SmartHelm: User Studies from Lab to Field for Attention Modeling

Mazen Salous, Dennis Küster, Kevin Scheck, Aytac Dikfidan, Tim Neumann, Felix Putze and Tanja Schultz
Cognitive Systems Lab
University of Bremen, Bremen, Germany
mazen.salous@uni-bremen.de

***Abstract*—We present three user studies that gradually prepare our prototype system *SmartHelm* for use in the field, i.e. supporting cargo cyclists on public roads for cargo delivery. SmartHelm is an attention-sensitive smart helmet that integrates none-invasive brain and eye activity detection with hands-free Augmented Reality (AR) components in a speech-enabled outdoor assistance system. The described studies systematically increased in ecological validity from lab to field. The first study consisted of an Augmented Reality preparation examination in the lab. The second study then investigated simulated attention distraction modeling, whereas the third study examined real-world attention distraction modeling while cycling in traffic. During these three studies, multimodal data (EEG, eye-tracking, video, GPS and speech) has been collected synchronously and analyzed in offline and online experiments. Machine Learning models were trained and optimized for attention modeling.**

Results: Analyses of self-report and objective data during the simulation study show the plausibility of the simulated internal and external distractions. The analysis of behavioral data captured by multimodal biosignals recorded in the field study further shows that real visual attention distractions can be automatically identified using synchronized video and eye-tracking data. Machine Learning methods based on long short-term memory models (LSTMs) indicate that simulated attention distractions can be automatically detected from EEG data, with the best detection performance for mental distractions. Finally, the self-report data suggest that the comfort of the SmartHelm helmet should be further improved for permanent use in road traffic.

***Index Terms*—Human-Machine-System, Assistant System, Attention Modeling, Multimodal Data**

I. Introduction

The "last mile" in city logistics is increasingly being covered by cargo cyclists who bring the goods from a transportation hub to the final destination. Day in and day out, cyclists have to transport large numbers of goods in a short amount of time through busy traffic. A complex task that requires full attention and constant adaptation to a wide variety of situations and distractions [24]. Cargo cyclists can therefore use any help that makes their task easier and increases their safety on the road. For this we propose eye- and hands-free assistance systems such as navigation, task planning and communication that present relevant information in a context-sensitive and least disruptive manner [21]. The SmartHelm[1] project aims to improve the safety, stress levels, and productivity of cargo cyclists using a suite of integrated innovative capabilities. Concretely, we are integrating Augmented Reality glasses, Automatic Speech Recognition (ASR), mobile Eye-Tracking (ET), and a none-invasive Brain-Computer-Interface (BCI) based on Electroencephalography (EEG) into a helmet. In this paper, we describe the design and implementation of the SmartHelm user studies for attention modeling. In this context, attention modeling means to automatically detect attention distractions (obstacles [25, 29]) from multimodal data [28]. The approach is based on the interpretation of the cyclists's multimodal biosignals derived from eye and brain activity. We define *biosignals* as autonomous signals produced by a living organism measured in physical quantities using sensors [32].

Driving and cycling are highly complex cognitive tasks, a situation that is exacerbated by the cognitive demands of navigating on busy roads and under time pressure. When employing cargo-bikes for delivering goods in our inner cities, the effects of time pressure and traffic are further compounded by having to mentally plan ahead for the next steps of the delivery process, or by having to pay attention to real-time information presented through hand-held devices. Attention-sensitive Augmented Reality (AR) systems could therefore make a positive contribution to road safety in cargo-cycling, because they could replace the spontaneous use of unsuitable mobile devices while simultaneously helping to avoid the worst spikes in cognitive demands [21]. For cognitive modeling, this implies that safe driving and "situation awareness" depend on the driver's full attention including perceptual abilities (visual and auditory perception), as well as cognitive memory processing, [18]. Attention represents a core property of all of our cognitive and perceptual operations [7]. Prior research has widely distinguished between internal and external attention in various domains [7, 8, 16, 31, 35], thus, we investigate attention distractions in the SmartHelm context accordingly: internal attention distractions, e.g., mental distractions such as mind wandering during cycling, and external attention distractions, such as visual or auditory distractions originating from the surrounding environment [7]. While traditional helmets exclusively aim at saving cyclists in case of accidents,

This work is part of the project SmartHelm, which is funded by the German Federal Ministry of Transport and Digital Infrastructure ("BMVI") as part of the research initiative mFUND" (19F2105F).

[1]SmartHelm (https://smart-helm.com/) is a collaborative project of seven partners from research (Cognitive Systems Lab CSL at the University of Bremen, Very Large Business Analysis VLBA at the University of Oldenburg), city logistics (Oldenburg city, CitiPost and Rytle in Bremen) and industry (TeamViewer and Uvex), that aims to develop an assistance system for cargo cyclist based on an attention-sensitive bicycle helmet, i.e., the "SmartHelm"

the SmartHelm project thus proposes a pro-active approach by modeling the cyclist's attention from multimodal data (EEG and eye-tracking) during biking. Multimodal attention modeling, in this context, means automatically and pro-actively detecting different kinds of attention distractions. Consequently, timely attention re-direction alerts can be issued in the AR application to ensure that the cyclist remains attentive to safety-relevant external events.

In this paper, we focus on the design, multimodal data collections, and analyses of the first evaluation studies that have employed the SmartHelm technology. The user studies detailed in the present work represent the current state of progress towards developing a functional prototype. For each user study, we collected and analyzed synchronized multi-modal data from subjective perspective (questionnaire) and objective perspective (biosignal uni-modal and multi-modal analyses). First, we briefly address the design and results of our initial in-lab user study, in which we investigated simulated attention distractions in an AR environment, [6]. Next, our second user study collected synchronized multimodal attention-related data in-practice (real biking sessions) using an AR environment. Finally, the field study investigated attention modeling in everyday traffic on the street, by studying real visual attention distractions during a postal delivery task. The experimental design of this work thus systematically moved from a controlled laboratory environment over an AR in-practice setup towards realistic annotation of real external visual attention distractions. The main contributions of this paper are: 1) The development of a robust setup for collecting multimodal attention-related synchronized data in cargo-cyclist scenario. 2) The design and validation of simulated internal and external attention distractions. 3) A successful multimodal Attention-load estimation and annotation of real distractions. 4) The paper addresses more general challenges in modeling cognitive processes from multimodal data in the wild [19, 20], demonstrating how combining different channels may lead to sufficiently robust real-world applications and contribute to improved safety of cyclists and pedestrians.

II. Related Works

Relevant works have studied multimodal data recording in cycling scenarios and related mobile recording situations, as well as attention modeling.

First, a few prior works have investigated situational awareness by recording multimodal data from cyclists. In the Multimer project, Ducau et al. [11] investigated the validity of multimodal biosignals data by recording EEG and accelerometer data from urban participants (including cyclists) in New York City. The Multimer app visualizes such multi-modal data to examine how humans cognitively and physically experience spatial environments. However, the authors did not discuss how participants' attention may be modeled from the recorded multimodal data. In a number of further studies, mobile data has been investigated in real world scenarios involving the assessment of mobile EEG. Andreas et al. explored body motion data to improve the interaction with eBikes [1]: While the authors did not investigate EEG data, they recently introduced Ena[2], an EEG-powered eBike, in which they investigated the rider's peripheral awareness form the recorded occipital-lobe EEG data to regulate engine control. Biking activity, however, can be assumed to strongly affect EEG data quality due to movement artifacts. E.g., Zink et al. [37] aimed to disentangle the influence of cognitive demands and distractions which arise from outdoor unconstrained recordings. The authors investigated the Event Related Potentials (ERP) and single trial characteristics of EEG data collected using a fixed indoor bike in comparison to EEG data from outdoors cycling sessions. While the authors found a decrease in P300 amplitude in outdoor biking compared to indoor fixed biking, they could not correlate the increase of movements to the differences in classification accuracy. Scanlon et al. [30] likewise compared the effects of indoor and outdoor biking on EEG-data, demonstrating that reliable ERPs could still be measured despite the increased noise and movement artifacts. The present work follows this approach to collect multimodal data in-lab (sitting, standing and walking) and outdoors during biking for further analysis of EEG data quality.

For attention modeling, prior works have typically aimed to distinguish between internal and external attention in various domains [7, 8, 16, 31, 35]. Vortmann et al. [34] investigated the discrimination of internal and external attention in Augmented-Reality (AR) scenario. The authors investigated different classifiers, features and window sizes, and they showed high accuracy by using a shallow neural net based on 4-second raw EEG data windows.

III. Data Collection, Analysis and Validation

We conducted user studies to collect data in controlled and uncontrolled conditions, leading up to real-world recordings on the road. In the description of the user studies, we are guided by the experiment model proposed in [27] for a description which allows replication of experiments. The data collections, for each of the present user studies, were approved by the ethics committee of the University of Bremen (Approval number: EK/2020/037-2). In each study, we explained the tasks and the setup for the participants, who then provided their informed written consent. All accidental data concerning non-participating individuals (e.g., bystanders, license plates, etc.) were automatically anonymized as defined by the respective ethics approvals.

A. In-lab Preparation User Study Using AR (Preparation Study)

This first user study, as detailed in [6], acted as a preparation in-lab phase for attention modeling. In this work, we only present a brief summary of this laboratory study, as our main focus will be on the subsequent studies outside of the laboratory. The main aim of this preliminary study was to investigate simulated internal and external attention distractions in an AR scenario. EEG Data from 16 dry electrodes was collected in the lab, wherein participants interacted with the AR scenario with simulated internal and external attention distractions using the HoloLens 2[3] device. We analyzed the EEG results from this starting phase [6], reduced the electrodes to only seven in attention-related positions, and designed a complete setup for collecting

[2]https://openbci.com/community/introducing-ena-the-eeg-powered-ebike/

[3]https://www.microsoft.com/en-us/hololens

synchronized multimodal attention-related data in the next user studies.

B. Simulated Attention Distractions (SIMULATION STUDY)

The aim of this study was to model cargo cyclists' attention under different simulated distractions using an AR application in-practice, i.e. in real delivery-like sessions: sitting, standing, walking and biking. To ensure a fully controlled environment with only the intended simulated distractions, the walking, sitting and standing sessions took place in the lab, while the biking sessions took place in the university parking spot. Participants experienced different kinds of simulated distractions in those delivery-like sessions. The simulated distractions include internal mental distractions (mental rotation task [12]) and external distractions (simulated visual and auditory stimuli). The participants were instructed to focus on the AR app and to avoid talking to the experimenter after the training phase when the experiment really starts. To minimize the learning factor of such simulated distractions from session to another, we randomised the order of sessions for each participant.

1) SIMULATION STUDY *Subjects:* Due to the Corona pandemic, we advertised the study only to our computer science students. Participants received monetary compensation for their participation (20 Euros). In total, 13 participants (students: 12 male and 1 female, mean age 23 ± 1.8) took part in this study. Only two participants were already familiar with AR glasses. None of the participants reported any history of brain disease or physical disabilities for cycling.

2) SIMULATION STUDY *Apparatus:* For collecting brain-activity EEG data, we used the g.Nautilus Research system with active EEG g.SAHARA electrode technology from g.Tec[4], a wireless, wearable EEG headset with dry-electrodes. It consists of the EEG cap and dry electrodes with cables connected to the EEG sender, which in turns sends the EEG signals wireless to the EEG-receiver connected to the recording laptop. Based on the EEG analysis in the PREPARATION STUDY [6] (in which we began with 16 electrodes in the 10-20 positioning system and investigated EEG-based attention prediction results from each electrode), we used the following 7 positions in the 10-20 positioning system for recording attention-related EEG data: Cz, C3, T7, P3, P4, PO7, PO8. Two electrodes were located at both earlobes as reference and ground. This study used Unity to implement the AR simulation, which was deployed on a HoloLens 2 device. The HoloLens 2 is the most recent version of AR glasses from Microsoft and is equipped with an optic mobile eye-tracker. The AR simulation continuously sends the eye-tracking data during the experiments. Additionally, it sends markers whenever a distraction occurs and subjects' answers to mental rotation tasks.

3) SIMULATION STUDY *Experimental Design:* At the beginning of the training phase before the actual experiment, the participants undertook a training phase in which they wore only the HoloLens 2, started the simulation application, and got familiar with the following simulated distractions:

- Internal, mental distraction: we aimed to simulate the cognitive processing of attention distractions via a mental rotational task [12]. The subject is shown two 3D objects in the AR glasses. The subject is then asked to answer with their voice (yes or no) whether the 3D objects are identical or not (See Figure 2). In order to answer, the subject has to mentally rotate the shown objects and judge. This manipulation aims to increase the cognitive load of participants to simulate a type of "mind wandering" effects as an example of internal attention distractions.
- External, visual distraction: traffic signs are shown to the subject at random positions on the AR glasses. The subject is instructed to not ignore the signs. Thus, when the subject follows the shown signs with their eyes, their visual attention is distracted by such an external stimulus.
- External, auditory distraction: Similar to the simulated external visual attention distractions, external auditory attention distractions are simulated via the HoloLens 2 by issuing everyday noise, e.g. barking dogs or ambulances. Hearing such external auditory stimulus is expected to distract the subject's auditory attention.

Experiment Setup: The experiment, with all the four sessions, lasts between 40 and 50 minutes per subject. For each session (sitting, standing, walking and biking) 10 blocks (each with a random order of simulated distractions) are shown to the participant, see Figure 1. Both the simulated

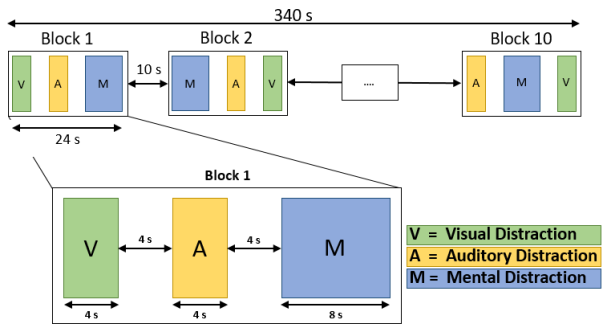

Fig. 1. Experiment Setup: 10 block with a random order of simulated distractions.

visual and auditory distractions last for 4 seconds, while the simulated mental distraction lasts for 8 seconds. Breaks (no distraction) last for 4 seconds between each two consecutive distractions within a block, see the zoomed Block 1 in Figure 1.

Experiment In-Practice: Figure 2 shows the design of SIMULATION STUDY experiments. The subject wears the g.Tec wearable headset with the 7 dry electrodes mentioned above and the HoloLens 2 device. Then, EEG and Eye-tracking calibrations are performed. Although the multi-moda data streams have different frequencies (EEG 250 Hz, Eye-tracking 40 Hz, video-frames 30 Hz), assigning a timestamp by a well-established centralized clock (Lab-Streaming-Layer LSL[5]) to each data measurement allows for synchronization.

4) SIMULATION STUDY *Questionnaire Analysis:* At the end of the experiment, participants completed an evaluation questionnaire. The distraction's validity statements to be examined were "How realistic were the simulated distractions:

[4]https://www.gtec.at/product/gnautilus-research/

[5]https://labstreaminglayer.readthedocs.io/info/intro.html

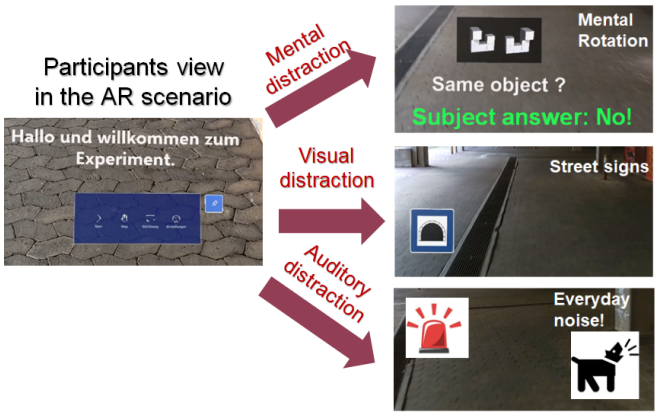

Fig. 2. SIMULATION STUDY experimental design. Left-side screenshot shows the greeting message shown to participants in the AR application just before starting. Right-side shows the different simulated attention distractions shown to participants in the AR application.

visual, mental and auditory". For each statement, participants could assign a score between 1 and 5, with 1 being the maximum approval (Likert-scale).
Table I shows that the participants were strongly convinced with the simulated mental (avg=1.7, std=0.8) and visual (avg=1.7, std=0.5) stimuli as realistic distractions (most responses were "1" or "2"), whereas the auditory stimulus was rated with a relatively low score (avg=3.1, std=1.3). A possible explanation could be the fact, that we did not provide a 3D audio support.

TABLE I
SIMULATION STUDY QUESTIONNAIRE RESULTS: MEAN AND STANDARD DEVIATION OF THE RESPONSES FROM ALL SUBJECTS.

Question	Mean(Std)
Realistic simulated visual distraction	1.7(0.5)
Realistic simulated mental distraction	1.7(0.8)
Realistic simulated auditory distraction	3.1(1.3)

5) SIMULATION STUDY *Objective Analysis:* In this section, we analyze the synchronized multimodal attention-related data (EEG and eye-tracking) which we collected in the SIMULATION STUDY. First, we discuss the EEG-based attention modeling, i.e the automatic detection of attention distractions from EEG data. Then, we show associations between the simulated distractions and the collected eye-tracking data, which in turn highlight advantages of multimodal attention modeling from both EEG and eye-tracking data.

EEG-based Attention Modeling: EEG data can be contaminated with noise and artifacts coming from different sources such as the type of power source, environment, eye blinks, heart rate, and muscle movements, [22]. We plotted and examined the raw EEG data from all the sessions (sitting, standing, walking and biking). Surprisingly, we did not observe major impairments due to the activity of biking in the EEG data quality, whereas walking seems to decrease the EEG quality by high amplitude outliers [17]. We used the well-established Independent Component Analysis (ICA) model in the ICLabel [26], which is integrated in the EEGLAB [9] tool to automatically classify none-brain data (artifacts) to be removed for improving the EEG data quality. Figure 3 illustrates the EEG quality before and after the ICA removal of eye artifacts.

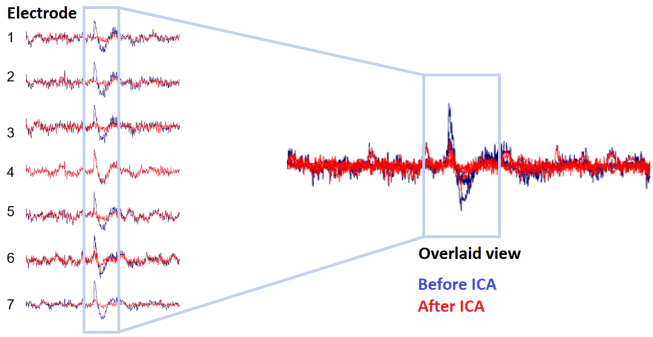

Fig. 3. EEG data before and after ICA artefacts removal: artefacts potentially coming from eye blinking have been removed.

We expect differences in brain activity depending on the experimental conditions (no distraction, distraction). Temporal dynamics exist between the EEG raw measurements [33]; thus, modeling such temporal dynamics in EEG data allows discriminating between different brain states. In the SmartHelm context, these brain states are attention and no attention. Researchers widely examined Long-Short Term Memory (LSTM [14]) to model temporal dependencies (dynamics) in EEG data for brain states classification in different domains, e.g. [13, 23, 36]. Therefore, we also examined LSTM models to automatically detect different attention distractions from raw EEG data. We segmented the EEG raw data in sequences of one-second bins (sampling-rate 250 Hz), aggregated the sequences from all the biking sessions from all subjects, and trained person-independent LSTM-based models which perform binary classification for different distractions (internal and external). Each LSTM-based model is a sequential model which consists of three layers: 1) LSTM input layer with 128 units (cells), 2) followed by a regularization layer (Dropout with rate=0.5) to avoid overfitting 3) and finally a fully connected dense layer with softmax activation acts as an output layer. To fit the model, we performed 500 epochs of Stochastic Gradient Descent using an early stop function with $patience = 50$ epochs. No further grid search optimization has been performed for the other hyper-parameters. The classification pipeline is implemented in Python. We used Keras [15] for machine learning and evaluation algorithms, and custom routines on Numpy and Matplotlib for data processing and visualization.

TABLE II
PRELIMINARY MACHINE LEARNING RESULTS: LSTM WITH RAW EEG DATA. EEG BEST SUITS THE DETECTION OF INTERNAL MENTAL DISTRACTIONS COMPARED TO EXTERNAL VISUAL AND AUDITORY DISTRACTIONS.

Binary Classifier	Before Down-Sampling		After Down-Sampling	
	Acc	F1-Score Dist.	Acc	F1-Score Dist.
None vs. Aud.	0.80	0.10	0.56	0.21
None vs. Vis.	0.78	0.12	0.64	0.23
None vs. Men.	0.61	**0.25**	0.59	**0.30**

We evaluated the LSTMs using leave-one-session-out cross-validation [5]. Due to the typical data imbalance in

attention-related data (major class is no distraction), we also down-sampled the data and only took sequences of measurements before and at the onsets of distractions. While the average accuracy is decreased as expected after down-sampling the data (Acc columns in Table II), the distraction detection performance improves because the models learn better from balanced data to detect more distractions (improved Recall, see the F1-scores in Table II). Table II shows that the EEG data, as expected [28], best suits the detection of internal mental attention distraction compared to external visual and auditory distractions, see the F1-scores. Although Figure 3 shows that the EEG quality has been improved by the ICA through deleting non-brain data from the raw EEG data, the LSTMs trained on ICA-processed data, do not show significant improvements compared to their counterparts trained on raw, unprocessed EEG data. A possible explanation for this is the strength of LSTMs for learning temporal dependencies in EEG data even if contaminated with "related" artifacts. E.g., eye-movements and EEG tend to correlate with sudden attention distractions [4].

Eye-tracking based Attention Modeling: according to the psychological literature, e.g. [10], a person will typically react with head and eye movements when their attention is distracted by an external stimulus. Thus, we expect eye-tracking features to perform well with respect to detecting the presence of external distractions (vs. no-distraction). In Figure 4, we plot the eye-tracking direction features (direction vectors which combine both head and eye movements) for one sample session, with the simulated attention distractions in 3D space: X, Y and Z. On the one hand, Figure 4 shows very clear associations between the eye-tracking direction data and the mental and visual distractions, i.e. we see eye-tracking peaks overlap with the distractions. On the other hand, no clear association can be seen between the eye-tracking data and the auditory distractions. This might be due to the subjects not responding with head and eye movements when they hear the mono (non 3D) audio during these distractions in our simulation. This weak association coincides well with the subjects' impressions analyzed in Section III-B4, where participants provided low ratings for the simulated auditory stimulus as a non-realistic distraction. Future simulation-based studies might be able to address this by employing 3D auditory stimuli.

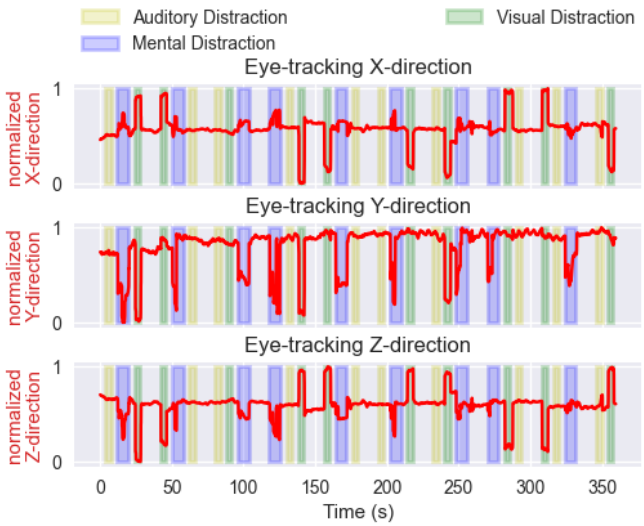

Fig. 4. Eye-Tracking sample session: Eye-tracking direction data (combination of head and eye movements) overlap with simulated attention distractions, especially with visual and mental distractions.

C. Real Field Attention Distractions (FIELD STUDY)

The aim of the FIELD STUDY was to investigate the modeling of real attention distractions which could happen during a short experimental delivery task. The task consisted of delivering three packages to three different addresses within a distance of about 2 km. The AR application shows the next delivery- and navigation steps, and allows to submit a delivered package. The AR application can be controlled completely "hands-free" via audio commands using an Automatic Speech Recognition (ASR) system which we trained for detecting wake words like "Hey Smarthelm" and application-specific voice commands, e.g., "Next navigation step!".

1) FIELD STUDY Subjects: Due to the corona pandemic, the study was only advertised to the project members' communities. 19 subjects participated in this study (16 Male and 3 Female, mean age 29.9 ± 9.8). Most of the participants stated that they had used VR or AR glasses before taking part in this study, so they had little problems getting used to the HoloLens 2. None of the participants reported any history of brain disease or physical disabilities for cycling.

2) FIELD STUDY Experimental Design and Apparatus: The experimental design of the FIELD STUDY is similar to that described in the SIMULATION STUDY but with a different AR application designed to support the postal delivery process.

AR postal delivery application: While we only simulated attention distractions in an AR application in the SIMULATION STUDY, this study aimed to annotate real distractions. We therefore used a professional AR postal delivery application[6] to support the participant with the postal delivery process and navigation steps.

Experiment In-Practice: Besides the AR delivery application, this study further employed an industrial route management system, as regularly employed by our commercial cargo-bike partner[7] to schedule tasks for delivery workers. Whenever a postal delivery task is scheduled and activated, the corresponding worker can see their task when starting the AR application. Thus, the experimenter schedules a simple postal delivery task in a route with 3 addresses for delivering three packages (Route is ca. 2 Km). The participant does not actually deliver any package, but instead drives to the corresponding addresses and uses the AR application as if they would carry out the delivery. The participant exercises the AR App at the office before the start of the main experiment. The participant is instructed to focus on their task and to avoid talking to the experimenter after the training phase when the experiment really starts. The participant wears the same devices described in the SIMULATION STUDY, with EEG and Eye-tracking being calibrated with the same

[6]This application was implemented by our project partner TeamViewer: https://www.teamviewer.com/de/unternehmen/

[7]RYTLE: https://www-staging.rytlecloud.com/de/login

procedure and settings. Finally, the experimenter starts the video recording of HoloLens 2 and the participant starts the task with the voice command "Hello SmartHelm, start the tour!".

3) FIELD STUDY *Questionnaire Analysis:* We prepared a questionnaire with Likert-scales and free text questions to check for subjective impressions from our participants. 8 from the 19 participants answered this questionnaire.
None of the participants reported perceiving the SmartHelm components as too heavy. There were also no overheating issues with the setup. However, the pressure of the dry electrodes combined with the HoloLens 2 was described as unpleasant by all except one participant. Four participants also expressed concern about the AR projections being distracting.
Attention: The participants were also asked questions about how much they got distracted during the study by either external visual or auditory distractions or their own thoughts (internal distractions). Overall, all participants reported not having been greatly distracted in general. This was expected because the delivery route and traffic conditions were not particularly challenging. Likewise in line with expectations, mental and visual distractions appeared to be perceived as most important. 5 participants stated they would get the most distracted by internal thoughts, 2 by visual distractions whereas no participant reported any substantial auditory distractions while performing this task.
Usability: The participants were not sure whether the SmartHelm could improve their safety on the road, with nearly everyone being neutral or slightly denying this statement. Consequently of the poor wearing comfort, most participants also would not want to wear the SmartHelm to work. Final questions asked about the driving experience of participants while wearing the SmartHelm components, as well as about wishes and possible design improvements (open answer-format). 3 participants found it interesting, where the main part that interested them about the SmartHelm was the technology. 3 other Participants found it 3 just "ok", while 2 participants abstained voting for this usability question. The main improvements requested by participants were an improved wearing comfort and a less crowded field of view (due to Holograms). Some subjects also stated that shorter and more distinct voice commands would be good.

4) FIELD STUDY *Objective Analysis:* Since there is no well-established ground-truth for attention in delivery tasks, we estimated a study-specific ground-truth for the FIELD STUDY attention data. Therefore, we analyze the synchronized multimodal attention-related data: video recordings and eye-tracking. We estimate the attention load according to video scenes and eye-tracking, and we annotate potential attention distractions from eye-tracking data and video scenes. Finally, we evaluate the annotated potential distractions using the estimated attention load of the corresponding video scenes and eye-tracking.
Traffic labeling: The first step in our analysis was a complete labeling for every video, using Yolov5[8] to automatically mark every object commonly found in traffic with a bounding box. This automatic labeling still has a few problems and cannot be used for our analysis without filtering some bounding boxes. E.g. subjects can look at a car driving in front of them as part of being completely attentive and focusing on the traffic, not distracted. We therefore implemented additional filters to classify real potential distractions (see below). Figure 5 shows a scene example in which we see all the automatically labeled objects using Yolov5, and our filters for focus area and candidate distractions border area.
Focus and Candidate Distraction Filters: We define two areas of interest in the video screen. The screen-middle rectangle represents the focus area, and the remaining surrounding border is considered as a candidate distractions area, see Figure 5. Thus, we filter the Yolov5 bounding boxes and only consider the border objects as candidate distractions: all objects that stay long (more than 2 seconds) in that outer area will count as candidate distractions. To be able to check this and track every object in the video, we used the Yolov5 and Deep Sort-based object tracker [3].

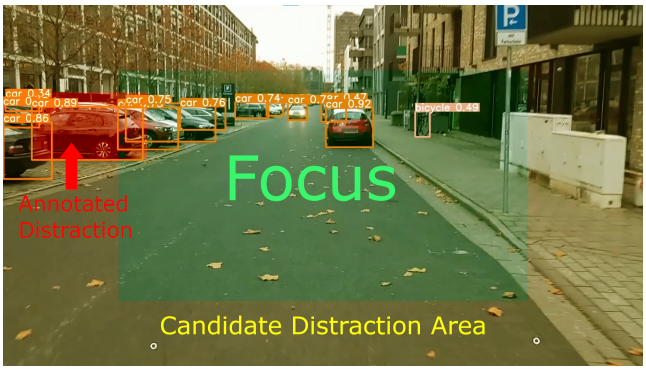

Fig. 5. Focus and candidate distractions areas in a video screen. A parked car at the left side has been automatically annotated as a distraction by our method. The annotated distraction looks very plausible as the participant had been looking at it for a long time (more than 2 seconds) although a moving car in the focus area had been driving towards the participant.

Annotation of Potential Real Distractions: In order to combine those candidate distractions with our ET data to detect distractions, we check for time-interval overlapping between ET and candidate objects rather than pixel-based overlapping, because of technical limitations of Hololens-2 ET, which has been designed for tracking holograms in 3D AR applications rather than tracking real physical objects in videos [2]. For the time-interval overlapping, we filter our ET data using a high-pass filter. This helps us to achieve a uniform baseline for our ET data, as well as peaks and negative peaks whenever there is considerable eye movement or some type of head movement. With this, if we see a substantial peak (exceeding a certain height threshold) during the time interval of a candidate distraction, we can assume that this candidate distraction is a real distraction. This annotation method applies for the x-direction data if the distracted participant looked long at a candidate distraction in the left or right sides of the screen. Similarly, looking long at candidate distractions on the bottom or upper sides of the screen will be annotated as a distraction.
Attention Load Evaluation: We estimate the attention load during biking from the multimodal attention-related data (ET and video recordings). Thus, we calculate, normalize and

[8]https://github.com/ultralytics/yolov5

plot ET and video attention load scores during biking. We analyze windows of five seconds, shift frame by frame, and calculate the attention load scores for the elapsed five seconds including the current frame.

ET Attention-Load Score: This score is computed based on the direction ET data with an applied high-pass filter. For each video frame, we accumulate every ET value in the last five seconds, so that time-frames with more frequent or stronger eye/head movements result in higher ET attention-load scores.

Video Attention-Load Score: This score is calculated for each frame as the average number of objects detected during the last five seconds including the current frame. Thus, a higher number of objects in the current time-frames results in a higher video score.

Figure 6 shows both attention-load scores together with the candidate attention distractions and annotated distractions in a sample session. Figure 6 shows that the candidate and annotated attention distractions come mostly, as expected, in time-frames with high estimated attention-load. We looked at all the sessions and found that 78% of the candidate distractions and 88% of the annotated distractions come in time-frames with high $VideoAttentionLoad > Mean$.

Moreover, Heat-maps on aggregated data from all the

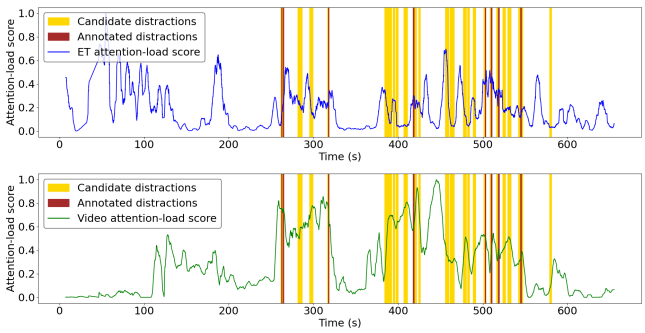

Fig. 6. ET and video Attention-load scores in window frames of 5 seconds length. Candidate and annotated distractions come mostly in time-frames with high estimated attention-load.

sessions highlight clear associations between the annotated distractions and high scores of video attention load (hot spots), see Figure 7.

Mutimodal Data Videos: We integrated the synchronized multimodal-data streams in the session's video, as depicted in Figure 8. In such videos,[9] synchronized peaks in the EEG alpha band and eye-tracking data are clearly observable, which may indicate a potential distraction. The video further shows the annotated distractions, based on which the plausible context of the annotated distraction can be described (see Figure 5).

IV. Conclusions

This paper presented the first evaluation results of the SmartHelm assistance system for cargo cyclists. Based on our formative design, the SmartHelm has evolved from the first laboratory tests to the current field-test prototype. We evaluated the validity of our simulated internal (mental)

[9]https://owncloud.csl.uni-bremen.de/s/yHyn9LgDbAWGW84

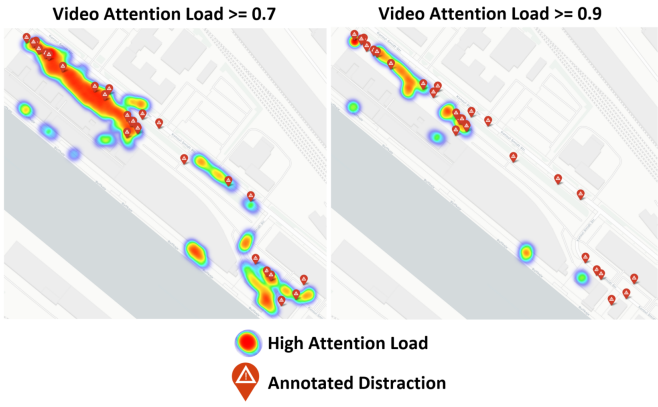

Fig. 7. Video attention load heat-maps with the annotated distractions on data aggregated from all the sessions. Annotated distractions associate to video attention load hot spots.

Fig. 8. Screenshot of a multimodal data video: we show all the synchronized multimodal data on the session video recording with yolo-based annotated objects. The video shows EEG raw data (right, top panel), EEG alpha power band (right, middle panel), eye-tracking direction data (right, bottom panel) and GPS positions with attention profile (left bottom, blue color shows the current position, green indicates no distraction, while yellow and red indicate estimated medium and high distractions respectively according to the synchronized data peaks).

and external (visual) attention distractions in the Simulation Study. We showed that LSTMs can learn temporal dependencies from the raw EEG data to detect different attention distractions. In the Field Study, we transferred our experience in this context to collect multimodal synchronized attention-related data: EEG, ET, and video recordings. Despite the limited capabilities of the Hololens-2 ET for tracking objects, we estimated candidate distractions and annotated real distractions from the ET and video recordings data. We compared those candidates and annotated distractions with estimated ET and video attention load scores and found, as expected, that attention distractions appear predominantly in high attention load time-frames.

We envision that novel assistive technologies such as SmartHelm will soon become a more common sight on our streets. Towards this aim, certain technical challenges, e.g., improved user-comfort, will still need to be addressed. While our current development aims specifically at cargo delivery cyclists, we believe that such systems are both viable and valuable contributions to the safety and well-being of a wide range of users. Systems like SmartHelm could therefore help take an important step towards attention-aware interactive

systems for empowering pedestrians and cyclists in everyday life.

REFERENCES

[1] Josh Andres, Julian De Hoog, Jürg Von Känel, Julian Berk, Bach Le, Xizi Wang, Marcus Brazil, and Florian'Floyd' Mueller. Exploring human: Ebike interaction to support rider autonomy. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pages 85–92, 2016.

[2] Samantha D Aziz and Oleg V Komogortsev. An assessment of the eye tracking signal quality captured in the hololens 2. *arXiv preprint arXiv:2111.07209*, 2021.

[3] Mikel Broström. Real-time multi-object tracker using yolov5 and deep sort. https://github.com/mikel-brostrom/Yolov5$_DeepSort_Pytorch$, 2020.

[4] Anne-Marie Brouwer, Maarten A Hogervorst, Bob Oudejans, Anthony J Ries, and Jonathan Touryan. Eeg and eye tracking signatures of target encoding during structured visual search. *Frontiers in human neuroscience*, 11:264, 2017.

[5] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107, 2010. Publisher: JMLR. org.

[6] Sromona Chatterjee, Kevin Scheck, Dennis Küster, Felix Putze, Harish Moturu, Johannes Schering, Jorge Marx Gómez, and Tanja Schultz. Smarthelm: Towards multimodal detection of attention in an outdoor augmented reality biking scenario. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 426–432, 2020.

[7] Marvin M. Chun, Julie D. Golomb, and Nicholas B. Turk-Browne. A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, 62(1):73–101, 2011.

[8] Désirée D Deiters, Stephan Stevens, Christiane Hermann, and Alexander L Gerlach. Internal and external attention in speech anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 44(2):143–149, 2013.

[9] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[10] Anup Doshi and Mohan M Trivedi. Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of vision*, 12(2):9–9, 2012.

[11] Arlene Ducao, Ilias Koen, and Zhiqi Guo. Multimer: validating multimodal, cognitive data in the city: towards a model of how the urban environment influences streetscape users. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, pages 1–8, 2018.

[12] Giorgio Ganis and Rogier Kievit. A new set of three-dimensional shapes for investigating mental rotation processes: validation data and stimulus set. *Journal of Open Psychology Data Files in this Item Files Size Format View 13-116-2-PB. pdf 567.5 Kb PDF View/Open*, 2015.

[13] Ryan G Hefron, Brett J Borghetti, James C Christensen, and Christine M Schubert Kabban. Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recognition Letters*, 94:96–104, 2017.

[14] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[15] Nikhil Ketkar. Introduction to keras. In *Deep learning with Python*, pages 97–111. Springer, 2017.

[16] Anastasia Kiyonaga and Tobias Egner. Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic bulletin & review*, 20(2):228–242, 2013.

[17] Julia E Kline, Helen J Huang, Kristine L Snyder, and Daniel P Ferris. Isolating gait-related movement artifacts in electroencephalography during human walking. *Journal of neural engineering*, 12(4):046022, 2015.

[18] Josef F. Krems and Martin R. K. Baumann. Driving and situation awareness: A cognitive model of memory-update processes. In Masaaki Kurosu, editor, *Human Centered Design*, volume 5619, pages 986–994. Springer Berlin Heidelberg, 2009. Series Title: Lecture Notes in Computer Science.

[19] Dennis Küster, Felix Putze, Patrícia Alves-Oliveira, Maike Paetzel, and Tanja Schultz. Modeling socio-emotional and cognitive processes from multimodal data in the wild. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 883–885. Association for Computing Machinery.

[20] Dennis Küster, Felix Putze, David St-Onge, Pascal E. Fortin, Nerea Urrestilla, and Tanja Schultz. 3rd workshop on modeling socio-emotional and cognitive processes from multimodal data in the wild. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, ICMI '21, pages 860–861. Association for Computing Machinery.

[21] Dennis Küster, Johannes Schering, Christian Janßen, Felix Putze, Jorge Marx Gómez, and Tanja Schultz. Intelligente und aufmerksamkeitssensitive Systeme in der Fahrradmobilität. In Alexandra Appel, Joachim Scheiner, and Mathias Wilde, editors, *Mobilität, Erreichbarkeit, Raum: (Selbst-)kritische Perspektiven aus Wissenschaft und Praxis*, Studien zur Mobilitäts- und Verkehrsforschung, pages 143–158. Springer Fachmedien.

[22] Chi Qin Lai, Haidi Ibrahim, Mohd Zaid Abdullah, Jafri Malin Abdullah, Shahrel Azmin Suandi, and Azlinda Azman. Artifacts and noise removal for electroencephalogram (eeg): A literature review. In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 326–332. IEEE, 2018.

[23] Yang Li, Zuyi Yu, Yang Chen, Chunfeng Yang, Yue Li, X Allen Li, and Baosheng Li. Automatic seizure detection using fully convolutional nested lstm. *International journal of neural systems*, 30(04):2050019, 2020.

[24] Yulan Liang, William J. Horrey, and Joshua D. Hoffman. Reading text while driving: Understanding drivers' strategic and tactical adaptation to distraction. 57(2):347–359, 2015.

[25] Anthony Mendil, Mazen Salous, and Felix Putze. Behaviour-based detection of transient visual interaction obstacles with convolutional neural networks and cognitive user simulation. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3348–3355, 2021.

[26] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.

[27] Felix Putze, Susanne Putze, Merle Sagehorn, Christopher Micek, and Erin T Solovey. Understanding hci practices and challenges of experiment reporting with brain signals: Towards reproducibility and reuse. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 29(4):1–43, 2022.

[28] Felix Putze and Mazen Salous. Multimodal differentiation of obstacles in repeated adaptive human-computer interactions. In *26th International Conference on Intelligent User Interfaces*, pages 260–269, 2021.

[29] Mazen Salous, Felix Putze, Markus Ihrig, and Tanja Schultz. Visual and memory-based hci obstacles: Behaviour-based detection and user interface adaptations analysis. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 1664–1671, 2019.

[30] Joanna EM Scanlon, Kimberley A Townsend, Danielle L Cormier, Jonathan WP Kuziek, and Kyle E Mathewson. Taking off the training wheels: Measuring auditory p3 during outdoor cycling using an active wet eeg system. *Brain research*, 1716:50–61, 2019.

[31] Hannah J Scheibner, Carsten Bogler, Tobias Gleich, John-Dylan Haynes, and Felix Bermpohl. Internal and external attention and the default mode network. *Neuroimage*, 148:381–389, 2017.

[32] Tanja Schultz, Christoph Amma, Dominic Heger, Felix Putze, and Michael Wand. Human-machine interfaces based on biosignale. *at - Automatisierungstechnik, 2013*, 61(11):760 – 769, 2013.

[33] Apoorva Sikka, Hamidreza Jamalabadi, Marina Krylova, Sarah Alizadeh, Johan N van der Meer, Lena Danyeli, Matthias Deliano, Petya Vicheva, Tim Hahn, Thomas Koenig, et al. Investigating the temporal dynamics of electroencephalogram (eeg) microstates using recurrent neural networks. *Human brain mapping*, 41(9):2334–2346, 2020.

[34] Lisa-Marie Vortmann and Felix Putze. Exploration of person-independent bcis for internal and external attention-detection in augmented reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–27, 2021.

[35] Lisa-Marie Vortmann, Moritz Schult, Mathias Benedek, Sonja Walcher, and Felix Putze. Real-time multimodal classification of internal and external attention. In *Adjunct of the 2019 International Conference on Multimodal Interaction*, pages 1–7, 2019.

[36] Chen Wei, Lan-lan Chen, Zhen-zhen Song, Xiao-guang Lou, and Dong-dong Li. Eeg-based emotion recognition using simple recurrent units network and ensemble learning. *Biomedical Signal Processing and Control*, 58:101756, 2020.

[37] Rob Zink, Borbála Hunyadi, Sabine Van Huffel, and Maarten De Vos. Mobile eeg on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of neural engineering*, 13(4):046017, 2016.