



UNIVERSITY OF LEEDS

This is a repository copy of *Methodologies to Understand the Road User Needs When Interacting with Automated Vehicles*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/153191/>

Version: Accepted Version

---

**Proceedings Paper:**

Portouli, E, Nathanael, D, Amditis, A et al. (8 more authors) (2019) Methodologies to Understand the Road User Needs When Interacting with Automated Vehicles. In: Human-Computer Interaction International (HCI) in Mobility, Transport, and Automotive Systems (Lecture Notes in Computer Science). 21st International Conference on Human-Computer Interaction, 26-31 Jul 2019, Orlando, Florida. Springer , pp. 35-45. ISBN 9783030226657

[https://doi.org/10.1007/978-3-030-22666-4\\_3](https://doi.org/10.1007/978-3-030-22666-4_3)

---

© Springer Nature Switzerland AG 2019. This is an author produced version of a paper published in Lecture Notes in Computer Science. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Methodologies to Understand the Road User Needs when Interacting with Automated Vehicles

Evangelia Portouli<sup>1</sup>, Dimitris Nathanael<sup>1</sup>, Angelos Amditis<sup>1</sup>, Yee Mun Lee<sup>2</sup>, Natasha Merat<sup>2</sup>, Jim Uttley<sup>2</sup>, Oscar Giles<sup>2</sup>, Gustav Markkula<sup>2</sup>, André Dietrich<sup>3</sup>, Anna Schieben<sup>4</sup> and James Jenness<sup>5</sup>

<sup>1</sup> Institute of Communication and Computer Systems, 15773 Zografou, Greece

<sup>2</sup> Institute for Transport Studies, University of Leeds, Leeds, United Kingdom

<sup>3</sup> Chair of Ergonomics, Technical University of Munich, 85748 Garching, Germany

<sup>4</sup> German Aerospace Center, Braunschweig, Germany

<sup>5</sup> Westat, 1600 Research Blvd., Rockville, MD 20850, USA

v.portouli@iccs.gr

**Abstract.** Interactions among road users play an important role for road safety and fluent traffic. In order to design appropriate interaction strategies for automated vehicles, observational studies were conducted in Athens (Greece), Munich (Germany), Leeds (UK) and in Rockville, MD (USA). Naturalistic behaviour was studied, as it may expose interesting scenarios not encountered in controlled conditions. Video and LiDAR recordings were used to extract kinematic information of all road users involved in an interaction and to develop appropriate kinematic models that can be used to predict other's behaviour or plan the behaviour of an automated vehicle. Manual on-site observations of interactions provided additional behavioural information that may not have been visible via the overhead camera or LiDAR recordings. Verbal protocols were also applied to get a more direct recording of the human thought process. Real-time verbal reports deliver a richness of information that is inaccessible by purely quantitative data but they may pose excessive cognitive workload and remain incomplete. A retrospective commentary was applied in complex traffic environment, which however carries an increased risk of omission, rationalization and reconstruction. This is why it was applied while the participants were watching videos from their eye gaze recording. The commentaries revealed signals and cues used in interactions and in drivers' decision-making, that cannot be captured by objective methods. Multiple methods need to be combined, objective and qualitative ones, depending on the specific objectives of each future study.

**Keywords:** Automated Vehicles, Interactions among road users, Methodologies

## 1 Introduction

Interactions among road users play an important role for road safety and fluent traffic [1]. A typical case is when Driver A wishes to turn left at a junction with oncoming traffic. The traffic in the oncoming lane may be so dense that Driver A is uncertain

when it is safe to turn left. Driver A turns on the left indicator and waits. One of the oncoming drivers, Driver B, notices the left indicator and slows down while flashing the vehicle headlights. Driver A perceives this and starts turning left, since they anticipate that this will now be safe. Through similar communicative interactions, drivers in a way purposefully agree or settle on a common future motion plan, each one adapting their own planned future trajectory, so as to enable the safe execution of a manoeuvre.

The above example is a typical case of how humans use multiple means of implicit cues, such as approach speed, and explicit communication, such as eye contact and gestures, as well as vehicle signals, to anticipate the intention of the other road users. Previous research has identified a number of factors influencing both pedestrian-vehicle interactions and vehicle-vehicle interactions in different settings. Drivers can engage in explicit communication with other road users through the use of eye contact, hand gestures, flashing lights and indicator signals, or implicit communication strategies such as speed reduction [2]. Mutual eye-contact has been identified as a factor in facilitating safe interactions between vehicles and Vulnerable Road Users (VRUs) [3], with some research suggesting that establishing eye contact with a driver increases the likelihood that the driver will yield to a pedestrian [4]. Interview data [5] showed that drivers make use of a variety of techniques to force pedestrians to yield, including refusing to decelerate, speeding up, and driving more in the centre of the road to avoid hitting a pedestrian while not stopping for them. Finally, environmental factors such as traffic volume [6], darkness and weather conditions [7, 8], are also likely to affect crossing behaviour. Although the exact means of such interactions may vary across different regions and cultures, it is through such means that effective coordination of future motion plans between different road users is achieved. The phenomenon has not been studied in detail yet, especially as regards interactions among drivers.

Automated vehicles currently lack such interaction capabilities and their behaviour is mostly dominated by the rational principle of collision avoidance. This results in non-human-like, (robotised) behaviour of the automated vehicles, whose actions are not predictable by other road users, and can actually be quite frustrating. Therefore, to safely integrate automated vehicles in complex, mixed, traffic environments, in the future, one must ensure that the automated vehicles can interact with other road users in an intuitive, expectation-conforming manner. This will allow the surrounding road users to correctly interpret the intentions of the automated vehicles, and coordinate their planned actions accordingly.

In order to design appropriate interaction strategies for AVs, observational studies were conducted in Athens (Greece), Munich (Germany) and Leeds (UK), as part of the interACT project “Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments”, funded from the European Union’s Horizon 2020 research and innovation programme, and in Rockville, MD (USA), as part of the NHTSA-sponsored project, “Automated Vehicle Communication and intent with Shared Road Users.” Both projects are connected by a twinning partnership organised by the EU and the US funding organisations. The aim of the studies was to identify interaction-demanding situations, and understand how road users resolve these in current traffic, focussing in particular on the explicit and implicit forms of communication. This paper presents the research objectives of each study, outlines the data collection

methods used and provides an overview of the advantages and disadvantages of each method and of the main research purpose served by each method.

## 2 Observing vehicles and pedestrians' interaction: Cameras and LiDAR observations at European test sites

Cameras were placed at elevated locations (e.g. upper floor or roof of multi-storey building) in Athens, Munich and Leeds to record interactions at predefined use cases. pedestrian-vehicle and vehicle-vehicle interaction at unregulated intersections and shared space parking areas were chosen to identify how road users interact in these differently regulated areas. All videos were recorded in accordance to the data privacy policies of the individual countries. Overall around 600 hours of video data was recorded across all locations using GoPro cameras in Athens and Munich and an HD wireless IP camera in Leeds.

The recorded videos served two purposes: a) to review interactions in traffic that were manually observed and b) to extract the positions of observed road users in each frame. Computer Vision algorithms can be used to extract kinematic information from videos. As no plug and play open source solution was found that was able to detect, track and classify road users in the recorded videos, existing algorithms were adapted and evaluated. These tracking algorithms ranged from simple blob tracking with background reduction to Histogram of Oriented Gradients [9] to convolutional neural networks trained on open source datasets.



**Fig. 1.** Example images of blob tracking

The kinematic data extracted from the videos will help to understand which situations actually require explicit interactions between road users and which situations can be resolved by adjusting the approach velocity.

To have a more accurate account of traffic participant position and velocity, a ground-based LiDAR was utilized to receive synchronized quantitative measurements [10]. The LUX LiDAR sensor provides an object tracking with object properties position, size and velocity of traffic objects. The sensor was integrated in a housing with

power supply, a hard disk storage and a GNSS receiver, to synchronize the LiDAR recordings with the video observations. Overall about 20h of LiDAR data was recorded across all locations.

The point clouds generated by the LiDAR are merged to objects, classified and tracked using Python scripts. Polygons are manually generated used to recreate the road geometry and allow the re-identification of objects that were lost due to short time occlusions. The generated data will be used to understand how the kinematic behaviour from yielding vehicles differentiates from not yielding ones and the condition for co-operative traffic encounters.

### **3 Observing pedestrians and driver behaviour: Manual observations at European test sites**

In addition to the video recordings, three researchers were positioned at each location in Athens, Munich and Leeds, to manually observe the vehicle-vehicle and pedestrian-vehicle interactions. The main purpose of the manual observations was to capture the presence and sequence of any explicit (e.g. hand gestures, signals, honking) and implicit event types (e.g. decelerated for pedestrian, stopped for traffic, accelerated) that was used between these observed road users while interacting with each other at the junctions.

During the data collection for pedestrian-vehicle interactions, three researchers positioned themselves at the designated location, where they were close enough to observe the interaction without interfering in the process. One of the researchers observed the behaviour of the pedestrian and one observed the behaviour of the driver/vehicle. The researchers also spoke out aloud about any event types that was being observed, and this material was recorded. After the end of the interaction (i.e. after a pedestrian had crossed the road), the two researchers then completed an HTML application that was specifically created to record any of these observed behaviours, demographic data of pedestrians observed, as well as the weather and infrastructure details of the observation site. The app also allowed an illustration of the trajectories of the observed road users, if required. The same procedure was conducted for vehicle-vehicle interactions, where one researcher observed the behaviour of one vehicle and one researcher observed the behaviour of the other vehicle.

For the pedestrian-vehicle interactions, a third researcher approached the pedestrian after they had completed their crossing, and asked if they wished to complete a short questionnaire, to provide a subjective measurement of their decision making while crossing the road. This questionnaire included questions about the types of information portrayed by the vehicle and driver that assisted in the crossing decision; how pedestrians themselves indicated their crossing intention; whether the presence of other pedestrian affected their crossing decision, and their familiarity of that particular crossing. These individuals were also asked to complete the Adolescent Road User Behaviour Questionnaire [11]. The data collected from the observation protocol was used to investigate which of these factors predict whether vehicle drove passed the pedestrian or whether pedestrian managed to cross in front of the vehicle, as well as the sequence of

behaviours which led to a crossing.

#### 4 Driving with an eyeglass mounted gaze sensor and retrospective commentary: An on-road study in Athens

An on-road, video-assisted observational study with retrospective commentary by drivers was designed and conducted so as to collect empirical evidence relevant to drivers' interactions with other drivers and pedestrians.

Twenty-one experienced drivers were asked to drive their own passenger car in a predefined urban course, while wearing an eye glass mounted gaze sensor. This system records the traffic scene from the driver's point of view and identifies the driver's eye-fixations points with a 50Hz sampling frequency and gaze position accuracy of 0.5°. The course consisted of a circular route of 0.75 km which was driven 5 times by each driver. The total course length was 3.75 km and the mean driving duration was 18 minutes. The course included left turning from a two-way street and right turning from a smaller to a two-way street. Turns were not regulated by a traffic light and given the traffic density it was expected that there would be a lot of interactions between drivers relevant to the left and right turns. Example traffic scenes are shown in Figure 2.



**Fig. 2.** Examples of eye gaze video recording relevant to left turn from two-way street with on-coming traffic (left) and right turn to two-way street (right)

After arriving at the lab, participants were introduced to the general setup and were calibrated on the eye-tracker, while seated on driver's seat their own passenger car, with a five-point procedure. Then they were instructed to drive at the selected site in their normal style and to repeat the selected course five times in a row. The driving duration was estimated to approximately 15 minutes.

Immediately following the driving session, participants returned to the lab and were asked to watch their eye-gaze video recording while commenting aloud on their behaviour and decision making for each case of interaction with another driver or pedestrian. The commentary was recorded through video and voice capture software. Verbal protocols offer a way to record the human thought process [12] and have been used in driving studies [1].

Afterwards, an analyst watched the participant's eye gaze and scene video as well as his/her retrospective commentary, and labelled the interactions between the participant and another driver. An interaction start with another driver was defined as the time

point when i) the participant had to wait for a gap in the oncoming traffic before turning or ii) the participant started turning knowing that the oncoming driver would have to modify his/her vehicle motion. For each interaction, the analyst labelled the type of the interacting vehicle and whether the other driver reacted. The signals or cues by the participant and his/her vehicle and by the other driver and his/her vehicle and their sequence were labelled for each interaction.

An interaction case with another pedestrian was defined when a pedestrian in the vicinity of the participant driver (i) affected the car movement and/or the driver's behaviour in an observable manner and (ii) received at least one eye-fixation from the driver. The starting point for each interaction case was defined by the observers according to the following criteria: either (i) the drivers' first fixation towards to the pedestrian or (ii) the first cue from the pedestrian interpreted as intention to cross. For each interaction case with a pedestrian, the video data were analysed by labelling the following indices: (i) participant-drivers' eye-fixations on the pedestrians, (ii) eye-contacts between pedestrian and participant-driver, (iii) cues denoting a pedestrian's projected direction (i.e. pedestrian's head orientation, body movement/orientation), (iv) cues denoting pedestrians awareness of the participant's vehicle (i.e. pedestrian's eye-gazes towards to the participant's vehicle). In addition, based on the video-assisted retrospective commentary (v) participants' expressed confidence about the future intended action of a pedestrian was noted when mentioned.

## **5 Real-Time Commentary Used to Study Shared Road Users' Interactions in Rockville, Maryland**

The objective of the Rockville study was to determine the cues that drivers, pedestrians, and bicyclists frequently use when interacting with traffic to perceive drivers' intent and to predict vehicle manoeuvres.

Forty study participants (automobile drivers, pedestrians, and bicyclists) were recruited and trained to perform verbal commentary procedures while engaged in travel through intersections, merge lanes, parking lots, and other situations where interactions between road users occur. For each participant, data collection took place over two sessions including a supervised session where a researcher communicated with the participant, and a naturalistic, unsupervised session where the participant travelled independently without any communication with researchers.

Participating drivers wore a head-mounted GoPro 6 video camera that captured their approximate field of view and recorded audio of both their comments and the accompanying researcher's follow-up questions. Participants drove their own vehicles in this study.

Pedestrians in the study wore a head-mounted GoPro 6 video camera that captured their approximate field of view and recorded audio of their comments. During the supervised data collection session, the accompanying researcher wore a chest-mounted GoPro video camera to capture a view of the participant within the traffic environment and to record audio of follow-up questions. During the supervised data collection

session, both the researcher and participant also used cell phones with wireless earpieces to maintain communications when they were not immediately next to each other.

For the bicyclist participants, two GoPro 6 video cameras were mounted on the participant's bicycle, one was pointed in the forward direction and one was pointed in the reverse direction. During the supervised data collection session, a smartphone was also mounted to the participant's bicycle handlebars with its face camera pointing toward the forward roadway. During the supervised data collection session, a researcher remotely viewed the forward scene and communicated with the participant through a cell phone application that provided a live video phone call. The participant wore a Bluetooth earpiece and microphone to hear and speak to the researcher. The cell phone was not used during the unsupervised data collection session.

For all participants data collection included a supervised session and a naturalistic, unsupervised session. In the first session (supervised) the participant travelled for approximately one hour along a predefined route that was chosen to include traffic situations where road user to road user communication may be necessary to avoid or resolve conflicts. Supervised sessions were scheduled on weekdays during time periods with greater vehicular traffic volumes including morning and afternoon commute hours and midday lunch hours.

At the start of the first data collection session, each participant was trained to use the video cameras and to perform the verbal commentary procedure. The researcher showed a video example of the verbal commentary procedure and then the participant engaged in approximately 10 minutes of practice traveling and commenting prior to starting data collection. The researcher provided navigation instructions. As needed, the researcher also prompted the participant to do more talking aloud, and reminded them to focus their comments on the cues that they were using to determine the actions of nearby vehicles and the intent of nearby drivers. Following traffic interactions where the participant commented, the researcher sometimes asked open-ended follow-up questions to elicit more information such as, "How did you know it was safe for you to cross the street?" or "You mentioned that the driver was going to stop for you, how did you know that?" In all interactions with participants, the researcher was careful to avoid biasing the participant toward reporting any particular vehicle-based or driver-based cues.

In the second data collection session (naturalistic, unsupervised), participants video recorded at least one hour of additional verbal commentary data as they travelled anywhere that they choose to go on public roads. Participants were instructed to restrict their travel to daylight hours and to travel during times of the day with moderate to heavy traffic volume. Pedestrians were urged to find routes that included many street crossings. Prior to conducting unsupervised session, the researcher reviewed instructions for conducting the verbal commentary procedure.

The predefined routes used in this study were located in and around Rockville, Maryland; an urban/suburban city with approximately 65,000 residents that is close to Washington, DC. Different routes were defined for drivers, pedestrians, and bicyclists.

Drivers. The 15-mile route involved driving on both local roads and highways. It took approximately one hour to complete. The route included merges, lane changes, navigating roundabouts, stop signs, right/left turns, U-turns, navigating parking lots,



and periods of driving straight. Along the route there were sixteen scenario locations where participants were prompted to engage in verbal commentary. The researcher rode in the front seat of the participant's vehicle throughout the entire drive.

**Pedestrians.** The 1.3-mile walking route designed for the supervised data collection session involved signalized intersections, controlled intersections with stop signs, mid-block crossings, crossings with pedestrian signals, driveways, entrances and exits to retail establishments, and parking lots. In total there were eighteen planned scenario locations on the route where participants were prompted to engage in verbal commentary. The researcher walked near the participant throughout the session. However, at certain times, the participant was separated from the researcher by a short distance, for example, standing on opposite sides of a street. During these times communication was maintained using hands-free cell phones.

**Bicyclists.** The riding route designed for the supervised data collection session involved signalized intersections, controlled intersections with stop signs, mid-block crossings, right/left turns, driveways, entrances and exits to retail establishments, and bike paths. The route was a loop, approximately 3.5 miles long and participants were instructed to complete the route twice, once in each direction. There were fifteen planned scenario locations along the route where participants were prompted to engage in verbal commentary. Although the researcher did not physically accompany the participant during the ride, communication was maintained using hands-free cell phones running a video calling application.

Video and audio data were downloaded from the GoPro cameras onto a computer where the files were edited using Adobe Premiere software. For pedestrians and bicyclists, when two cameras were used for data collection, the two videos were synchronized and composited into a single split screen view.

Data were analysed separately for drivers, pedestrians, and bicyclists, and for supervised and unsupervised sessions. Researchers reviewed and manually coded the video data from both supervised and unsupervised sessions using Morae Manager software. Participants' comments concerning the cues that they used to assess vehicle manoeuvres and drivers' intentions were categorized and marked with their time-referenced position in the video. For supervised sessions, the comment markers were also referenced to the appropriate set of predefined scenario locations, such as "Roundabout 1" or "Midblock crossing 1." Similarly, cues identified in the data from unsupervised sessions were assigned to generic scenario locations, such as "midblock crossing," or "driving straight."

For the supervised sessions, data collected at each planned scenario location were combined across participants. For the unsupervised sessions, data were combined within the generic scenario locations. Subsequent analyses focused on determining the frequencies for use of different types of cues and how these frequencies differed by types of traffic scenarios. Implicit cues that signal intent such as vehicle movements were compared to explicit cues, such as use of a turn signal. Of special interest was determining how frequently cues coming directly from the driver, such as eye contact, or gestures were used because these cues will not be available or reliable from occupants of highly automated vehicles.

## 6 Discussion

The objective of the presented studies was (i) to support the development of safe AVs by developing kinematic models of behaviour during interactions and (ii) to inform system designers about important existing cues available from human-driven vehicles that may need to be replicated, or replaced by cues from the AV's operational behaviours and/or explicit external signals from an external human machine interface designed to communicate with nearby humans. It is assumed that to ensure safe and efficient interactions between AVs and shared road users, the AV's intent, for example, must be clearly legible, and compatible with shared road users' expectations. Communication through such cues is important for safe and efficient mixed traffic. There is little evidence, however, concerning the signals and cues used by the drivers to infer the future intention of other road users.

Naturalistic behaviour was studied, as it may expose interesting scenarios not encountered in controlled conditions.

A first type of methods used in the studies was video and LiDAR recordings. The aim of these recordings was to extract kinematic information of all road users involved in an interaction and to develop appropriate kinematic models that can be used to predict other's behaviour or plan the behaviour of an AV. While a stationary ground-based LiDAR suffers heavily from occlusion, it directly records points in space with relative positions matching the real world. This is very beneficial compared to the video recordings, as the video data loses accuracy due to distortion and homography. Furthermore, the LiDAR does not generate any personal data and can be used in situations, where installing an elevated camera is unfeasible. On the other hand, video data is simpler to understand and offers more information, such as head rotation or posture of pedestrians. For stationary high angle videos, blob tracking with background reduction works well in different lighting conditions but suffers from ID loss when a traffic participant stands still or gets too close to another road user. Furthermore, classifying tracked objects using the blob size is inaccurate. Detectors using deep learning methods will enhance the tracking results but require large training datasets.

Manual on-site observations of interactions provided additional behavioural information that may not have been visible via the overhead camera or LiDAR recordings. For instance, the camera would not be able to detect any hand movements from pedestrians who were facing away from the camera. On the other hand, these observations could be used to confirm the presence of certain behaviours, recorded by the videos. Using two different methodologies to collect the same data could therefore provide a redundancy gain. The manual on-site observations posed the difficulty that the researchers needed to complete a protocol very quickly, as the interaction frequently evolved very quickly. To assist this process, extensive effort was invested in creating a standardised observation protocol that could be used in different countries, with repeated piloting by members of the team. Following the pilot studies, a list of 98 observable event types was drawn up. Examples of these event types are: drivers' and pedestrians' looking behaviour, observed hand gestures, as well as signals used and movements observed, during an "approaching phase" (approaching the junction) and "crossing phase" (at the junction). At first, the protocols were developed in Microsoft (MS)

Excel and tested by using a pen on printouts. To simplify the data extraction from the observation protocols, enable measurements synchronized in time and reduce the amount of paper used within the observation, the protocols were transferred into an HTML app that was programmed and usable on a variety of smartphones and tablets. The app enabled the researchers to quickly record any of these observed behaviours, as well as demographic data of pedestrians observed, as well as the weather and infrastructure details of the observation site. The app also supported a sketching of the trajectories of the observed road users, if needed. Furthermore, the app enabled the synchronization of the different observation methods, by displaying the device's UNIX time to the observing cameras and logging every input by the ground observers with a timestamp and a sequence number. To be usable in areas with low mobile reception, the HTML app was specifically designed to work offline, saving each observed traffic interaction in a .csv file.

The data from observation protocol and questionnaires also provided an overview of the most common behaviours observed by pedestrians and car drivers during crossings at un-signalised junctions, across the three European cities, allowed the exploration on the frequency and to what extent that a particular event type occurred when the pedestrian and vehicle are at the junction and parking space.

To get a more direct recording of the human thought process, verbal protocols were also applied. Real-time verbal reports seem to provide a more complete and richer representation of pre-reflective cognition and deliver a richness of information that is inaccessible by purely quantitative data [13]. Still, they may pose excessive cognitive workload and remain incomplete. Participants in Maryland were trained to perform the verbal commentary in real time and were encouraged as much as possible to report what they were attending to in real time. In some of the data collection sessions, a researcher also asked participants follow-up questions to clarify and expand upon what had just been said. Such real-time commentaries may remain incomplete in complex traffic environments. This was the case in the Athens study, where due to the density of traffic, participants in the trial runs very frequently did not perform well in the real-time commentary. For this reason, a retrospective commentary was applied, which however carries an increased risk of omission, rationalization and reconstruction. This is why it was applied while the participants were watching videos from their eye gaze recording. The commentaries revealed signals and cues used in interactions and in drivers' decision-making, that cannot be captured by objective methods.

Understanding interactions and behaviour is a complex process and multiple methods need to be combined, objective and qualitative ones, depending on the specific objectives of each study.

**Acknowledgement:** This article is developed from the twinning partnership between the EU project interACT and the US project AVintent. Some work is a part of the interACT project. interACT has received funding from the European Union's Horizon 2020 research & innovation programme under grant agreement no 723395. Content reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains. Funding for the study conducted

in Rockville, MD was provided by the National Highway Traffic Safety Administration under contract DTNH2214D00329L, Task Order 0007.

## References

1. Portouli, E., Nathanael, D., Marmaras, N.: Drivers' Communicative Interactions: On-road observations and modelling for integration in future automation systems. *Ergonomics*, DOI: 10.1080/00140139.2014.952349 (2014).
2. Fuest, T., Sorokin, L., Bellem, H., & Bengler, K.: Taxonomy of traffic situations for the interaction between automated vehicles and human road users. *International Conference on Applied Human Factors & Ergonomics*, 708-719, Cham: Springer (2017)..
3. Schneemann, Friederike and Irene Gohl.: Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments, 2016 IEEE Intelligent Vehicles Symposium (IV): 38-43 (2016).
4. Guéguen, N., Meineri, S., & Eyssartier, C.: A pedestrian's stare and drivers' stopping behavior: a field experiment at the pedestrian crossing. *Safety Science*, 75, 87–89 (2015).
5. Šucha, M.: Road users' strategies and communication: Driver pedestrian interaction. *Proceedings 5th Conference Transport Solutions from Research to Deployment*, Transport Research Arena, Paris (2014).
6. Harrell, W. A.: Factors influencing pedestrian cautiousness in crossing streets. *Journal of Social Psychology*, 131, 367–372 (1991).
7. Klop, J.R., & Khattak, A.J.: Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina. *Transportation Research Review*, 1674, 78–85 (1999).
8. Sayed, T., Zaki, M.H., & Autey, J.: Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis. *Safety Science*, 59, 163–172, (2013)
9. Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, pp. 886-893 vol. 1 (2005).
10. Dietrich A., Ruenz J. (2019) Observing Traffic – Utilizing a Ground Based LiDAR and Observation Protocols at a T-Junction in Germany. In: Bagnara S., Tartaglia R., Albolino S., Alexander T., Fujita Y. (eds) *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. IEA 2018. *Advances in Intelligent Systems and Computing*, vol 823. Springer, Cham
11. Elliott, M.A. & Baughan, C.J.: Developing a self-report method for investigating adolescent road user behaviour. *Transportation Research Part F*, 7, 373-393, (2004).
12. Ericsson, K.A., Simon, H.A.: *Protocol analysis: Verbal reports as data*. MIT Press, Cambridge, MA (1993).
13. Banks, V.A., Stanton, N.A. and Harvey C.: What the drivers do and do not tell you: using verbal protocol analysis to investigate driver behaviour in emergency situations. *Ergonomics*, DOI:10.1080/00140139.2014.884245 (2014).