and Rosalind W. Picard. 2011. Call center stress recognition with person-specific models. In *Affective Computing and Intelligent Interaction*. Lecture Notes in Computer Science, Vol. 6974, Springer, 125–134. DOI: https://doi.org/10.1007/978-3-642-24600-5_16

[47] Stefan Hoch, Frank Althoff, G. McGlaun, and G. Rigoll. 2005. Bimodal fusion of emotional data in an automotive environment. In *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing*. 1085–1088.

[48] Klas Ihme, Christina Dömeland, Maria Freese, and Meike Jipp. 2018. *Frustration in the Face of the Driver: A Simulator Study on Facial Muscle Activity During Frustrated Driving. Interaction Studies.* John Benjamins Publishing Company.

[49] Myounghoon Jeon. 2016. Don't cry while you're driving: Sad driving is as bad as angry driving. *International Journal of Human-Computer Interaction* 32, 10 (2016), 777–790. DOI: https://doi.org/10.1080/10447318.2016.1198524

[50] Myounghoon Jeon, Jason Roberts, Parameshwaran Raman, Jung-Bin Yim, and Bruce N. Walker. 2011. Participatory design process for an in-vehicle affect detection and regulation system for various drivers. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. 271–272.

[51] In Cheol Jeong, Dong Hee Lee, Shin Woo Park, Jae Il Ko, and Hyung Ro Yoon. 2007. Automobile driver's stress index provision system that utilizes electrocardiogram. In *Proceedings of the 2007 IEEE Intelligent Vehicles Symposium*. 652–656. DOI: https://doi.org/10.1109/IVS.2007.4290190

[52] Christian Jones and Ing Marie Jonsson. 2008. Using paralinguistic cues in speech to recognise emotions in older car drivers. In *Affect and Emotion in Human-Computer Interaction*. Lecture Notes in Computer Science, Vol. 4868. Springer, 229–240. DOI: https://doi.org/10.1007/978-3-540-85099-1_20

[53] Christian Martyn Jones and Ing-Marie Jonsson. 2005. Automatic recognition of affective cues in the speech of car drivers to allow appropriate responses. In *Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. 1–10.

[54] Christian Martyn Jones and Ing Marie Jonsson. 2007. Performance analysis of acoustic emotion recognition for in-car conversational interfaces. In *Universal Access in Human-Computer Interaction: Ambient Interaction*. Lecture Notes in Computer Science, Vol. 4555. Springer, 411–420. DOI: https://doi.org/10.1007/978-3-540-73281-5_44

[55] Ing-Marie Jonsson, Clifford Nass, Helen Harris, and Leila Takayama. 2005. Matching in-car voice with driver state: Impact on attitude and driving performance. In *Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. 173–180. DOI: https://doi.org/10.17077/drivingassessment.1158

[56] O. Karaduman, H. Eren, H. Kurum, and M. Celenk. 2013. An effective variable selection algorithm for aggressive/calm driving detection via CAN bus. In *Proceedings of the 2013 International Conference on Connected Vehicles and Expo (ICCVE'13)*. IEEE, Los Alamitos, CA, 586–591.

[57] Salman Karimi and Mohammad Hossein Sedaaghi. 2013. Robust emotional speech classification in the presence of babble noise. *International Journal of Speech Technology* 16, 2 (2013), 215–227. DOI: https://doi.org/10.1007/s10772-012-9176-y

[58] Tomokazu Kato, Haruki Kawanaka, Md. Shoaib Bhuiyan, and Koji Oguri. 2011. Classification of positive and negative emotion evoked by traffic jam based on electrocardiogram (ECG) and pulse wave. In *Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC'11)*. IEEE, Los Alamitos, CA, 1217–1222.

[59] Christos D. Katsis, N. Katertsidis, George Ganiatsas, and Dimitrios I. Fotiadis. 2008. Toward emotion recognition in car racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 38, 3 (2008), 502–512. DOI: https://doi.org/10.1109/TSMCA.2008.918624

[60] Nobuo Kawaguchi and Shigeki Matsubara. 2001. Multimedia data collection of in-car speech communication. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech'01)*. 3–6.

[61] N. Keshan, P. V. Parimi, and I. Bichindaritz. 2015. Machine learning for stress detection from ECG signals in automobile drivers. In *Proceedings of the 2015 IEEE International Conference on Big Data (IEEE Big Data'15)*. 2661–2669. DOI: https://doi.org/10.1109/BigData.2015.7364066

[62] Abhiram Kolli, Alireza Fasih, Fadi Al Machot, and Kyandoghere Kyamakya. 2011. Non-intrusive car driver's emotion recognition using thermal camera. In *Proceedings of the 3rd International Workshop on Nonlinear Dynamics and Synchronization (INDS'11) and the 16th International Symposium on Theoretical Electrical Engineering (ISTET'11)*. IEEE, Los Alamitos, CA, 1–5.

[63] Arun Sai Krishnan, Xiping Hu, Jun-Qi Deng, Li Zhou, Edith C.-H. Ngai, Xitong Li, Victor C. M. Leung, and Yu-kwong Kwok. 2015. Towards in time music mood-mapping for drivers: A novel approach. In *Proceedings of the 5th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications*. 59–66. DOI: https://doi.org/10.1145/2815347.2815352

[64] H. Leng, Y. Lin, and L. A. Zanzi. 2007. An experimental study on physiological parameters toward driver emotion recognition. *Ergonomics and Health Aspects of Work with Computers* 4566 (2007), 237–246. DOI: https://doi.org/10.1007/978-3-540-73333-1_30

[65] Linda J. Levine. 1997. Reconstructing memory for emotions. *Journal of Experimental Psychology: General* 126, 2 (1997), 165.

[66] Y. Lin, H. Leng, G. Yang, and H. Cai. 2007. An intelligent noninvasive sensor for driver pulse wave measurement. *IEEE Sensors Journal* 7, 5 (2007), 790–799. DOI:https://doi.org/10.1109/JSEN.2007.894923

[67] C. Lisetti and F. Nasoz. 2005. Affective intelligent car interfaces with emotion recognition. *Proceedings of 11th International Conference on Human Computer Interaction.* 1–10. https://www.eurecom.fr/fr/publication/1797/download/mm-lisech-050722.pdf.

[68] Andreas Löcken, Klas Ihme, and Anirudh Unni. 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).* 88–93. DOI:https://doi.org/10.1145/3131726.3131744

[69] Zhiyi Ma, Marwa Mahmoud, Peter Robinson, Eduardo Dias, and Lee Skrypchuk. 2017. Automatic detection of a driver's complex mental states. In *Proceedings of the International Conference on Computational Science and Its Applications.* 678–691.

[70] Diana MacLean, Asta Roseway, and Mary Czerwinski. 2013. MoodWings. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA'13).* DOI:https://doi.org/10.1145/2504335.2504406

[71] Marek Malik, J. Thomas Bigger, A. John Camm, Robert E. Kleiger, Alberto Malliani, Arthur J. Moss, and Peter J. Schwartz. 1996. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal* 17, 3 (1996), 354–381.

[72] L. Malta, P. Angkititrakul, C. Miyajima, and K. Takeda. 2008. Multi-modal real-world driving data collection, transcription, and integration using Bayesian network. In *Proceedings of the 2008 IEEE Intelligent Vehicles Symposium.* 150–155. DOI:https://doi.org/10.1109/IVS.2008.4621141

[73] Lucas Malta, Chiyomi Miyajima, Norihide Kitaoka, and Kazuya Takeda. 2011. Analysis of real-world driver's frustration. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (2011), 109–118. DOI:https://doi.org/10.1109/TITS.2010.2070839

[74] D. McNair, M. Lorr, and L. F. Droppleman. 1991. *POMS: Profile of Mood States.* Educational and Industrial Testing Service, San Diego, CA.

[75] Jolieke Mesken, Marjan P. Hagenzieker, Talib Rothengatter, and Dick de Waard. 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 (2007), 458–475. DOI:https://doi.org/10.1016/j.trf.2007.05.001

[76] Tsuyoshi Moriyama. 2012. Face analysis of aggressive moods in automobile driving using mutual subspace method. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR'12).* 2898–2901.

[77] Nermine Munla, Mohamad Khalil, Ahmad Shahin, and Azzam Mourad. 2015. Driver stress level detection using HRV analysis. In *Proceedings of the 2015 International Conference on Advances in Biomedical Engineering (ICABME'15).* 61–64. DOI:https://doi.org/10.1109/ICABME.2015.7323251.

[78] Kevin P. Murphy. 2009. The Bayes Net Toolbox for Matlab. Retrieved May 12, 2020 from https://www.cs.utah.edu/~tch/notes/matlab/bnt/docs/bnt_pre_sf.html

[79] Fatma Nasoz, Christine L. Lisetti, and Athanasios V. Vasilakos. 2010. Affectively intelligent and adaptive car interfaces. *Information Sciences* 180, 20 (2010), 3817–3836. DOI:https://doi.org/10.1016/j.ins.2010.06.034

[80] Fatma Nasoz, Onur Ozyer, Christine L. Lisetti, and Neal Finkelstein. 2002. Multimodal affective driver interfaces for future cars. In *Proceedings of the 10th ACM International Conference on Multimedia.* 319–322. DOI:https://doi.org/10.1145/641007.641074

[81] Clifford Nass, Ing-Marie Jonsson, Helen Harris, Ben Reaves, Jack Endo, Scott Brave, and Leila Takayama. 2005. Improving automotive safety by pairing driver emotion and car voice emotion. In *Proceedings of CHI'05 Extended Abstracts on Human Factors in Computing Systems (CHI EA'05).* 1973. DOI:https://doi.org/10.1145/1056808.1057070

[82] Khairun Nisa'Minhad, Sawal Hamid Md. Ali, Jonathan Ooi Shi Khai, and Siti Anom Ahmad. 2016. Human emotion classifications for automotive driver using skin conductance response signal. In *Proceedings of the 2016 International Conference on Advances in Electrical, Electronic, and Systems Engineering (ICAEES'16).* IEEE, Los Alamitos, CA, 371–375.

[83] M. Oehl, F. W. Siebert, T.-K. Tews, R. Höger, and H.-R. Pfister. 2011. Improving human-machine interaction: A non invasive approach to detect emotions in car drivers. In *Human-Computer Interaction: Towards Mobile and Intelligent Interaction Environments.* Lecture Notes in Computer Science, Vol. 6763. Springer. 577–585. DOI:https://doi.org/10.1007/978-3-642-21616-9_65

[84] Jonathan Shi Khai Ooi and Siti Anom Ahmad. 2016. Driver emotion recognition framework based on electrodermal activity measurements during simulated driving conditions. In *Proceedings of the Conference on Biomedical Engineering and Sciences.* DOI:https://doi.org/10.1109/IECBES.2016.7843475

[85] Pablo Enrique Paredes, Nur Al Huda Hamdan, Dav Clark, Carrie Cai, Wendy Ju, and James A. Landay. 2017. Evaluating in-car movements in the design of mindful commute interventions: Exploratory study. *Journal of Medical Internet Research* 19, 12 (2017), e372. DOI: https://doi.org/10.2196/jmir.6983

[86] Pablo E. Paredes, Francisco Ordonez, Wendy Ju, and James A. Landay. 2018. Fast and furious: Detecting stress with a car steering wheel. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[87] Thomas D. Parsons and Christopher G. Courtney. 2016. Interactions between threat and executive control in a virtual reality Stroop task. *IEEE Transactions on Affective Computing* 9, 1 (2016), 66–75. DOI: https://doi.org/10.1109/TAFFC.2016.2569086

[88] M. Paschero, G. Del Vescovo, L. Benucci, A. Rizzi, M. Santello, G. Fabbri, and F. M. Frattale Mascioli. 2012. A real time classifier for emotion and stress recognition in a vehicle driver. In *Proceedings of the IEEE International Symposium on Industrial Electronics*. 1690–1695. DOI: https://doi.org/10.1109/ISIE.2012.6237345

[89] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA.

[90] Ming Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. 2011. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering* 58, 1 (2011), 7–11. DOI: https://doi.org/10.1109/TBME.2010.2086456

[91] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37 (2017), 98–125. DOI: https://doi.org/10.1016/j.inffus.2017.02.003

[92] Hamidur Rahman, Shaibal Barua, and Begum Shahina. 2015. Intelligent driver monitoring based on physiological sensor signals: Application using camera. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC'15)*. 2637–2642. DOI: https://doi.org/10.1109/ITSC.2015.424

[93] Genaro Rebolledo-Mendez, Angelica Reyes, Sebastian Paszkowicz, Mari Carmen Domingo, and Lee Skrypchuk. 2014. Developing a body sensor network to detect emotions during driving. *IEEE Transactions on Intelligent Transportation Systems* 15, 4 (2014), 1850–1854. DOI: https://doi.org/10.1109/TITS.2014.2335151

[94] Andreas Riener, Alois Ferscha, and Mohamed Aly. 2009. Heart on the road: HRV analysis for monitoring a driver's affective state. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'09)*. 99–106. DOI: https://doi.org/10.1145/1620509.1620529

[95] George Rigas, Yorgos Goletsis, and Dimitrios I. Fotiadis. 2012. Real-time driver's stress event detection. *IEEE Transactions on Intelligent Transportation Systems* 13, 1 (2012), 221–234. DOI: https://doi.org/10.1109/TITS.2011.2168215

[96] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Björn Schuller, and Rosalind W. Picard. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics* 3, 19 (2018), eaao6760.

[97] Daniele Ruscio, Luca Bascetta, Alessandro Gabrielli, Matteo Matteucci, and Lorenzo Mussone. 2017. Collection and comparison of driver/passenger physiologic and behavioural data in simulation and on-road driving. In *Proceedings of the IEEE International Conference on Models and Technologies for Intelligent Transportation Systems*. 403–408.

[98] James A. Russell. 1980. A circumplex model of affect. *Personality and Social Psychology* 39 (1980), 1161–1178.

[99] Aaqib Saeed and Stojan Trajanovski. 2017. Personalized driver stress detection with multi-task neural networks using physiological signals. In *Proceedings of the Conference on Neural Information Processing Systems*. http://arxiv.org/abs/1711.06116

[100] Arun Sahayadhas, Kenneth Sundaraj, and Murugappan Murugappan. 2012. Detecting driver drowsiness based on sensors: A review. *Sensors (Switzerland)* 12, 12 (2012), 16937–16953. DOI: https://doi.org/10.3390/s121216937

[101] Bjorn Schuller, Felix Friedmann, and Florian Eyben. 2013. Automatic recognition of physiological parameters in the human voice: Heart rate and skin conductance. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13)*. 7219–7223. DOI: https://doi.org/10.1109/ICASSP.2013.6639064

[102] Björn Schuller, Manfred Lang, and Gerhard Rigoll. 2006. Recognition of spontaneous emotions by speech within automotive environment. *Tagungsband Fortschritte der Akustik (DAGA'06)*. 57–58. http://www.mmk.ei.tum.de/publ/pdf/06/06sch5.pdf.

[103] Björn Schuller, Gerhard Rigoll, and Manfred Lang. 2004. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. *Acoustics, Speech, and Signal Processing* 1 (2004), 577–580. DOI: https://doi.org/10.1109/ICASSP.2004.1326051

[104] Bjoern Schuller, Matthias Wimmer, Dejan Arsic, Tobias Moosmayr, and Gerhard Rigoll. 2008. Detection of security related affect and behaviour in passenger transport. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH'08)*. 265–268.

[105] Bjoern W. Schuller. 2008. Speaker, noise, and acoustic space adaptation for emotion recognition in the automotive environment. In *Proceedings of the ITG Conference on Voice Communication (SprachKommunikation'08)*. 1–4. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5759973

[106] Liping Shen, Minjuan Wang, and Ruimin Shen. 2009. Affective e-learning: Using emotional data to improve learning in pervasive learning environment related work and the pervasive e-learning platform. *Educational Technology & Society* 12 (2009), 176–189. DOI: https://doi.org/citeulike-article-id:7412147

[107] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological momentary assessment. *Annual Review of Clinical Psychology* 4, 1 (2008), 1–32. DOI: https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[108] Felix W. Siebert, Michael Oehl, and H.-R. Pfister. 2010. The measurement of grip-strength in automobiles: A new approach to detect driver's emotions. In *Advances in Human Factors, Ergonomics, and Safety in Manufacturing and Service Industry,* W. Karwowski and G. Salvendy (Eds.). CRC Press, Boca Raton, FL, 775–782.

[109] Mohamad Hoseyn Sigari, Mahmood Fathy, and Mohsen Soryani. 2013. A driver face monitoring system for fatigue and distraction detection. *International Journal of Vehicular Technology* 2013 (2013), 73–100. DOI: https://doi.org/10.1155/2013/263983 arxiv:263983

[110] Rajiv Ranjan Singh and Rahul Banerjee. 2010. Multi-parametric analysis of sensory data collected from automotive drivers for building a safety-critical wearable computing system. In *Proceedings of the 2010 International Conference on Computer Engineering and Technology (ICCET'10),* Vol. 1.355–360. DOI: https://doi.org/10.1109/ICCET.2010.5486110

[111] Rajiv Ranjan Singh, Sailesh Conjeti, and Rahul Banerjee. 2011. An approach for real-time stress-trend detection using physiological signals in wearable computing systems for automotive drivers. In *Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC'11).* 1477–1482. DOI: https://doi.org/10.1109/ITSC.2011.6082900

[112] Rajiv Ranjan Singh, Sailesh Conjeti, and Rahul Banerjee. 2012. Biosignal based on-road stress monitoring for automotive drivers. In *Proceedings of the 2012 National Conference on Communications (NCC'12).* 8–9. DOI: https://doi.org/10.1109/NCC.2012.6176845

[113] Rajiv Ranjan Singh, Sailesh Conjeti, and Rahul Banerjee. 2013. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control* 8, 6 (2013), 740–754. DOI: https://doi.org/10.1016/j.bspc.2013.06.014

[114] Kare Sjölander. 2004. The snack sound toolkit. http://www.speech.kth.se/snack.

[115] Kåre Sjölander and Jonas Beskow. 2000. Wavesurfer—An open source speech tool. *Interspeech* 4 (2000), 464–467.

[116] SourceForge. 2012. The Open Racing Car Simulator (TORCS). Retrieved May 12, 2020 from http://torcs.sourceforge.net/.

[117] SourceForge. 2019. The BioSig Project. Retrieved May 12, 2020 from http://biosig.sourceforge.net/.

[118] Olga Sourina, Yisi Liu, Qiang Wang, and Minh Khoa Nguyen. 2011. EEG-based personalized digital experience. In *Universal Access in Human-Computer Interaction: Users Diversity.* Lecture Notes in Computer Science, Vol. 6766. Springer, 591–599. DOI: https://doi.org/10.1007/978-3-642-21663-3_64

[119] Ronnie Taib, Jeremy Tederry, and Benjamin Itzstein. 2014. Quantifying driver frustration to improve road safety. In *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI EA'14).* 1777–1782. DOI: https://doi.org/10.1145/2559206.2581258

[120] Ashish Tawari and Mohan Trivedi. 2010. Speech emotion analysis in noisy real-world environment. In *Proceedings of the International Conference on Pattern Recognition.* DOI: https://doi.org/10.1109/ICPR.2010.1132

[121] Ashish Tawari and Mohan M. Trivedi. 2010. Speech based emotion classification framework for driver assistance system. In *Proceedings of the IEEE Intelligent Vehicles Symposium.* 174–178. DOI: https://doi.org/10.1109/IVS.2010.5547956

[122] Ashish Tawari and Mohan Manubhai Trivedi. 2010. Speech emotion analysis: Exploring the role of context. *IEEE Transactions on Multimedia* 12, 6 (2010), 502–509. DOI: https://doi.org/10.1109/TMM.2010.2058095

[123] OpenCV Team. 2019. Open Source Computer Vision Library. Retrieved May 12, 2020 from https://opencv.org/.

[124] Masaharu Terasaki, Youlchl Klshimoto, and Alto Koga. 1992. Construction of a multiple mood scale. *Japanese Journal of Psychology* 62, 6 (1992), 350–356. DOI: https://doi.org/10.4992/jjpsy.62.350

[125] Tessa Karina Tews, Michael Oehl, Felix W. Siebert, Rainer Höger, and Helmut Faasch. 2011. Emotional human-machine interaction: Cues from facial expressions. In *Human Interface and the Management of Information: Interacting with Information.* Lecture Notes in Computer Science, Vol. 6771. Springer, 641–650. DOI: https://doi.org/10.1007/978-3-642-21793-7_73

[126] M. Tischler, C. Peter, M. Wimmer, and J. Voskamp. 2007. Application of emotion recognition methods in automotive research. In *Proceedings of the Workshop on Emotion and Computing—Current Research and Future Impact.* 50–55. http://ias.cs.tum.edu/_media/spezial/bib/tischler07application.pdf.

[127] Geoffrey Underwood, Peter Chapman, Sharon Wright, and David Crundall. 1999. Anger while driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 2, 1 (1999), 55–68. DOI: https://doi.org/10.1016/S1369-8478(99)00006-6

[128] Paul Viola and M. J. Jones. 2004. Robust real-time face detection. *International Journal of Computer Vision* 57, 2 (2004), 137–154. DOI: https://doi.org/10.1023/B:VISI.0000013087.49260.fb arxiv:arXiv:1011.1669v3

[129] Jinjun Wang and Yihong Gong. 2008. Recognition of multiple drivers' emotional state. In *Proceedings of the 2008 19th International Conference on Pattern Recognition.* 1–4. DOI: https://doi.org/10.1109/ICPR.2008.4761904

[130] Jeen Shing Wang, Che Wei Lin, and Ya Ting C. Yang. 2013. A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing* 116 (2013), 136–143. DOI: https://doi.org/10.1016/j.neucom.2011.10.047

[131] D. Watson, L. A. Clark, and A. Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales.*Journal of Personality and Social Psychology* 54, 6 (1988), 1063–1070.

[132] Frank H. Wilhelm and Paul Grossman. 2010. Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology* 84, 3 (2010), 552–569. DOI: https://doi.org/10.1016/j.biopsycho.2010.01.017

[133] Kenton Williams, José Acevedo Flores, and Joshua Peters. 2014. Affective robot influence on driver adherence to safety, cognitive load reduction and sociability. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'14).* 1–8. DOI: https://doi.org/10.1145/2667317.2667342

[134] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking.* ACM, New York, NY, 95–108.

Received November 2019; revised March 2020; accepted March 2020