ABSTRACT
While affective non-verbal communication between pedestrians and drivers has been shown to improve on-road safety and driving experiences, it remains a challenge to design driver assistance systems that can automatically capture these affective cues. In this early work, we identify users’ emotional self-report responses towards commonly occurring pedestrian actions while crossing a road. We conducted a crowd-sourced web-based survey (N=91), where respondents with prior driving experience viewed videos of 25 pedestrian interaction scenarios selected from the JAAD (Joint Attention for Autonomous Driving) dataset, and thereafter provided valence and arousal self-reports. We found participants’ emotion self-reports (especially valence) are strongly influenced by actions including hand waving, nodding, impolite hand gestures, and inattentive pedestrian(s) crossing while engaged with a phone. Our findings provide a first step towards designing in-vehicle empathic interfaces that can assist in driver emotion regulation during on-road interactions, where the identified pedestrian actions serve as future driver emotion induction stimuli.

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI).

KEYWORDS
empathic, emotion, affective, in-vehicle interface, drivers, pedestrians, actions

ACM Reference Format:

1 INTRODUCTION
Advances in affective computing within the automotive domain have allowed for the design of empathic cars that can automatically infer drivers’ emotions [13, 21, 29]. The primary motivation for the development of empathic cars arises from the need to improve road safety, given that drivers’ emotional states can impact their driving behavior [4, 13, 21, 23, 29, 47]. While state-of-the-art focuses on environmental factors such as traffic and weather, prior work has shown that pedestrian non-verbal actions have an important yet relatively understated impact on drivers [5, 11, 14]. Therefore, there is a potential scope for enhancing vehicles with empathic interfaces that can automatically understand drivers’ affective states and convey this information to drivers using emotion regulation techniques (e.g., ambient displays, bio-feedback) [4, 17, 24, 44]. Actively informing drivers about their emotional states during driving scenarios such as pedestrian intersections can allow for drivers to regulate their emotions in a timely manner and reduce chances of driving errors that include improper control over the steering wheel and pedals, delayed reaction times, and tunnel vision [12].

For this study, we investigated the emotional responses of people with prior driving experience towards frequently occurring pedestrian actions. We conducted a web-based study (N = 91) involving respondents worldwide, who watched short video clips of urban driving scenarios and rated them for valence and arousal. A significant challenge in observing respondents’ affective states is the selection of suitable pedestrian actions that can elicit an emotional response in a respondent [10, 25, 36]. Additionally, several pedestrians may be present at the crosswalk at any given point so identifying the relevant pedestrian impacting the respondent is also a non-trivial task [38]. For this early study, we narrow our scope to the JAAD (Joint Attention for Autonomous Driving) dataset, where we select videos with fewer pedestrians, and pre-determine a fixed set of pedestrian actions. As part of key findings, our study identified 10 videos showing 7 different pedestrian actions from the JAAD dataset that may be used as stimuli to induce driver affect. We demonstrate that respondents’ valence ratings are more influenced by pedestrian actions. In particular, pedestrians waving and nodding...
positively influence respondents’ emotion responses while *impolite hand gestures* or *engaging with phone* have a negative influence. Our findings provide a first step towards designing in-vehicle empathic interfaces that can automatically infer a driver’s affective state as part of an emotion recognition pipeline and thereby assist in driver emotion regulation during pedestrian interactions to improve on-road safety.

2 RELATED WORK

Prior work has shown that pedestrian non-verbal actions (e.g., body posture) play a key role in influencing driving behavior [11, 43]. Many studies investigated different aspects of pedestrian actions at crosswalks such as eye contact before crossing [10, 40, 45], including how often and when they typically occurred [5]. Pedestrian conditions like standing or walking before crossings were also studied [33, 41]. Researchers demonstrated that body language like hand, leg, and head movements are important cues of pedestrian actions [18, 40, 42] and often help drivers to yield [10, 36], thereby generating a positive or negative interaction [9, 16].

Several studies have been conducted to infer drivers’ affective states using different modalities and settings. Driving simulators have been used to observe and measure physiological signals or acoustic responses elicited by drivers based on traffic and pedestrian interactions [15, 20]. Systematic reviews have identified a few *in-the-wild* studies that observe affective states of the driver [48]. Reiner et al. [2009] inferred arousal states of drivers using heart rate variability (HRV), electrocardiography (ECG) and global positioning system (GPS) data [37]. Bethge et al. [2021] developed a novel application to classify drivers’ emotions based on contextual driving data and drivers’ facial expressions. [1]. 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 interactions. This study explores the relationship between self-reported valence and arousal scores and the type of pedestrian actions respondents observed to aid in designing empathic in-vehicle interfaces that can automatically infer driver affective states.

3 WEB-BASED SURVEY

3.1 Survey Design

We designed a web survey using videos from the publicly available and annotated JAAD dataset [33, 34]. This dataset comprises of 5-15 seconds long video clips of urban driving scenarios across different contexts (e.g., weather, time of day) from a driver’s perspective. The dataset also contains 25 videos of pedestrians crossing and performing actions (e.g. *hand wave*, *nod*, *impolite-hand-gesture*) [8, 33], which were selected for the study.

Emotion ratings were based on the Circumplex model of emotion [39], where we look at valence and arousal. Respondents were asked to watch each video and report how pleased (or displeased) and excited (or calm) they would feel as a driver if they observed such an action from a pedestrian. Respondents rated their valence (pleasantness) and arousal (excitement) on a 5-point scale where 1 indicates low valence or arousal, and 5 indicates high valence or arousal [46]. A 5-point scale based on the Self Assessment Manikin (SAM) was used given its popularity in emotion measurement studies [2].

We analyzed the survey questions for reliability and obtained a Cronbach’s alpha score of 0.82, and 0.53 for valence and arousal, respectively.

3.2 Study Procedure

The survey was launched on an online crowd-worker platform\footnote{https://www.microworkers.com/}. Respondents watched the pre-selected pedestrian action videos and provided valence and arousal ratings. To ensure data quality, we (a) eliminated responses completed in less than 3 minutes, as this indicates that the respondent may not have seen all the videos, and (b) eliminated responses with identical valence and arousal ratings for all videos. This resulted in 91 valid respondent responses.

3.3 Study Respondents

The study recruited respondents with prior driving experience, who were awarded $1 to fill an approximately 8 minutes long survey\footnote{https://forms.gle/VQrbwi2hwkeFof6X6}. From the 91 finalized respondents, 82.4% identified as male, and 17.6% as female. 18% were within 17 to 25 years of age, 58% were within 25 to 35 years, and around 25% were within 35 to 50 years of age. 71% drove at least once per week in the last six months prior to filling in the survey. Most respondents (50%) were located in Europe, followed by Asia (25.3%), Africa (16.5%), and North America (7.7%). Figure 1 summarizes the demographic details.

![Figure 1: Distribution of survey participants: (a) gender-wise (b) age-wise (c) driving-frequency wise and (d) continent-wise](image)

4 SURVEY FINDINGS

4.1 Influence of Pedestrian Actions on Emotion Self-Reports

To identify if an observed pedestrian action influenced respondents’ emotion self-reports, we computed arithmetic and weighted average of valence and arousal ratings corresponding to each video.
Weighted average was calculated as follows: 

\[ m_{\text{wt}} = \frac{\sum_{i=1}^{N} i \cdot n_i}{N} \]

where \( i \) denotes the rating provided by the user, \( n_i \) indicates the number of users who provided the rating \( i \), and \( N \) indicates the total number of users. Given that the arithmetic and weighted average values were the same, we report the arithmetic average in Figures 2a and 2b respectively. We observe that while valence ratings for the videos vary, this is not so for arousal ratings. This demonstrates that pedestrian actions observed in the videos can influence respondents’ emotion self-reports, particularly the valence component of emotion.

4.2 Selection of Emotion Eliciting Actions

We identified the top-5 and bottom-5 videos based on the average valence score. Table 1 that identifies actions (seven in total) for corresponding JAAD video IDs, shows high agreement in respondent ratings for the top-5 and bottom-5 videos. Two top-5 videos (video_0299, video_0165) were rated with very high (>=4) valence scores of 4.03 and 4.00, while none of the bottom-5 videos were rated for very low (<=2) valence scores. Thus, no video was perceived to be very negative. Figure 3 shows frames from the top-2 (handwave, nod), and bottom-2 videos (impolite-hand-gesture, engage-with-phone) respectively.

Figure 4 shows variation of valence ratings per video across respondent characteristics of gender and driving frequency. For all cases, the top-5 videos (JAAD ID: video_0165, video_0299, video_0303, video_0135, video_0249) were rated with high valence, and the bottom-5 videos (JAAD ID: video_0272, video_0066, video_0092, video_0107, video_0054) were typically rated with low valence ratings. These indicate that the selected 10 videos (c.f., Table 1) may be used as stimuli to induce different emotions towards pedestrian actions. To investigate variance in valence ratings across gender in these videos, we performed an unpaired Mann-Whitney test, given that scores were not normally distributed in both cases. Since driving frequency is a relevant variable in automotive studies, we also grouped driving frequency into high (every day, and every week) versus low frequency (every two weeks, every month, rarely, did not drive in the last 6 months [30, 31]). However, we found no significant effect of gender (\( U = 59304, p = 0.761 \)) or driving frequency (\( U = 39, p = 0.215 \)) on valence scores for these videos.

4.2.1 Action-wise Valence and Arousal. We additionally examined average valence and arousal scores by actions (Table 2). We observe that average valence scores for actions like handwave, nod, and eye contact are perceived to be pleasant, while actions like impolite-hand-gesture and engage-with-phone are perceived as non-pleasant. Average arousal ratings however appear to be uniform across actions (neutral). We also compared the variation in valence and arousal scores based on action types by grouping them into two categories - pos and non_pos. Since no action was rated as very negative (c.f., Table 1), we refer to these bottom-5 actions as non-positive actions, rather than negative actions. The pos actions group

Table 1: Ranking top-5 and bottom-5 videos based on average valence score for corresponding action. Participants had tendency to rate top-5 videos with higher score (>3) while rating bottom-5 videos with lower scores (<3).

<table>
<thead>
<tr>
<th>Rank (JAAD Id)</th>
<th>Action</th>
<th>Average Valence</th>
<th>Participants (%) rated &lt;3</th>
<th>Participants (%) rated =3</th>
<th>Participants (%) rated &gt;3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (video_0299)</td>
<td>handwave</td>
<td>4.03</td>
<td>8.8</td>
<td>18.7</td>
<td>72.5</td>
</tr>
<tr>
<td>2 (video_0165)</td>
<td>nod</td>
<td>4.0</td>
<td>7.7</td>
<td>18.7</td>
<td>73.6</td>
</tr>
<tr>
<td>3 (video_0135)</td>
<td>handwave</td>
<td>3.92</td>
<td>7.7</td>
<td>27.5</td>
<td>64.8</td>
</tr>
<tr>
<td>4 (video_0249)</td>
<td>nod</td>
<td>3.99</td>
<td>12.1</td>
<td>19.8</td>
<td>68.1</td>
</tr>
<tr>
<td>5 (video_0249)</td>
<td>eye_contact</td>
<td>3.88</td>
<td>13.2</td>
<td>22.0</td>
<td>64.8</td>
</tr>
<tr>
<td>21 (video_0054)</td>
<td>handwave</td>
<td>2.79</td>
<td>42.9</td>
<td>20.9</td>
<td>36.3</td>
</tr>
<tr>
<td>22 (video_0107)</td>
<td>hesitant_crossing</td>
<td>2.77</td>
<td>45.1</td>
<td>26.4</td>
<td>28.6</td>
</tr>
<tr>
<td>23 (video_0092)</td>
<td>running_in_the_middle</td>
<td>2.47</td>
<td>56.0</td>
<td>19.8</td>
<td>24.2</td>
</tr>
<tr>
<td>24 (video_0066)</td>
<td>impolite_hand_gesture</td>
<td>2.3</td>
<td>59.3</td>
<td>16.5</td>
<td>24.2</td>
</tr>
<tr>
<td>25 (video_0272)</td>
<td>engage_with_phone</td>
<td>2.13</td>
<td>64.8</td>
<td>16.5</td>
<td>18.7</td>
</tr>
</tbody>
</table>

Table 2: Participants’ average valence and arousal scores for actions reveals greater variation for valence than arousal.

<table>
<thead>
<tr>
<th>Action</th>
<th>Average valence</th>
<th>Average arousal</th>
</tr>
</thead>
<tbody>
<tr>
<td>handwave</td>
<td>3.58</td>
<td>3.12</td>
</tr>
<tr>
<td>nod</td>
<td>3.95</td>
<td>3.00</td>
</tr>
<tr>
<td>impolite_hand_gesture</td>
<td>2.30</td>
<td>3.32</td>
</tr>
<tr>
<td>engage_with_phone</td>
<td>2.13</td>
<td>3.15</td>
</tr>
<tr>
<td>running_in_the_middle</td>
<td>2.47</td>
<td>3.15</td>
</tr>
<tr>
<td>eye_contact</td>
<td>3.88</td>
<td>3.00</td>
</tr>
<tr>
<td>hesitant_crossing</td>
<td>2.77</td>
<td>3.01</td>
</tr>
</tbody>
</table>
Figure 4: Comparison of average valence scores for top-5 and bottom-5 ranked videos from participants grouped by (a) gender, and (b) driving frequency. All groups generally provided higher ratings for top-5 videos and lower ratings for bottom-5 videos.

Figure 5: Comparison of (a) valence scores and (b) arousal scores for pos, non_pos action types. A significant (p < 0.001) effect of action type is only observed for valence scores.

5 DISCUSSION

5.1 Implications

Identifying drivers’ affective states during on-road interactions with pedestrians is important for ensuring overall driving and pedestrian on-road safety. This is especially given that positive, implicit communication between drivers and pedestrians has been known to influence driving behavior [5, 22, 43]. Our key findings first and foremost demonstrate that pedestrians’ non-verbal actions influence respondents’ emotions across different respondent groups. Moreover, we identified 10 videos with 7 pedestrian actions (waving, nodding, eye contact, hesitant crossing, running in the middle, impolite hand gesture, engaging with phone) from the JAAD dataset that may be used as stimuli to induce driver affect. Particularly, positive actions of hand waving and nodding induce pleasant self-reports, while non-positive actions such as impolite hand gestures or engaging with phone induce non-pleasant ones. These cues can later be automatically inferred and communicated to drivers as part of driving intervention strategies, making the driver emotionally cognizant during driver-pedestrian interactions leading to safer road environments.

We have also shown that self-reported valence scores vary more than arousal scores (Figure 2). This could be due to the lack of real-world driving context, or lack of extreme actions in the study (e.g., aggressive pedestrian behavior was not present in the JAAD dataset). Findings also demonstrate low reliability of arousal scores. This may be attributed to variability in understanding the arousal dimension. Studies have closely linked high arousal levels with road safety [7, 28], where watching short interaction videos in the survey may not present sufficient ecological validity to elicit real-world arousal responses.

5.2 Open Challenges and Future Work

There are several open challenges that emerged from this study, which we aim to consider for our future work. First, given that this work was limited to survey respondents and not real-world actions, it remains a challenge to ensure ecological validity for pedestrian crossing scenarios, while ensuring safety of both drivers and pedestrians. Therefore, in our future work we aim to explore a hybrid (real and virtual) simulator setup where real pedestrians can perform crossing interactions.

Second, there are different pedestrian interactions in real world scenarios that were not accounted for in this study (e.g., facial expressions, different hand actions) [25]. Instead, we identified a set of most frequently occurring actions as a starting point in identifying the effect of pedestrian actions on drivers’ affective states [35]. Future work could therefore be designed to include a greater variety of pedestrian actions, facial expressions, and verbal remarks. Furthermore, certain videos in the study contained multiple pedestrians crossing at the same time, so identifying the relevant pedestrian impacting a respondent’s affective state remains a challenge. Nevertheless, positive and non positive actions from pedestrians in the videos had an influence on participants’ self-reported valence and types ($U = 150812, Z = -13.549, p < 0.001, r = 0.45$). The median arousal score for both groups is 3, resulting in no significant effect of action type on arousal scores.
arousal scores. Future work could isolate the impact of pedestrian actions on drivers’ affective states by having only one pedestrian cross the road at a time.

Given that driving experience has been shown to impact driving behavior, analysis could be further enhanced by collecting users’ driving experience and measuring its impact on self-reported scores [6, 19, 26]. Regardless, we collected demographic data and driving frequency which are also known socio-demographic factors that influence driving behavior [27, 30, 31]. We intend to conduct future work by collecting driving experience, and additional influential variables such as education level and geographical regions predominantly driven in [30, 31]. Finally, we do not account for cultural differences in non-verbal communication between pedestrians and drivers that may vary across countries [32]. For this survey, we only selected videos from the JAAAD dataset, and while cultural differences may exist, we see a pattern in responding towards positive and non-positive actions. For future work, we intend to recruit participants from culturally similar regions.

Our work enables future research to better understand the impact of positive and non-positive actions enacted by pedestrians towards drivers. These driver affect inducing cues can then be automatically inferred as part of an in-vehicle emotion regulation pipeline to predict drivers’ affective states and intervene, particularly in case of negative emotions. By being aware of their emotions, drivers can obtain the opportunity to regulate them and experience safer on-road driving. Together, our preliminary findings provide a first step towards designing in-vehicle empathic interfaces that can recognize and respond to such pedestrian actions.

REFERENCES


