Investigating Affective Responses toward In-Video Pedestrian Crossing Actions using Camera and Physiological Sensors

Shruti Rao
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
shruti.rao@cwi.nl

Surjya Ghosh
BITS Pilani
Goa, India
surjya.ghosh@gmail.com

Gerard Pons Rodriguez
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
gerardponsr@gmail.com

Thomas Röggla
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
t.roggla@cwi.nl

Abdallah El Ali
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
aea@cwi.nl

Pablo Cesar
Centrum Wiskunde & Informatica
Delft University of Technology
Amsterdam, The Netherlands
p.s.cesar@cwi.nl

ABSTRACT

Automatically inferring drivers' emotions during driver-pedestrian interactions to improve road safety remains a challenge for designing in-vehicle, empathic interfaces. To that end, we carried out a lab-based study using a combination of camera and physiological sensors. We collected participants' (N=21) real-time, affective (emotion self-reports, heart rate, pupil diameter, skin conductance, and facial temperatures) responses towards non-verbal, pedestrian crossing videos from the Joint Attention for Autonomous Driving (JAAD) dataset. Our findings reveal that positive, non-verbal, pedestrian crossing actions in the videos elicit higher valence ratings from participants, while non-positive actions elicit higher arousal. Different pedestrian crossing actions in the videos also have a significant influence on participants' physiological signals (heart rate, pupil diameter, skin conductance) and facial temperatures. Our findings provide a first step toward enabling in-car empathic interfaces that draw on behavioural and physiological sensing to in situ infer driver emotions during non-verbal pedestrian interactions.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**.

KEYWORDS

empathic car, pedestrian behaviour, driver emotion recognition

ACM Reference Format:
Shruti Rao, Surjya Ghosh, Gerard Pons Rodriguez, Thomas Röggla, Abdallah El Ali, and Pablo Cesar. 2022. Investigating Affective Responses toward In-Video Pedestrian Crossing Actions using Camera and Physiological Sensors. In *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22), September 17–20, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3543174.3546842

AutomotiveUI '22, September 17–20, 2022, Seoul, Republic of Korea

ACM ISBN 978-1-4503-9415-4/22/09.
https://doi.org/10.1145/3543174.3546842

1 INTRODUCTION

There is an increasing interest within the automotive industry to develop *empathic* cars[1,2], which can infer driver emotions [24]. This is because human emotions that may arise during driving scenarios (particularly anger or stress) are known to adversely impact driving behaviour [9, 21]. Therefore, identifying these emotions during driving scenarios and conveying this information to drivers such that emotions may be regulated in a timely manner are considered crucial factors for improving road safety [23, 58, 59]. While environmental (weather, roads) and situational (traffic) factors have previously been considered for inferring drivers' emotional states [4, 15, 22], the non-verbal interaction between a driver and pedestrian(s) has received less attention. Considering that pedestrian non-verbal behaviour is often a source of negative driver emotion [58], automatically capturing drivers' affective responses toward pedestrian non-verbal actions can aid in designing empathic, in-vehicle interfaces, thus leading to increased road safety.

To investigate the influence of such pedestrian crossing actions using physiological and camera sensors, we adopt a highly controlled experimental approach, where we show such pedestrian crossing actions through recorded videos. Given the foregoing, in this paper we ask **(RQ)**: how do people's affective responses vary in response to different non-verbal, pedestrian crossing actions shown through video stimuli? To answer this, we conduct an in-lab study where participants with driving experience (*N=21*) watched 10 short videos of driving scenarios (involving different pedestrian actions) from the publicly available *Joint Attention for Autonomous Driving (JAAD)* dataset [11, 41, 42]. Non-verbal pedestrian actions (e.g. a nod or eye contact) towards drivers persist for a very brief period of time and therefore pose a significant challenge when capturing a driver's affective response induced by such pedestrian actions [13, 31, 43]. Moreover, several pedestrians performing different actions may be present at the crosswalk at any given point. Therefore, identifying the relevant pedestrian impacting the driver's affective state is also a non-trivial task [45]. Given this, we select participant (driver) emotion-inducing stimuli (positive and non-positive) from

[1] https://www.theguardian.com/business/2018/jan/23/a-car-which-detects-emotions-how-driving-one-made-us-feel
[2] https://www.irishtimes.com/business/transport-and-tourism/researchers-developing-empathic-car-technology-1.3900701

the JAAD dataset [3] [11]. To ensure that all relevant participant signals are captured, we collect participant signals throughout the entire duration of the experiment and adapt techniques from activity annotation to segment and identify relevant participant signals induced by the stimuli (videos). Specifically, we recorded participants' responses to the videos in the form of emotion self-reports (valence and arousal, based on Russell's Circumplex model of emotion [46]), facial temperatures (using a FLIR Duo Pro R camera), pupil diameter (using a Pupil Labs eye tracker), and physiological signals (heart rate and skin conductance using an Empatica E4 wristband).

Our findings show that our in-lab setup can effectively capture driver affective states from watching videos of non-verbal, pedestrian crossing actions. Specifically, we observe that participants reported higher valence (pleasantness) upon observing positive pedestrian crossing action videos. On the other hand, participants reported higher arousal (excitement) upon watching non-positive pedestrian crossing videos. Additionally, participants' physiological signals (heart rate, skin conductance and pupil diameter) are significantly influenced ($p < 0.05$) by the different (positive versus non-positive) non-verbal, pedestrian crossing actions. These signals also vary significantly ($p < 0.05$) for different levels of participants' valence (positive versus non-positive) and arousal (high versus non-high) scores. Finally, participants facial temperatures also vary significantly for different levels of participants' valence (positive versus non-positive) and arousal (high versus non-high) scores.

Our exploratory work offers two key contributions: **(1)**: Validation of non-verbal, pedestrian crossing stimuli (JAAD videos) that influence participants' affective states though multi-modal physiological and camera sensors. **(2)**: Empirical findings which reveal that non-verbal, pedestrian actions influence participants' self-reported emotions (valence and arousal), physiological signals and facial temperatures. In-car emotion recognition research is particularly interested in determining a driver's high arousal as well as low valence states associated with risky driving [5, 47]. Quantitative results from our study identify positive and non-positive non-verbal pedestrian crossing actions that results in high arousal and low valence participant states. These non-verbal pedestrian actions may aid in identifying risky driving behaviour arising from driver-pedestrian interaction. Moreover, the participant affective cues (physiological, behavioural, and emotion self-reports) may also be used by empathic, in-car interfaces to automatically infer drivers' affective states during driver-pedestrian interactions, as part of an emotion self-regulation framework for improving road safety.

2 RELATED WORK

Several research strands contributed in shaping our work, including prior research on: (a) driver-pedestrian non-verbal interactions in daily driving scenarios, and (b) measurement of in-vehicle drivers' affective responses.

[3]https://data.nvision2.eecs.yorku.ca/JAAD_dataset/

2.1 Non-Verbal Driver-Pedestrian Interactions

Prior work indicates that non-verbal communication (e.g., body posture) between drivers and pedestrians is a key factor influencing driving behaviour [15, 53]. Studies also investigated different aspects of driver-pedestrian interactions at zebra crossings e.g. eye contact before crossing [14, 49, 56]. Researchers demonstrated that pedestrian body language (eg. hand, leg and head movement) towards drivers are important cues that influence positive or negative driver-pedestrian interactions [12, 25, 51]. However, research on driver-pedestrian interaction often tends to overlook the impact of pedestrian actions on drivers' emotional states. Therefore, in this study we contribute to a better understanding of the role that non-verbal pedestrian actions play in influencing a person's affective state by measuring their self-reports, physiological signals and facial temperatures in a controlled video-watching setting.

2.2 Emotion Models and Self-Reports

There are broadly two emotion models - discrete emotion models (e.g. Ekman's six basic emotions model [8], Plutchik's emotion wheel [39]), and dimensional emotion models (e.g. Circumplex emotion model [46], which consider human emotions as a combination of valence and arousal; Pleasure-Arousal-Dominance model [33], which considers human emotions to be a combination of valence (displeasure vs pleasure), arousal (calm vs excitement), and dominance. In an automotive context, a few studies have explored the most frequently occurring discrete emotions during driving scenarios. For example, Mesken et al. [34] found that anxiety occurred most frequently, followed by anger and happiness. Based on users ease of use and popularity across emotion-measurement studies, we employ the Self-Assessment Manikin (SAM) model with valence and arousal dimensions, where each dimension runs on a discrete 9-point scale [3].

2.3 Sensing Emotion Cues from Multi-modal Physiological and Behavioural Signals

In our work, we capture signals from the participants' face and eyes, as well as bio-physiological markers. To capture facial changes, existing approaches identify the regions of interest (ROIs) from thermal images of the face and head region [30, 37, 59]. Bio-physiological signals include cardiography (for eg., electrocardiograph (ECG), heart-rate variability (HRV), heart rate (HR)), electrodermal activity such as Galvanic Skin Response (GSR), as well as respiratory and skin temperature related signals [59]. These signals when captured from the driving context, contain significant noise due to car movements and so is addressed using pre-processing steps like spike removal [52], bandpass filtering [35] and normalisation (between *0* to *1*) to counter the effect of different baselines and physiological ranges (e.g.[52]). Work has shown that physiological signals such as EDA and HR show higher autonomic activity during favourable driving scenarios with the opposite trend during unfavourable situations [1, 19].

Very few works combined multiple modalities to measure drivers' affective responses. For example, Malta et al. combined EDA and Controller Area Network (CAN) behaviour signals to study driver irritation [32]; Rigas et al. combined several bio-physiological signals, CAN-bus data, and the Global Positioning System (GPS) signal to study driver stress [44]. Hoch et al., and Schuller et al. combined speech and face to study different sets of driver emotions[20, 50]. Finally, Bethge et al. developed a novel application to classify drivers' emotions based on contextual driving data and drivers' facial expressions [2]. While the foregoing work has focused on identifying and classifying drivers' emotions using contextual factors such as traffic or environmental conditions, there has been less emphasis on driver-pedestrian non-verbal interactions. Our study provides the first investigation on the relationship between multi-modal physiological and behavioural signals, and pedestrian crossing actions in videos, which contribute a validated set of emotion induction videos.

3 USER STUDY

We designed a lab-based study to investigate participants' affective responses towards video stimuli containing non-verbal, pedestrian crossing actions.

3.1 Study Design

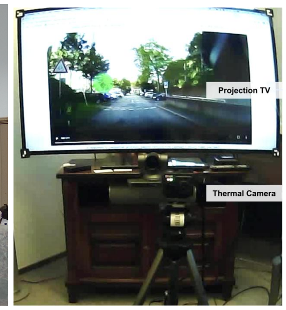
Our experiment is a 1 (IV1: Emotion Rating Task) x 2 (IV2: Pedestrian Crossing Action Video: Positive Action vs. Non-positive Action) within-subjects design, tested in a controlled, laboratory environment. Participants with driving experience watched 10 videos from the JAAD dataset recorded from the driver's perspective. These videos show pedestrians crossing the road and performing non-verbal actions towards the driver such as hand waving, nodding etc [26]. For each video, participants rated pedestrian actions for valence and arousal using the 9-point discrete Self-Assessment Manikin (SAM) [3]. During the study, participants' facial temperatures, pupil diameter and physiological signals were recorded. Our study followed strict guidelines from our institute's ethics and data protection committee.

3.2 Experiment Setup

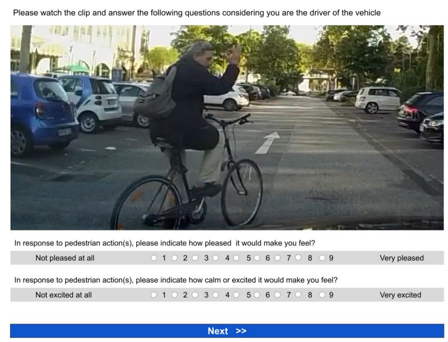
Our in-lab experimental setup consists of the following key components: (a) video stimuli, (b) web interface for viewing video stimuli, and (c) sensors and sensor synchronisation module. Participants are presented with video stimuli through the web interface, that in turn triggers the sensors module to record participants' physiological signals, pupil diameter, and facial temperatures. Figure 1 shows the setup with the web-based user interface for displaying video stimuli and recording participants' emotion ratings.

3.2.1 Video Stimuli. To induce different types (positive, non-positive) of emotions among participants, we draw on a validated set of JAAD dataset videos from a prior study by Ghosh et al. [11]. In this prior study, 91 participants viewed 25 pedestrian action videos from a driver's perspective and rated them for valence (pleasant) and arousal (excitement) on a 5-point scale. Ghosh et al. [11] thereby identified the top-five most positive, and bottom-five most non-positive videos, which we selected for our study. Table 1 shows these 10 JAAD videos, their corresponding pedestrian action and action type along with the average valence ratings obtained in the prior work by Ghosh et al. [11].

(a) Participant wearing the Pupil Labs eye tracker while watching video stimuli.



(b) Thermal Camera and projection screen placement.



(c) The web-based user interface displays the video stimuli and records the participant's valence and arousal ratings after each video.

Figure 1: Experiment setup with thermal camera and projection screen. Participants wear Pupil Labs eye tracker and watch the projection screen which shows the web-based user interface for viewing and rating driver-affect inducing stimuli.

JAAD Video ID	Pedestrian Action	Action Type	5-Point Avg. Valence Rating [11])
video_0299	handwave	Positive	4.03
video_0165	nod	Positive	4.0
video_0135	handwave	Positive	3.92
video_0303	nod	Positive	3.89
video_0249	eye_contact	Positive	3.88
video_0054	handwave	Non-positive	2.79
video_0107	hesitant_crossing	Non-positive	2.77
video_0092	running_in_the_middle	Non-positive	2.47
video_0066	impolite_hand_gesture	Non-positive	2.3
video_0272	engage_with_phone	Non-positive	2.13

Table 1: The 10 JAAD videos used as participant (driver) emotion-inducing stimuli in this study. These 10 videos were identified in a prior study based on the average 5-point valence ratings [11]. Additionally, a positive handwave action video (video_0054) was rated as non-positive by participants [11].

3.2.2 Web-based User Interface. To display the video stimuli and collect valence and arousal self-reports from participants, we developed a web-based interface (Figure 1c), that was projected on a 46" television (full HD, LCD, 1920x1080, 100Hz) shown in Figure 1b. Upon entering demographic details using the laptop mouse, participants pressed the *Next* button that triggers the interface to send a signal to the hardware setup to start recording data from all sensors. Given the stimuli was video only, no audio output was collected from any speakers.

3.2.3 Hardware Setup for Sensor Logs. The hardware setup comprises of the FLIR Duo Pro R thermal camera [4], Empatica E4 wristband [5], and the Pupil Labs Core wearable eye tracker[6] (Figure 1a).

The thermal camera facing the participant (without obstructing their view) is connected to a custom ESP8266 ESP-12 micro-controller, which runs the software for initiating sensor data recording. When powered, the micro-controller starts an HTTP server via WiFi, and awaits commands from the central server. The E4 wristband is connected to an Android mobile device running the EmpaticaRelay application. Once the wristband is switched on, it connects to the software running on the micro-controller, and starts a TCP server to which the central server will connect to fetch data. Finally, the eye tracker is connected to a laptop (MacBook Pro, 1.4 GHz quad core Intel i5, 16GB RAM) running the Pupil Labs Capture software. Once the tracker is connected and calibrated, the setup is complete. Thereafter, the experimenter starts the central recording application, and connects to the sensors via each specified IP address. The setup triggers recording of skin conductance, heart rate, facial temperature, and pupil diameter (PD). Additionally, since pupil diameter is also quite sensitive to light conditions, we fixed the illumination in the lab (350± 5lx) to ensure that participants' pupil would be unaffected by illumination changes [38].

3.3 Study Procedure

Before the experiment, an explanation of the study task was provided to participants, after which we obtained participants' informed consent. After the sensor setup was complete, participants entered their demographic (age, gender, location) and driving experience details (years of experience, country where they mostly drove) on the web interface. Upon entering their details, the sensors were synced and the first video stimuli was shown. Following prior work [29], we ensured 10 seconds long black screens before and after each video to decrease the effects of participants' emotions overlapping between different videos. After each video, participants entered their valence and arousal ratings using a *9-point* discrete SAM scale (Figure 1c). Positive and non-positive action conditions were counterbalanced across all participants, with the subsequent trials randomized. Upon completion of the study session, a brief, semi-structured interview was conducted to gather participants' overall impression of the experiment. The complete experiment lasted approximately 60 minutes and participants were provided with a 10 Euro gift card for participation.

[4]https://www.flir.eu/support/products/duo-pro-r/
[5]https://www.empatica.com/en-gb/research/e4/
[6]https://pupil-labs.com/products/core/

3.4 Participants

For this study, participants were required to be at least 21 years of age and have a minimum driving experience of 1 year. Participants were also required to not wear eyeglasses that may otherwise impact eye tracking. 21[7] participants (7f, 14m) aged between 22-64 *(M=32.4, SD=11.6)* were recruited. Participants were recruited from academic institutes, and comprised diverse cultural backgrounds (66% European, 24% Asian, and 10% North American). 76% of participants had at least three years of driving experience in Western Europe *(M=9.8, SD=10.7)*. None reported visual (including colour blindness), auditory, or motor impairments.

4 STUDY FINDINGS

In this section, we discuss data pre-processing steps undertaken and report results of participants' affective response analysis. Specifically, we discuss: (a) variation across emotion self-reports (b) variation across physiological signals, and (c) variation across facial temperature in different regions of the face with respect to different pedestrian action types. We also summarise the post-study feedback obtained from participants.

4.1 Data Pre-processing

We performed different pre-processing steps prior to analyzing the data (resulting dataset is shown in Table 2). These steps are described below:

4.1.1 Pedestrian Action Segmentation from Video Stimuli. Given our interest in studying affective responses of participants toward pedestrian crossing actions, videos from the JAAD dataset had to be segmented to the relevant aspect of the video ie., the part where the pedestrian action occurred. The duration of the pedestrian actions in the videos were identified by adapting the temporal localisation method which is used in activity annotation [18].

Annotators (*N=7*) from our institute were asked to mark the beginning *start set* and end *end set* of a pedestrian action in the 10 JAAD videos. From this, K-means clustering was used and the centroid of majority clusters (clusters having most data points) for the *start set* and the *end set* were used to mark the beginning and the end of an action [27]. To ensure validity of the annotations, and since time is on a continuous scale, we computed the intra-class correlation coefficient (ICC) for the action start and end points which are 0.997 and 0.980, respectively [40]. We used the time values to extract participants' physiological signals corresponding to the segmented pedestrian action videos. Finally, for each pedestrian, we normalized the physiological signal values to handle inter-subject variability [6, 54]. We scaled as follows:

$$x' = \frac{x - min(X)}{max(X) - min(X)} \tag{1}$$

where X is the set of values recorded for a signal across all individuals, x is one instance of the set X, $min(X)$, $max(X)$ indicate minimum and maximum of the set X.

[7]For effect size f=0.25 under α = 0.05 and power (1-β) = 0.95, with 10 repeated measurements within factors, we need 20 participants.

4.1.2 Valence-Arousal Ratings Transformation. Valence and arousal self-reports corresponding to each video were collected from every participant. In line with the study by Ghosh et al. [11] which revealed that no videos were rated as very negative (valence scores <= 2), valence scores were grouped into *positive* or *non-positive* categories depending on whether they were ≥ 3 or < 3, respectively. Similarly, arousal scores were categorised as *high* or *non-high* scores. Following Russell's dimensional model of emotion, positive versus non-positive valence and high versus non-high arousal relates to emotion categories mapped along the axes of valence and arousal [46].

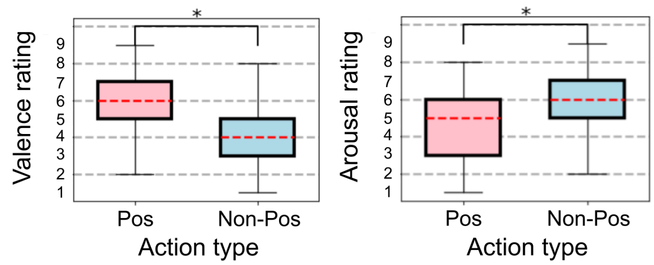
4.1.3 Signal Cleaning and Sensor Sampling. We streamed continuous data from the FLIR thermal camera that recorded thermal images; the wearable eye tracker which recorded pupil diameter, and Empatica wristband which recorded skin conductance in the form of galvanic skin response (GSR) and heart rate in the form of blood volume pulse (BVP). First, missing and incorrectly captured values (for eg. `NaN`) were removed from sensor readings (approximately 3% samples). Furthermore, since the signals had different sampling rates (thermal camera: 30 FPS, eye tracker: 200 Hz, wristband - GSR: 4 Hz and BVP: 64 HZ), we sampled every signal at a uniform rate of 30 Hz (corresponding to facial thermal camera). BVP was further filtered using second order Butterworth lowpass filtering and Stationary Wavelet Transform (SWT) 7th level Daubechies mother wavelet [36]. Inter-beat Interval (IBI) that represents intermittent heart rate [8] was extracted from BVP and used for the analysis. The raw GSR signals were first filtered using a low-pass filter with a 2 Hz cutoff frequency to remove noise. Then, changes were calculated using the mean of the non-negative, first-order differential of GSR signals [10, 55].

Parameter	Values
Total thermal frames	6,594
Total GSR samples	6,594
Total IBI samples	6,594
Total pupil diameter (PD) samples	6,594 (for each eye)
Total valence self-reports	210 (Pos: 60.0%. Non-pos: 40.0%)
Total arousal self-reports	210 (High: 68.6%, Non-high: 31.4%)

Table 2: Final dataset details after pre-processing.

4.2 Emotion Self-report Variation across Pedestrian Action Videos

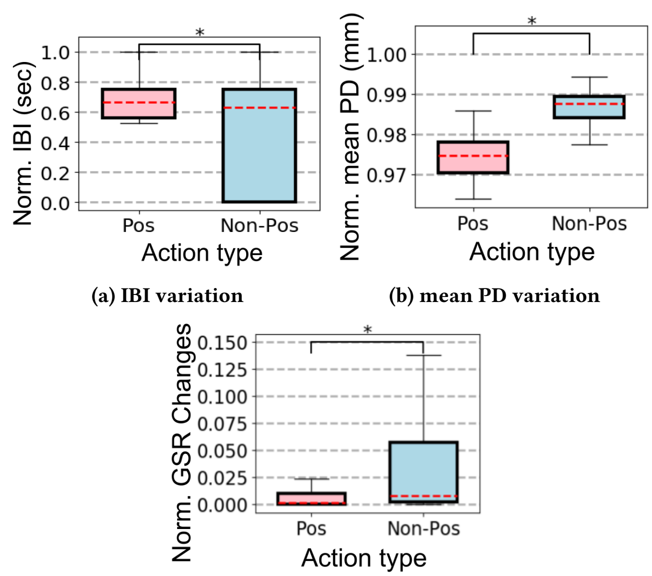
We first examined the variance in emotion self-report (valence, arousal) ratings across *positive* and *non-positive* pedestrian crossing action types as observed in the videos. The median valence ratings for positive and non-positive actions are 6 and 4, respectively. Since the Shapiro-Wilk test revealed that the responses did not follow a normal distribution ($p < 0.05$), we ran a Mann-Whitney U test to evaluate the difference in the responses from the 9-point Self Assessment Manikin (SAM) scale. Figure 2a shows a significant effect of action type on valence ratings (*U = 8317, Z = -6.44, p < 0.05, r = 0.44*). Similarly, the median arousal ratings for the positive and non-positive actions are 5 and 6, respectively. Once again, Mann-Whitney U test revealed a significant effect of action type on the arousal ratings (*U = 3023.5, Z = 5.74, p < 0.05, r = 0.40*), as seen in Figure 2b.



(a) Valence rating comparison **(b) Arousal rating comparison**

Figure 2: Comparison of emotion self-report ratings across different action types present in the video: (a) valence rating (b) arousal rating. Mann-Whitney U test shows both valence and arousal self-report scores to vary significantly ($p < 0.05$) between two action types.

4.3 Physiological Signal Variation across Pedestrian Action Videos



(a) IBI variation **(b) mean PD variation**

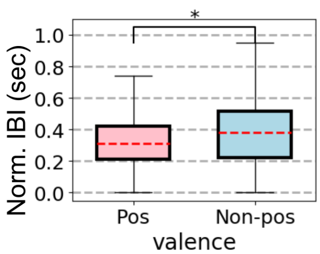
(c) GSR change variation

Figure 3: Variation in physiological signals for different pedestrian action types: (a) IBI variation (b) mean PD variation (c) GSR change variation. All values are found to vary significantly ($p < 0.05$) using Mann-Whitney U test.

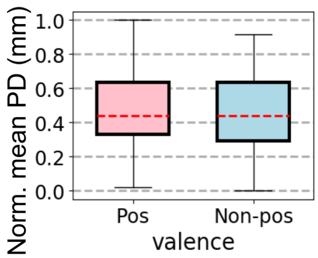
We next investigated the variations in IBI, mean pupil diameter (PD) and GSR changes across *positive* and *non_positive* action types. Box-plots in Figures 3a, 3b, and 3c show these changes between

[8]https://support.empatica.com/hc/en-us/articles/360030058011-E4-data-IBI-expected-signal

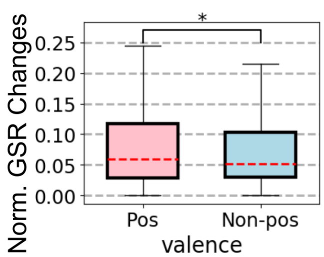
two action types (as observed in the videos) for the IBI, mean PD, and GSR signals respectively. A Shapiro-Wilk test showed that IBI values are not normally distributed ($p < 0.05$)[9]. Despite having an equal number of positive and non-positive actions, the variability in the action duration resulted in unequal samples being collected from the two action types. As a result, we performed an unpaired Mann-Whitney U test. The median IBI (normalized) for positive and non-positive actions are 0.664 and 0.629, respectively. Here, we find a significant effect of action type on the IBI values (*U = 4031220, Z = -4.10, p < 0.05, r = 0.05*). Next, the median value of mean PD (normalized) for positive and non-positive actions are 0.974 and 0.988, respectively. We find a significant effect of action type on mean PD (*U = 209373, Z = 62.574, p < 0.05, r = 0.81*). Finally, median changes in GSR are found to be 0.002 and 0.006 for positive and non-positive actions. We again observe a significant effect of action type on GSR change (*U = 159294, Z = 14.479, p < 0.05, r = 0.37*).



(a) IBI variation



(b) Mean PD variation

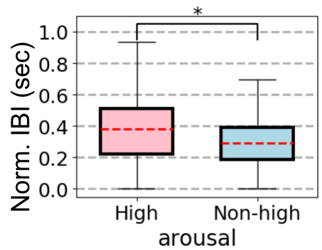


(c) GSR variation

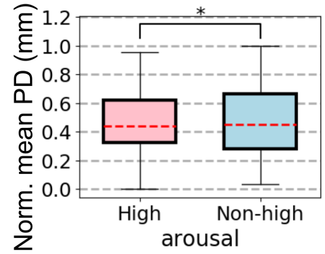
Figure 4: Variation in physiological signals for different level of valence: (a) IBI variation (b) Mean PD (Pupil Diameter) variation (c) GSR variation. GSR and IBI values are found to vary significantly ($p < 0.05$) for valence using Mann-Whitney U test.

We also compared the changes in physiological signals across two levels of self-reported valence (positive vs non-positive) and arousal (high vs non-high) scores. Figure 4 shows the IBI, mean pupil diameter (PD), and GSR changes boxplots across *positive* and *non-positive* levels of valence. Mann Whitney's U tests revealed a significant effect of valence level on IBI values ($U = 2485728$, $p = 0.000$, $r = 0.295$) and GSR values ($U = 6444711$, $p = 0.015$, $r = 0.094$). However, we do not find a significant effect of valence level on mean PD. Similarly, Figure 5 shows the variance in physiological signals for *high* and *non-high* levels of arousal scores. The Mann-Whitney U test shows that arousal level has a significant effect on all three signals: IBI (U = 1939099, $p = 0.000$, $r = 0.357$), mean PD (U = 4693302, $p = 0.017$, $r = 0.098$) and GSR (U = 4249954, $p = 0.000$, $r = 0.075$).

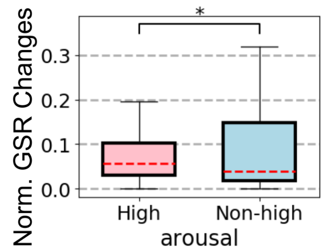
[9]We have the same finding for mean PD and GSR changes.



(a) IBI variation



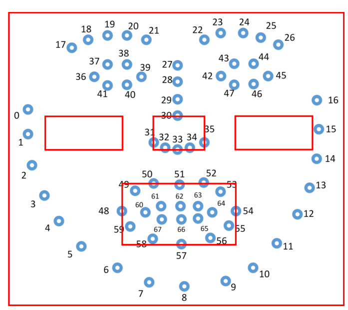
(b) Mean PD variation



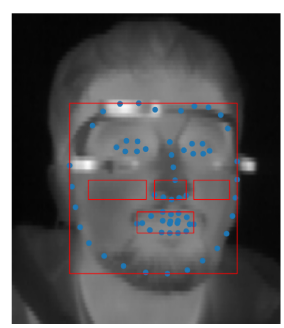
(c) GSR variation

Figure 5: Variation in physiological signals for different level of arousal: (a) IBI variation (b) Mean PD (Pupil Diameter) variation (c) GSR variation. GSR, IBI and mean PD values are found to vary significantly ($p < 0.05$) using Mann-Whitney U test.

4.4 Facial Temperature Variation across Pedestrian Action Videos



(a) ROIs given facial landmarks.

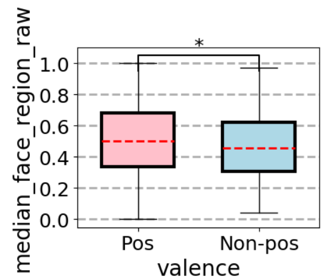


(b) Automatically generated thermal ROIs.

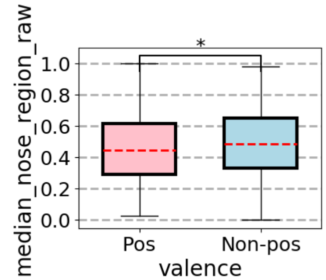
Figure 6: Facial landmarks are used to automatically generate different ROIs on the face from which the thermal features are extracted.

Facial images (frames) captured by the thermal camera were analysed to understand variance in facial temperatures across participant self-reports (valence and arousal). For this, we extracted frame-level median values from different regions of interest (ROIs) of the face: *face, mouth, nose,* and *cheeks (both sides)*.Figure 6 shows a representative thermal image with different ROIs tagged for a subject. We extracted and aggregated median values of ROIs on each frame for all users. We grouped median values independently into two categories based on the self-reported values of valence (positive vs non-positive) and arousal (high vs non-high) and examined the valence-wise and arousal-wise variation for different ROIs. A Shapiro-Wilk test revealed that median values did not follow a normal distribution ($p < 0.05$). The Mann-Whitney U test therefore revealed that the frame-wise median values for all ROIs vary significantly ($p<0.05$ for two levels of valence (Figure 7) and arousal (Figure 8) respectively.

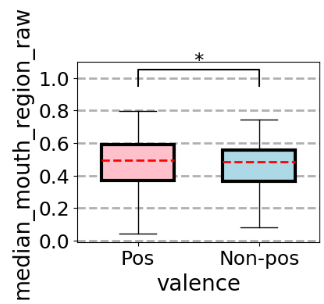
The summary statistics for median valence variation are as follows - Face: $U = 276813407.0, p < 0.05, r = 0.249$; Nose: $U = 285484450.5, p < 0.05, r = 0.156$; Mouth: $U = 291888256.0, p < 0.05, r = 0.154$; Cheek1: $U = 304665302.0, p < 0.05, r = 0.107$; Cheek2: $U = 299927749.5, p < 0.05, r = 0.114$. The summary statistics for median arousal variation are as follows - Face: $U = 252934818.5, Z = 0.000, p < 0.05, r = 0.251$; Nose: $U = 258534915.0, p < 0.05, r = 0.153$; Mouth: $U = 269701814.5, p < 0.05, r = 0.134$; Cheek1: $U = 274233501.0, p < 0.05, r = 0.105$; Cheek2: $U = 274306008.5, p < 0.05, r = 0.068$.



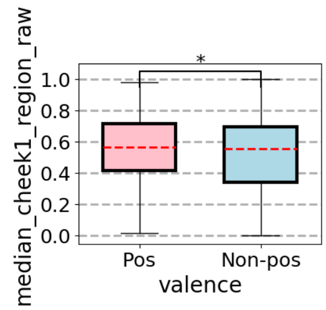
(a) Variation in frame-wise median values



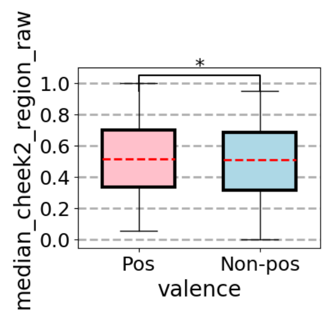
(b) Variation in frame-wise median values



(c) Variation in frame-wise median values



(d) Variation in frame-wise median values



(e) Variation in frame-wise median values

Figure 7: Median valence variation in frame-wise ROIs for: (a) Face (b) Nose (c) Mouth (d) Cheek1 (e) Cheek2. All ROI frame-wise median values vary significantly ($p < 0.05$) across two levels of valence using Mann-Whitney U test.

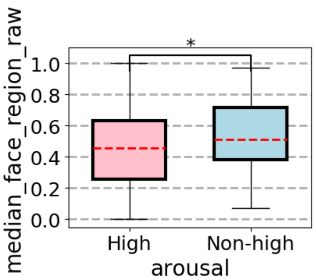
5 DISCUSSION

In this section, we discuss the key findings from our controlled lab study and highlight future steps to be undertaken to address limitations in our work.

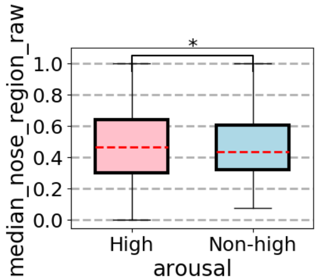
5.1 Key Findings and Implications

To validate whether the positive and non-positive JAAD video stimuli can induce emotion in participants (cf. Section 3.2.1), we designed and executed an exploratory, in-lab setup using a combination of thermal, physiological and eye tracking sensors to record participants' affective states in response to non-verbal, pedestrian crossing videos. Our in-lab study showed the influence of non-verbal, pedestrian actions on participants' physiological responses, facial temperature as well as emotion self-reports: (a) First, we observe that participants' self-reported emotions vary across positive and non-positive pedestrian crossing actions (Figure 2). Positive, non-verbal actions (as shown in the videos) elicit higher valence ratings, whereas non-positive actions (as shown in the videos) elicit higher excitement. (b) We observe that physiological signals (IBI, mean PD, and GSR) vary significantly for positive versus non-positive pedestrian actions (Figure 3). Furthermore, different levels of valence (positive, non-positive) are influenced by pedestrian action types (positive, non-positive) for IBI and GSR signals; while different levels of arousal (high, non-high) are influenced by pedestrian action type (positive, non-positive) for all signals. (c) Similarly, we find variation in facial temperatures across different emotion self-reports. Median values observed at different ROIs (face, mouth, nose, cheeks) of the thermal images are found to vary significantly between different types of actions and valence and arousal self-reports.

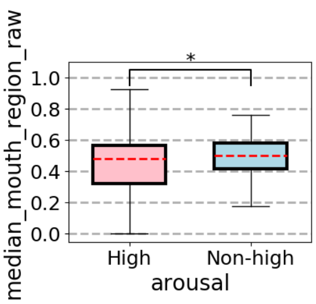
Results from our study have implications in the field of automatic, timely, in-vehicle driver emotion detection and recognition that use machine learning models to infer driver emotions given various driver behavioural and bio-physiological signals. First, our in-lab study validated the suitability of the selected 10 JAAD videos in inducing participant affect, which can be used in future studies as driver affect-inducing stimuli. Moreover, identification of suitable driver affective cues from this study (IBI, GSR, mean PD and facial temperature) can aid researchers in selecting the appropriate sensing modality for detecting emotion signals related to non-verbal, pedestrian crossing actions. This can facilitate the development of (supervised) machine learning models for automatic emotion recognition and subsequent emotion-regulation. Our study also demonstrated the influence of non-positive, pedestrian crossing



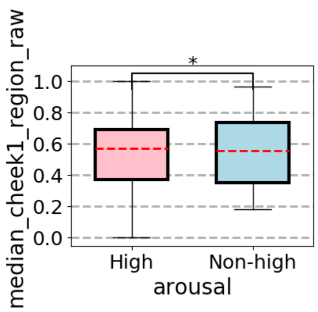
(a) **Variation in frame-wise median values**



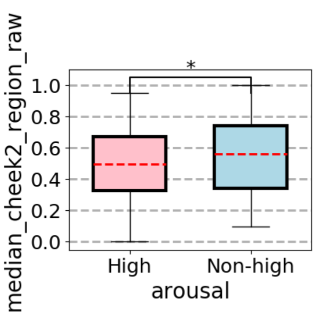
(b) **Variation in frame-wise median values**



(c) **Variation in frame-wise median values**



(d) **Variation in frame-wise median values**



(e) **Variation in frame-wise median values**

Figure 8: Median arousal variation in frame-wise ROIs for: (a) Face (b) Nose (c) Mouth (d) Cheek1 (e) Cheek2. All ROI frame-wise median values vary significantly ($p < 0.05$) across two levels of arousal using Mann-Whitney U test.

actions on participants who reported higher arousal upon viewing such non-positive videos. These pedestrian crossing actions can thereby aid in identification of potential on-road factors that may elicit risky driving behaviour [5, 47], which is a significant area of affective automotive research.

5.2 Towards Just-in-time Interventions using Physiological and Camera Sensors for Emotion Regulation

We observed significant variation in participants' physiological signals upon observing pedestrian crossing action videos. However, the extent to which such signals are robust enough to provide just-in-time interventions, necessary for an empathic vehicle that can facilitate drivers to self-regulate their emotion in situ remains an open question [5]. Our initial results provides a first step towards the development of machine learning models that can leverage such physiological signals for automatic emotion recognition. In a self-regulation context, this can become a binary classification task (for eg. real-time stress detection [17]) during encounters with such pedestrian actions, which can aid subsequent emotion-regulation.

Furthermore, in a real-world driving context expecting drivers to provide self-reports across different intervals is impractical. While our study necessitated the need for establishing a ground truth to investigate if such effects exist in the first place, real-world contexts would benefit from considering other sensing modalities, including camera-based sensors, positioning sensors (e.g., GPS), mapping data (e.g., open street maps[16]), and driving characteristic (e.g., average speed, road type, CAN bus data, etc.) [24]. Cameras in the vehicle allow detecting not only driver facial expressions (which can support the task of automatically identifying in situ emotion expressions), but may also be used for remote physiological marker detection (using for eg. remote Photoplethyography (rPPG) [57] to automatically estimate heart rate). To circumvent the need for widely annotated datasets and extract useful end-to-end features, self-supervised feature learning techniques [48] can be leveraged to make predictions based on the current physiological state of a driver, given the traffic encounter they find themselves in. The specific context of pedestrian crossings would be inferred using a combination of positioning and mapping data. While this allows scaling our approach, some physiological signals such as GSR do require contact-based wearable sensors, which may limit their scaling potential.

5.3 Limitations and Future Work

There are four open challenges that emerged from this study: First, as the study was an in-lab controlled setup with participants that had primarily an academic background, it lacks ecological validity. This is because we do not test users in a real driving context, and participants' background may have influenced their perceptions of the pedestrian crossing actions. Nevertheless, our study aimed to firstly identify whether such pedestrian actions influence emotion perceptions, and how these are reflected using camera and physiological sensors. Our future work will involve designing a hybrid simulator setup with participants comprising diverse professional and educational backgrounds, driving in a simulator but interacting with real-world pedestrian crossing actions.

Second, the stimuli (videos) in our study came from an existing public dataset that depicts real-world driving events. Videos often contained background objects (e.g., cars) or more than one pedestrian crossing at the same time. This made identifying the relevant pedestrian or event from the video impacting a participant's affective state a challenge. Nevertheless, our analysis reveals that the positive and non-positive crossing actions in the videos influenced participants' self reported emotions, physiological signals (IBI, mean PD and GSR) and facial temperature. Future work could isolate the impact of visual stimuli on participants' affective states by also tracking eye movements across the stimuli to pinpoint the regions of the stimuli causing affective changes in participants. Third while this work focused on behavioural and physiological analysis of participants, the effectiveness of our dataset for driver affect prediction such that it may be used for automatic in-vehicle driver emotion detection and regulation remains open. However, our quantitative analysis did reveal that there is statistically significant variation across affective responses to pedestrian actions, which allows future training of supervised machine learning models for carrying out such affect state inferences.

Lastly, while we showed that self-reported valence and arousal levels vary according to the videos observed, we cannot make further inferences regarding the exact emotions drivers may experience in real world. For example, inferring that low valence and high arousal relates to general aggressive driving (cf., [47]) versus a specific situation that elicited such states, would be erroneous. Such inferences would require considering other sensed data, including scene understanding, driving characteristics (e.g., from CAN bus data), and positioning and mapping data. Nevertheless, even with a combined sensing sensing approach, we believe that for any automated emotion regulation intervention stemming from an empathic car, the interaction may still require a final verification from the user to avoid any false positives, which subsequently helps build more robust self-report emotion annotations.

6 CONCLUSION

Inferring driver affective states during non-verbal driver-pedestrian interactions is key for developing empathic, in-car interfaces. This is especially so given that positive, implicit communication between drivers and pedestrians has been known to influence driving behaviour [7, 28, 53]. In this exploratory work, we investigated the impact of non-verbal, pedestrian crossing actions from the JAAD dataset on affective responses (emotion self-reports, physiological responses and facial temperatures) of participants with driving experience. Our in-lab study (*N=21*) revealed the influence of video stimuli (pedestrians crossing and performing non-verbal actions) on participants' valence and arousal self-reports, IBI, GSR, mean pupil dilation and facial temperatures. Specifically, participants reported higher valence and arousal from watching positive and non-positive pedestrian crossing actions. Participants' heart rate (IBI), mean pupil movements (mean PD), skin conductance (GSR) and facial temperatures (across face, mouth, nose and cheeks) significantly varied in response to the pedestrian crossing action observed from the videos. By validating the suitability of video stimuli of non-verbal, pedestrian crossing actions through empirical evidence, our results serve as the basis for development of automatic, in-car empathic interfaces. These interfaces as part of a real-time emotion recognition pipeline can infer drivers' affective states (based on observed pedestrian actions) and aid in "just-in-time" driver emotion regulation for improved road safety.

REFERENCES

[1] Michela Balconi and Adriana Bortolotti. 2012. Empathy in cooperative versus non-cooperative situations: the contribution of self-report measures and autonomic responses. *Applied Psychophysiology and Biofeedback* 37, 3 (2012), 161–169.

[2] David Bethge, Thomas Kosch, Tobias Grosse-Puppendahl, Lewis L Chuang, Mohamed Kari, Alexander Jagaciak, and Albrecht Schmidt. 2021. VEmotion: Using Driving Context for Indirect Emotion Prediction in Real-Time. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York United States, Virtual Event USA, 638–651.

[3] 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.

[4] Michael Braun, Jonas Schubert, Bastian Pfleging, and Florian Alt. 2019. Improving driver emotions with affective strategies. *Multimodal Technologies and Interaction* 3, 1 (2019), 21.

[5] Michael Braun, Florian Weber, and Florian Alt. 2021. Affective Automotive User Interfaces?Reviewing the State of Driver Affect Research and Emotion Regulation in the Car. *ACM Comput. Surv.* 54, 7, Article 137 (sep 2021), 26 pages. https://doi.org/10.1145/3460938

[6] Michael E. Dawson, Anne M. Schell, and Diane L. Filion. 2016. *The Electrodermal System* (4 ed.). Cambridge University Press, Cambridge, 217–243. https://doi.org/10.1017/9781107415782.010

[7] Debargha Dey and Jacques Terken. 2017. Pedestrian interaction with vehicles: roles of explicit and implicit communication. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. Association for Computing Machinery New York NY United States, New York NY United States, 109–113.

[8] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.

[9] Paul Ed Ekman and Richard J Davidson. 1994. *The nature of emotion: Fundamental questions.* Oxford University Press, UK.

[10] Julien Fleureau, Philippe Guillotel, and Izabela Orlac. 2013. Affective benchmarking of movies based on the physiological responses of a real audience. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, IEEE, Geneva, Switzerland, 73–78.

[11] Surjya Ghosh, Gerard Pons Rodriguez, Shruti Rao, Abdallah El Ali, and Pablo Cesar. 2022. Exploring Emotion Responses toward Pedestrian Crossing Actions for Designing In-vehicle Empathic Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (submitted)* (New Orleans, LA, USA) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3491101.3519764

[12] Nicolas Gueguen, Chloé Eyssartier, and Sébastien Meineri. 2016. A pedestrian's smile and drivers' behavior: When a smile increases careful driving. *Journal of safety research* 56 (2016), 83–88.

[13] Nicolas Guéguen, Sébastien Meineri, and Chloé Eyssartier. 2015. A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety science* 75 (2015), 87–89.

[14] Nicolas Guéguen, Sébastien Meineri, and Chloé Eyssartier. 2015. A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety science* 75 (2015), 87–89.

[15] Azra Habibovic. 2018. Communicating intent of automated vehicles to pedestrians. *Frontiers in psychology* 9 (2018), 1336.

[16] Mordechai (Muki) Haklay and Patrick Weber. 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing* 7, 4, Article 1536-1268 (oct 2008), 7 pages. https://doi.org/10.1109/MPRV.2008.80

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

[18] Fabian Caba Heilbron and Juan Carlos Niebles. 2014. Collecting and annotating human activities in web videos. In *Proceedings of International Conference on Multimedia Retrieval*. Association for Computing Machinery New York NY, Taipei Taiwan, 377–384.

[19] Jonathan L Helm, David Sbarra, and Emilio Ferrer. 2012. Assessing cross-partner associations in physiological responses via coupled oscillator models. *Emotion* 12, 4 (2012), 748.

[20] Stefan Hoch, Frank Althoff, Gregor McGlaun, and Gerhard Rigoll. 2005. Bimodal fusion of emotional data in an automotive environment. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Vol. 2. IEEE, IEEE, Philadelphia, PA, ii–1085.

[21] Myounghoon Jeon. 2015. Towards affect-integrated driving behaviour research. *Theoretical Issues in Ergonomics Science* 16, 6 (2015), 553–585.

[22] 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.

[23] Myounghoon Jeon, Jung-Bin Yim, and Bruce N Walker. 2011. An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload. In *Proceedings of the 3rd international conference on automotive user interfaces and interactive vehicular applications*. ACM, ACM NY, NY, 137–142.

[24] Kevin Koch, Varun Mishra, Shu Liu, Thomas Berger, Elgar Fleisch, David Kotz, and Felix Wortmann. 2021. When do drivers interact with in-vehicle well-being interventions? an exploratory analysis of a longitudinal study on public roads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–30.

[25] Julian Francisco Pieter Kooij, Nicolas Schneider, Fabian Flohr, and Dariu M Gavrila. 2014. Context-based pedestrian path prediction. In *European Conference on Computer Vision*. Springer, LNCS, Zurich, Switzerland, 618–633.

[26] Iuliia Kotseruba, Amir Rasouli, and John K Tsotsos. 2016. Joint attention in autonomous driving (JAAD). *arXiv preprint arXiv:1609.04741* NA, NA (2016), 264–269.

[27] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. 2003. The global k-means clustering algorithm. *Pattern recognition* 36, 2 (2003), 451–461.

[28] Andreas Löcken, Carmen Golling, and Andreas Riener. 2019. How should automated vehicles interact with pedestrians? A comparative analysis of interaction concepts in virtual reality. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery New York NY United States, Utrecht, Netherlands, 262–274.

[29] Antoine Lutz, Julie Brefczynski-Lewis, Tom Johnstone, and Richard J Davidson. 2008. Regulation of the neural circuitry of emotion by compassion meditation: effects of meditative expertise. *PloS one* 3, 3 (2008), e1897.

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

[31] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Associated Computing Machinery New York NY United States, Montreal, Canada, 429.

[32] Lucas Malta, Pongtep Angkititrakul, Chiyomi Miyajima, and Kazuya Takeda. 2008. Multi-modal real-world driving data collection, transcription, and integration using Bayesian network. In *2008 IEEE Intelligent Vehicles Symposium*. IEEE, ACM, Eindhoven, Netherlands, 150–155.

[33] Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14, 4 (1996), 261–292.

[34] 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.

[35] Nermine Munla, Mohamad Khalil, Ahmad Shahin, and Azzam Mourad. 2015. Driver stress level detection using HRV analysis. In *2015 international conference on advances in biomedical engineering (ICABME)*. IEEE, IEEE, Beirut, Lebanon, 61–64.

[36] Guy P Nason and Bernard W Silverman. 1995. The stationary wavelet transform and some statistical applications. In *Wavelets and statistics*. Springer, NY, 281–299.

[37] Maurizio Paschero, Guido Del Vescovo, L Benucci, Antonello Rizzi, Marco Santello, Gianluca Fabbri, and FM Frattale Mascioli. 2012. A real time classifier for emotion and stress recognition in a vehicle driver. In *2012 IEEE international symposium on industrial electronics*. IEEE, IEEE, Hangzhou, Zhejiang, China, 1690–1695.

[38] Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. 2016. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, ACM, NY, 5776–5788.

[39] Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89, 4 (2001), 344–350.

[40] Priya Ranganathan, CS Pramesh, and Rakesh Aggarwal. 2017. Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in clinical research* 8, 4 (2017), 187.

[41] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, IEEE, CA, USA, 264–269.

[42] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Are they going to cross? A benchmark dataset and baseline for pedestrian crosswalk behavior. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, CA, USA, 206–213.

[43] Zeheng Ren, Xiaobei Jiang, and Wuhong Wang. 2016. Analysis of the influence of pedestrians' eye contact on drivers' comfort boundary during the crossing conflict. *Procedia engineering* 137 (2016), 399–406.

[44] George Rigas, Yorgos Goletsis, and Dimitrios I Fotiadis. 2011. Real-time driver's stress event detection. *IEEE Transactions on intelligent transportation systems* 13, 1 (2011), 221–234.

[45] Malte Risto, Colleen Emmenegger, Erik Vinkhuyzen, Melissa Cefkin, and Jim Hollan. 2017. Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. National Advanced Driving Simulator, Vermont, Iowa City, IA, 411–420.

[46] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[47] Susan Raouf Hadadi Sani, Zahra Tabibi, Javad Salehi Fadardi, and Despina Stavrinos. 2017. Aggression, emotional self-regulation, attentional bias, and cognitive inhibition predict risky driving behavior. *Accident Analysis & Prevention* 109 (2017), 78–88.

[48] Pritam Sarkar and Ali Etemad. 2020. Self-Supervised Learning for ECG-Based Emotion Recognition. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Barcelona,Spain, 3217–3221. https://doi.org/10.1109/ICASSP40776.2020.9053985

[49] Sarah Schmidt and Berthold Faerber. 2009. Pedestrians at the kerb–Recognising the action intentions of humans. *Transportation research part F: traffic psychology and behaviour* 12, 4 (2009), 300–310.

[50] Björn Schuller, Matthias Wimmer, Dejan Arsic, Tobias Moosmayr, and Gerhard Rigoll. 2008. Detection of security related affect and behaviour in passenger transport. In *Ninth Annual Conference of the International Speech Communication Association*. IEEE, Brisbane, Australia, 265–268.

[51] Andreas Th Schulz and Rainer Stiefelhagen. 2015. Pedestrian intention recognition using latent-dynamic conditional random fields. In *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, IEEE, Seoul, South Korea, 622–627.

[52] 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.

[53] Matus Sucha, Daniel Dostal, and Ralf Risser. 2017. Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention* 102 (2017), 41–50.

[54] Ronnie Taib, Jeremy Tederry, and Benjamin Itzstein. 2014. Quantifying driver frustration to improve road safety. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, Toronto, Canada, 1777–1782.

[55] Chen Wang and Pablo Cesar. 2017. The Play Is a Hit: But How Can You Tell?. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition* (Singapore, Singapore) *(C&C'17)*. Association for Computing Machinery, New York, NY, USA, Article NA, 12 pages. https://doi.org/10.1145/3059454.3059465

[56] Tianjiao Wang, Jianping Wu, Pengjun Zheng, and Mike McDonald. 2010. Study of pedestrians' gap acceptance behavior when they jaywalk outside crossing facilities. In *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, IEEE, Funchal, Portugal, 1295–1300.

[57] Bing-Fei Wu, Yun-Wei Chu, Po-Wei Huang, Meng-Liang Chung, and Tzu-Min Lin. 2017. A Motion Robust Remote-PPG Approach to Driver's Health State Monitoring. In *Computer Vision – ACCV 2016 Workshops*, Chu-Song Chen, Jiwen Lu, and Kai-Kuang Ma (Eds.). Springer International Publishing, Cham, 463–476.

[58] Sebastian Zepf, Monique Dittrich, Javier Hernandez, and Alexander Schmitt. 2019. Towards Empathetic Car Interfaces: Emotional Triggers while Driving. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, Hawai, US, 1–6.

[59] Sebastian Zepf, Javier Hernandez, Alexander Schmitt, Wolfgang Minker, and Rosalind W Picard. 2020. Driver Emotion Recognition for Intelligent Vehicles: A Survey. *ACM Computing Surveys (CSUR)* 53, 3 (2020), 1–30.