This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3333336

HazARdSnap: Gazed-based Augmentation Delivery for Safe Information Access while Cycling

Guanghan Zhao, Jason Orlosky, Joseph Gabbard, and Kiyoshi Kiyokawa

Fig. 1. Images showing (a) the HazARdSnap snapped to a pedestrian in the gaze path of the cyclist, (b) the computer vision algorithms we designed to detect pedestrians and potholes in real-time, (c) the UI fixed to the forward direction when no hazards are detected, (d) an example of placement in a more complex environment, in this case on a crowded sidewalk, and (e) a screenshot from an AR based user study in which participants avoided collisions with virtual hazards while reading virtual content.

Abstract—During cycling activities, cyclists often monitor a variety of information such as heart rate, distance, and navigation using a bike-mounted phone or cyclocomputer. In many cases, cyclists also ride on sidewalks or paths that contain pedestrians and other obstructions such as potholes, so monitoring information on a bike-mounted interface can slow the cyclist down or cause accidents and injury. In this paper, we present HazARdSnap, an augmented reality-based information delivery approach that improves the ease of access to cycling information and at the same time preserves the user's awareness of hazards. To do so, we implemented real-time outdoor hazard detection using a combination of computer vision and motion and position data from a head mounted display (HMD). We then developed an algorithm that snaps information to detected hazards when they are also viewed so that users can simultaneously view both rendered virtual cycling information and the real-world cues such as depth, position, time to hazard, and speed that are needed to assess and avoid hazards. Results from a study with 24 participants that made use of real-world cycling and virtual hazards showed that both HazARdSnap and forward-fixed augmented reality (AR) user interfaces (UIs) can effectively help cyclists access virtual information without having to look down, which resulted in fewer collisions (51% and 43% reduction compared to baseline, respectively) with virtual hazards.

Index Terms—Augmented Reality, Cycling, Safety, Eye Tracking, Object Detection and User Interfaces

1 Introduction

Cycling is a very popular form of transportation around the world, and it is environmentally friendly and generally safe. However, cyclists must often ride on sidewalks, narrow or rugged paths, or situations in which obstacles can be crash hazards. To stay safe, riders must pay attention to what is in front of them without looking down and reduce their speed to avoid collisions when necessary [4, 38]. Maintaining visual attention to the forward view can also be challenging when cyclists are distracted by navigation information or smartphone messages [31]. To complicate the matter, many countries, including Japan and various states the US, allow or partially allow cycling on sidewalks according to cycling regulations [13, 14, 18]. In addition, both professional and recreational cyclists typically make use of a cyclocomputer or phone that is usually mounted to a bicycle's stem, handlebars, or a forward-projecting extension to access biometric or map data. Although safety preservation and head-up information delivery approaches for driver assistance have been well-studied in the automotive domain, assistance for cyclists has not been well addressed. To alleviate the dangers of looking down at a smartphone, cyclocomputer, or other digital cycling interface for everyday and competitive cycling alike, we propose an information delivery approach for safer and more efficient cycling that combines eye tracking with adaptive augmented visualization. Our primary objective in this work is to improve and test this type of information presentation.

AR systems have increasingly been implemented with HMDs to augment user vision and provide more accessible digital content, such as the system by Chatterjee et al. [2] and Matviienko et al. [23]. These studies explored the impact of AR methods on attention and hazard perception during driving and cycling, suggesting that future research should delve further into counterbalancing information access and environmental awareness. Head-worn technology is an excellent candidate for cycling since riders already wear helmets while riding, and HMDs can easily be integrated into helmets. In addition, bike-mounted screens (e.g. smartphones or cyclocomputers) that function as cycling support systems require cyclists to look down frequently, potentially causing distractions and accidents [7, 34].

Optical see-through headworn AR displays have the potential to alleviate this problem since information can be overlaid in the user's natural cycling field of view (FoV). In order to develop an optical see-through interface for cyclists, the detection of, and interaction between, digital content and real-world hazards must be handled by the system in real time. For this purpose, machine learning and deep neural network (DNN) models are well developed for human and object detection, and they have already been widely applied in the fields of unmanned aerial vehicles, robotics, and autonomous driving, such as the implementation by Lan et al. [16] and Redmon et al. [30].

Although the performance of the obstacle detecting DNN models remains non-ideal in high-speed scenarios [8,32], they can be incredibly useful for bicycles, which travel at relatively low speeds compared to vehicles. In addition, AR approaches are widely applied in driving and

- *Guanghan Zhao is with Osaka University, Japan. E-mail: zhao.guanghan@lab.ime.cmc.osaka-u.ac.jp.*
- *Jason Orlosky is with Augusta University and Osaka University*
- *Joseph Gabbard is with Virginia Tech, United States.*
- *Kiyoshi Kiyokawa is with NAIST, Japan.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

hazard avoidance experiments since they can improve immersion while reducing danger to the participants [21, 25, 44].

In this paper, we introduce a novel approach called "HazARdSnap" that not only improves accessibility of bike-related UIs but also preserves awareness of hazards (e.g. pedestrians, cyclists and potholes). Several images outlining HazARdSnap are shown in Fig. 1. First, to determine the location and size of potential hazards, we gather scene data with an RGB-D camera and process scenery in front of the bike through a set of computer vision based algorithms. This data is sent to a HoloLens 2 via TCP/IP socket communication. Then, scene data, motion, gaze, and head position are gathered from the HMD.

Using this data as input, we designed an algorithm that detects hazards and snaps any relevant UI information onto those hazards when they become gaze targets. This allows simultaneous visual access to the information and awareness of the hazard's location and speed. Simply put, we provide better information access to cyclists without compromising their perception of forward-located hazards. In an experiment using real bicycles with virtual hazards (for participant safety), we evaluated HazARdSnap, smartphone-based presentation, and a forward-fixed UI delivered by an HMD. Our contributions in this paper are summarized as follows:

- The design, testing, and refinement of a UI snapping technique, HazARdSnap, designed to decrease the chance of collisions when accessing digital information during cycling.
- A bike-mounted system that makes use of computer vision, depth information, and environment tracking to detect oncoming hazards such as potholes and pedestrians.
- An AR based user study that compares cyclist performance while using HazARdSnap, a forward-fixed UI, and a bike-mounted smartphone UI as well as analysis and discussion of metrics like collisions with virtual objects, time spent viewing the road, and subjective ratings.

In the following sections, we discuss related work, detail the design process and implementation of HazARdSnap, describe our user study, and provide an analysis of the results. We conclude with a discussion of our findings and observations.

2 Related Work

When in motion, cyclists can often be distracted when checking their smartphones or cyclocomputers for navigation instructions, biometric information, or messages, especially when riding on low-quality surfaces or passing pedestrians or other cyclists. To help address these problems, many studies have been conducted for cycling and driving support. In this section, we discuss previously proposed approaches for improving both safety and information accessibility for cyclists and drivers.

2.1 Information Delivery

Cycling through narrow tracks or terrain with obstacles often requires high attention to the forward direction, frequent steering and low speed [38]. These challenges are especially pronounced for people who lack spatial awareness, such as children [39]. Moreover, Van der Horst et al. [36] demonstrated that cyclists' directional control behavior is usually influenced by pedestrians, particularly on shared bike paths, highlighting potential conflicts and safety concerns in such scenarios. On the other hand, using supporting functions (i.e. navigation) in mobile phones or other displays causes distractions [31] for cyclists and drivers. To solve this problem, Dancu et al. [6] evaluated a projection-based interface featuring a head-up display (HUD) that displayed map information and concluded that the HUD was only subjectively considered as safer and easier to use. Further, AR displays were proposed as a viable solution: Ginters [11] developed an AR cycling support system that displays route information on AR glasses. However, his work mainly focused on computing route design rather than interactive information delivery. Sawitzky et al. [40] investigated the effectiveness of augmentation concepts such as warnings with a cycling simulation. Their results suggest that the augmented information should be provided in a way that is easily accessible in order to benefit cyclists, which further inspired our work. With further consideration of usability, Berge et al. [1] found that on-bike human-machine interfaces may fulfill a cyclist's need for visually receiving recognition information from automated vehicles, however their interview studies revealed that the cyclists were hesitant about using such devices for communicating with automated vehicles. Lopik et al. [37] demonstrated that handheld devices provide higher perceived usability, while in-sight monocular displays, such as AR glasses, enhance awareness of hazards by leading to a greater number of hazards being acknowledged during navigation tasks. This further underscores the potential of AR glasses as a device to enhance hazard perception.

According to the concept of "processing resources," when the combined demand of two tasks exceeds available attention resources, time-sharing efficiency decreases, especially as the difficulty of either task increases [43]. In this context, the two tasks are information access and spatial awareness. To better counterbalance the potential negative effects of AR-based information presentation that may draw excessive attention away from spatial awareness, suitable content length and delivery timing were studied by Uchiya and Futamura [35].The results of their experiment indicated that the appropriate presentation timing depended on the participants' usual modes of transportation. Regarding driving safety, HMD based information delivery systems were also considered as more competitive than classical head-up displays [17]. Park et al. [28] proposed an in-vehicle AR system that delivers driving-safety information within the driver's field of view and excels in recognizing obstacles under adverse weather conditions. Their research confirms the effectiveness of combining AR in-view safety information delivery with real-time hazard detection algorithms. As an evolutionary branch of AR, MR approaches can also be both interactive and immersive for driving support while preserving efficiency and safety, as demonstrated by Ghiurau et al. [10]. They extended the safety support system of Volvo cars to incorporate MR technology, thereby improving users' awareness of upcoming hazards. Additionally, Matsunaga et al. [22] developed an MR-based assistance system for welfare vehicle users and validated its effectiveness in collision avoidance. Leveraging the characteristics of MR, the system projects a virtual vehicle in the user's field of view, allowing them to control the real vehicle as if it were following the virtual one from behind.

2.2 Hazard Detection

Computer vision based AR methods were widely accepted as efficient and low-cost solutions for detecting pedestrians, thus they were applied to vehicles: Early work by Gavrila [9] applied a two-step machine learning algorithm on moving vehicles as a pedestrian avoiding support and suggested further improvements and integration to meet the performance requirements for practical deployment of driving safety systems. On the basis of Garvrila's study, Shashua et al. [33] developed a single-frame classification approach for day-time driving conditions. Their system showed satisfactory performance for some functions under specific conditions, but further improvements were needed for more challenging scenarios. Hariyono et al. [12] have improved the performance of moving vehicle pedestrian detection with a histogram of oriented gradients. The way that they take advantage of contours and histograms in these works inspired the development of our real-time hazard detection algorithms.

To date, DNN models have been applied as an advanced computer vision method for pedestrian detection on vehicles. Wang et al. [41] demonstrated a DNN based pedestrian detecting system with semantic information of body parts and contextual information. By leveraging body part semantic information and contextual data, it significantly reduced missed detections and improved localization accuracy, particularly for occluded pedestrians. Nataprawira et al. [24] evaluated the performance of a DNN model for pedestrian detection on an urban road during both day and night time conditions. Their results indicated that the model performed significantly worse in low-lighting conditions.

Reagrding the danger caused by low-quality road surfaces with obstacles, Labayrade et al. [15] implemented a robust computer vision based method for obstacle detection on non-flat roads that estimates the road's longitudinal profile, identifies objects above the road surface as poten-

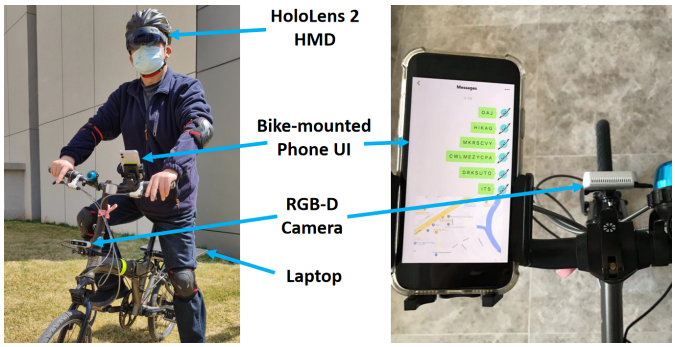

Fig. 2. Images of the bicycle setup, including the HMD, bike-mounted phone UI, RGB-D camera, and laptop from a front (left image) and top (right image) view.

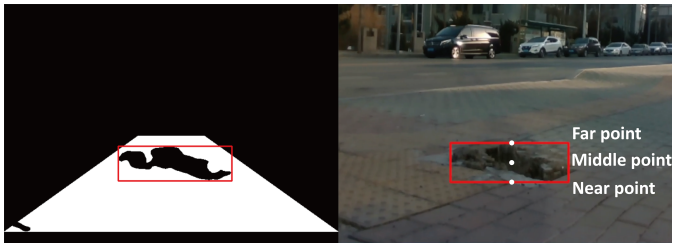

Fig. 3. Images showing the detection process for a pothole using binary thresholding of the RGB-D camera with contour detection (left) and the resulting detected pothole with the three points indicating the coordinates used for specifying the distance calculation and resulting distance-based pothole detection.

tial obstacles, and accurately detects road obstacles. Similarly, Rateke et al. [29] proposed a convolutional neural network based road quality classifier, aiming to benefit road maintenance departments and enhance road safety. However, this approach does not take hazard indication to the driver side into consideration. In line with the concerns raised by previous studies regarding the hazards posed by low-quality roads, we believe that pothole detection is also essential to our topic. Additionally, a smartphone based hazard warning system for pedestrians was proposed by Wang et al. [42]. In this work, they take advantage of the smartphone's back camera to detect vehicles with machine learning, however the performance can be easily affected by the smartphone's gesture. This issue highlights the importance of securely positioning the RGB-D camera in a position where it is minimally affected by cycling behavior in our research. Similar to the work by Wang et al., Li et al. [19] proposed a system for smartphones which focus more on detecting the pedestrian's walking behaviour by machine learning algorithms with video inputs from the front camera of the smartphone.

2.3 Further Motivation

Current approaches on AR based information delivery mostly use fixed UIs (i.e. screen-relative graphics) on HUDs or take advantage of auditory outputs. As UIs are fixed in the field of view, they may block the user's vision and require the gaze to move away from the road or obstacles. On the other hand, although the spatial audio can be delivered by bone conduction headsets, the user still has to process the sounds received and translate those sounds into directional cues, which may be difficult to process or correctly map to a hazard's location. These current limitations further motivated us to come up with a method that allows for simultaneous perception of information and road hazards.

When DNN models and other hazard detection algorithms are typically used, it is often for the safety support of motor vehicles. However, cycling presents unique challenges, as it requires additional motor control, involves more frequent tilts and turns, and often involves close proximity to other cyclists or pedestrians. Therefore, there is a need to adapt and evaluate the performance of DNN models and other hazard detection algorithms specifically for cycling scenarios. This issue highlights the importance of developing a method that allows for simultaneous perception of information and road hazards during cycling.

3 Design and Implementation

In order to detect potential hazards in front of the bike and place the interface such that it smoothly snaps to each hazard, we developed two subsystems: a *visual recognition subsystem* programmed with Python and an *interface delivery subsystem* programmed with C#. The visual recognition subsystem was implemented with an RGB-D camera and a backseat attached laptop, as shown on the left of Fig. 2. The interface delivery subsystem was implemented using an AR HMD. These two subsystems exchange data using asynchronous TCP/IP socket communication through a smartphone WiFi hotspot. Algorithms in these two subsystems run in separate threads for efficient parallel processing.

3.1 Equipment

In the implementation (Fig. 2), we applied a Microsoft HoloLens 2 HMD, an Alienware m15 Laptop, an Intel Realsense D435i RGB-D camera, an iPhone 11 smartphone and a Dahon Jetstream P5 bicycle. The algorithms were driven by Unity 2021.1.20f1 and Python 3.9. The HMD's FoV is 43°×29°. The RGB-D camera has minimum 28cm depth distance, frame rate of 30 frames per second (fps), and 69°×42° FoV.

3.2 Visual Recognition Subsystem

To start, we capture both the RGB and depth frames using an RGB-D camera fixed to the front of the bicycle, which faces directly forward with respect to the bike frame. Specifically, the RGB-D camera is fixed to the main frame of the bicycle rather than the handlebar in order to maintain a forward-looking pose independent of small movements in the handlebar. We align these captured frames with the camera's application program interface and then apply two algorithms: a Caffe-based MobileNet-SSD detection network DNN model [5] for human detection and a customized algorithm for pothole detection.

In addition, in the pothole algorithm we first calculate the peak gray value of the ground area using a histogram and then apply a binary threshold with a value of the peak value minus 30. Contours of probable potholes are gathered from the binary threshold results using a contour finder. To filter non-pothole results such as dark stains or manholes, we further compare the average depth of the far point and near point with the depth of the middle point (Fig. 3). If the difference is higher than 0.2 meters, the contour will be passed to the next step. With the gathered regions of interest (ROIs) and their vertex coordinates from both algorithms, we calculate the center pixel coordinate and depth of each ROI. In addition, the available depth range is 0.3 meters to 12 meters due to the RGB-D camera, but this covers a wide range of possible hazards and stopping distances (roughly 5 seconds of stopping time at 20km/h at 12 meters). After gathering vertex coordinates and center depth, world coordinates (x, y, z) of each vertex are defined as:

$$z = centerdepth \tag{1}$$

$$x = (u - cx) * z/fx \tag{2}$$

$$y = -(v - cy) * z/fy \tag{3}$$

where the (u, v) is the pixel coordinate of each vertex, and the cx, cy, fx, fy are from the intrinsic matrix of the RGB-D camera. To convert the v value on the top down v axis in pixel coordinates to the y axis in projection coordinates, we calculate $-(v - cy)$. This set of equations are originally from the 2D-3D coordinate transformation models of SLAM. Since we only focus on the front view and the camera is always considered to be the origin in this study, the rotation and transformation matrices are not applied.

Once the world coordinates of the region of interest are defined, we pass them through an asynchronous TCP/IP socket running in another thread. In this socket, we set the subsystem to be the host and the sending interval to 0.05 seconds in order to mitigate jitter and packet sticking.

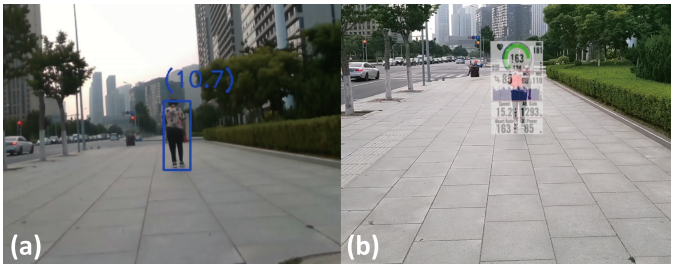

Fig. 4. Images showing (a) the detection result from the visual recognition system with the depth value marked and (b) the UI snapped to an in-focus pedestrian (the red dot is a gaze indicator).

3.3 Interface Delivery Subsystem

This subsystem is developed in Unity and deployed to an HMD to afford an AR environment (Fig. 4). The data from the visual recognition subsystem is received via Socket client as messages and are cut based on a header in order to further avoid packets getting stuck. To compute the intersection of the HoloLens Gaze vector with hazards, transparent quads are generated in the AR environment based on the received world coordinates of detected pedestrians and potholes. An 8-meter long, 0.7-meter wide (corresponding to the bicycle width) transparent bicycle heading indicator is implemented in order to detect obstructions that stand in the way within a range of 4-12 meters.

We have also developed a method to compensate for the drift phenomenon that occurs in gyroscopes and other direction errors that result from the tilt motions that happen often while cycling. In our implementation, the bicycle's heading direction is calculated by taking the HMD's velocity vector in the AR environment, which is further corrected using the face-forward direction of the head with a 5-degree difference range. In simple terms, the heading direction will shift only when the HMD is moving in a similar direction to the face-forward direction of the head. With this justified heading direction, we are able to rotate the quads and heading indicator such that they are forward-fixed. The height of the quads are set relative to the HMD height plus 0.3 meters in order to match the hazards' locations in-view.

3.4 Gaze-contingent Rendering

Gaze data is collected from the built-in eye tracker of the HMD, which allows the interface to be snapped to the quad of the hazard that the user is focusing on. In particular, as the gaze ray comes close enough to a detected hazard (less than a 5-degree difference on the Y-axis), the UI's position will be updated to the hazard's position, providing a frame-to-frame transition from the current UI position to the next as the user's gaze follows the hazard. Considering that erratic or instantaneous motion of the UI is likely to cause confusion or distract the cyclist, a smoothing function is applied. In particular, the UI is set to move toward the hazard with a speed of 7m/s, which provides smooth and more consistent movement. To prevent variations in the scale of the text of the interface, which can degrade readability [3], the scale of the interface is set to change relative to the distance to the gaze-hit position in order to keep the same scale in the view. In other words, the text will appear to remain the same size, but binocular cues will still give the cyclist proper depth judgement. With further logical testing, we defined the scale relationship as:

$$w = d * 0.12 \tag{4}$$

$$h = w * 1.7778 \tag{5}$$

where the d is the distance between the gaze-hit position and (w,h) are the width and height of the interface. The w-d ratio is defined by testing the minimum scale for displaying our UI content with the HMD's resolution and a ratio of 0.1 (w : d) is found as the just readable value. In order to ensure that the UI was fully readable in a variety of scenarios (especially our experiment), we defined the ratio as 0.12. The aspect ratio was fixed to 16:9 (1.7778) in order to match the aspect ratio of smartphone applications. Regarding situations in which the user is not focusing on any hazard or there is no hazard in a forward location, we provide an interface that is fixed in front of and 20 degrees below the heading direction (below the horizon from the user's viewpoint) in order to prevent blocking the user's forward vision and preserve accessibility [26] at the same time.

4 System Evaluation

In this section, we describe two evaluations of the system's technical performance, including evaluations of detection accuracy and the system's process latency.

4.1 Detection Accuracy

To set up this test, members of our research team went cycling in three different environments for approximately 18 minutes in each one, recording data for the duration of the trip. Then we gathered the every 10th frame from the data and set up two frame pools that included a total of 2000 frames in each pool for testing human and pothole detection. Forty representative images taken from this dataset are shown in Fig. 15. From each frame pool, we picked every second frame (1000 for each condition) for testing in terms of true and false positives and negatives. Then, we manually labeled all of these frames as having or not having a human/pothole and compared the detection output with our manually labeled results. For human detection, results were 97.48% precision, 93% recall, and a 2.4% false-positive rate. For pothole detection, results were 96.27% precision, 82.6% recall, and 3.2% false-positive rate.

4.2 Process Latency

We next tested the end-to-end latency of the system by printing timestamps in each step of the system processes for 10 iterations and averaging the results. We found that the latency was on average 273.1ms from the detection of a hazard to a quad (i.e., the gaze-contingent bounding-box) is placed. The average latency of each step was: 82.6ms from detection until sending a socket message (with the 50ms sending interval), 186.5ms from sending a socket message until it is received by the display, and 4ms from receiving a socket message until a quad is placed.

5 User Study

Our primary design considerations for the HazARdSnap algorithm were to 1) ensure fewer collisions with virtual objects while reading UI content, 2) enhance the participant's ability to read the information on the interface, and 3) reduce distraction while improving the user's feeling of safety, as compared to other conditions.

In the experiment, we compared HazARdSnap to a smartphone UI, as shown on the right of Fig. 2, and a forward-fixed AR interface. Because of the common use of smartphones or cyclocomputers for information access during cycling, the smartphone UI condition in this experiment was taken as a baseline (control) condition. Participants rode a bicycle in an open space, while attending to the UIS and avoiding virtual hazards. In addition, participants were asked to go through an AR track with assigned tasks. We examined their primary task performance (i.e., cycling and reading performance), head and gaze behavior, and subjective preference with the three UI conditions.

5.1 Equipment

We used a Microsoft HoloLens 2 HMD, a Dahon Jetstream P5 bicycle, an iPhone 11 and a set of a cycling helmet with joint pads. The subsystems were separately developed and driven by Python 3.9 and Unity 2021.1.20f1.

5.2 Experimental Design

Although HazARdSnap is designed for real-world cycling scenarios, considering the potential risks to participants and pedestrians when navigating through real crowds while cycling, we aimed to create a testing environment that is both safe and closely mimics actual cycling

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3333336

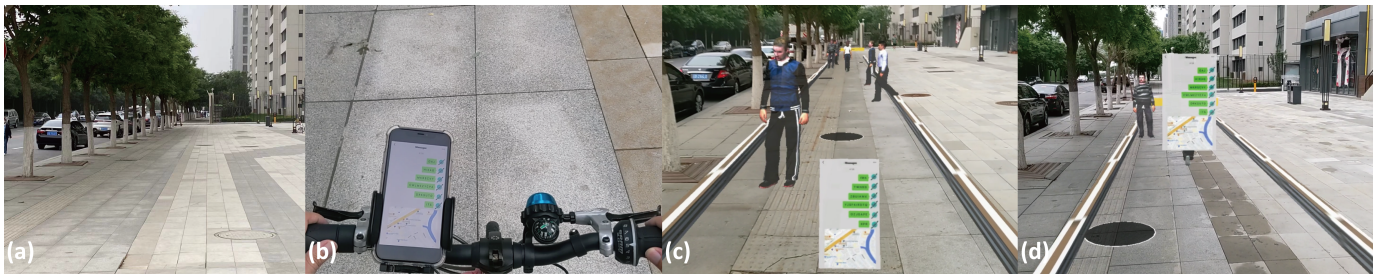

Fig. 5. Images showing (a) the sidewalk where we conducted the user study and the three conditions in the experiment, including (b) the smartphone UI, (c) the forward-fixed UI, and (d) HazARdSnap.

Fig. 6. Images showing (a) a view of the AR environment in the real world recorded through the HoloLens 2's streaming function and examples of a (b) virtual pedestrian and (c) virtual pothole from Unity.

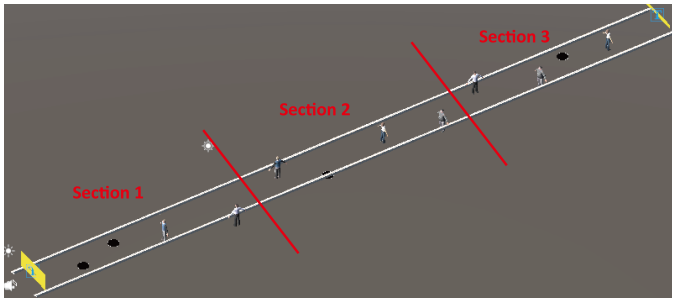

Fig. 7. Overview of the designed virtual environment in the Unity Inspector, scaled to 50 meters long and 2.5 meters wide. This was segmented into 3 sections with 4 objects in each section.

conditions. In order to preserve safety in bicycle related works, researchers have conducted experiments with virtual reality environments and fixed bicycles [20, 25]. However, this approach of user testing may cause motion sickness and could miss some motions triggered by losing body balance. To provide a better tradeoff, our experiment was implemented by letting participants actually ride a bicycle in an open and flat space with virtual pedestrians, potholes and a track delivered by the AR HMD. Since the user study was conducted as a directly forward scenario and no direction adjustment was needed, we did not apply the RGB-D camera and laptop in order to mitigate the load on the bicycle.

The user study used a repeated measures within-subjects design to test HazARdSnap (Fig. 5-d) against two other conditions: An interface delivered by a smartphone mounted on the left side of the handwheel (Fig. 5-b) and an interface which was fixed front-below the forward direction (same as the situation that the user is not focusing on any hazard or there is no hazard located forward of our proposed approach) (Fig. 5-c). The order of conditions was assigned based on a balanced Latin-Square in order to minimize order effects.

Considering the possibility that participants may remember the locations of the hazards and the content on the interface, six different experimental environments and six different content on the interface were randomly assigned as each trial start. For each trial, participants were required to ride forward while avoiding the virtual hazards and boundaries in the experimental pathway. Furthermore, participants were asked to read out the content presented via UI condition and complete the UI reading task before reaching the finish line. Detailed descriptions about the hazards and content can be found in section 4.3. Raw data (e.g. collisions, speed, focus, and distance) were recorded during each trial in order to help us to understand the accessibility of each method and how participants' riding behavior were affected.

5.3 Experimental Environment

To test HazARdSnap and the AR UI, we utilized a system that made use of virtual humans and potholes to ensure that participants would not be at risk of crashing into real hazards. Instead of detecting real objects, we adapted our method to react to the virtual objects within a Unity environment for the experiment. In the AR environment, we rendered 8 pedestrians, 4 potholes and 2 boundaries (one on each side) where the size of these objects were set according to real-world sizes (Fig. 6). In order to simulate real pedestrian behaviours, we programmed the scenario such that at least one pedestrian would move across the track repeatedly, one pedestrian across the entire track when approached by the bicycle, and one pedestrian who is already on the road but moves into the path of the bicycle when approached. Note that the version of HazARdSnap described above includes a smoothing function to move one object to the next. The initial version of the algorithm in the experiment quickly snapped the content from one object to another, which we improved upon later based on participant feedback.

In the virtual environment, the track was set to be 50 meters long and 2.5 meters wide to increase the likelihood that participants would finish the UI reading task and the objects would not fully block the way. To alleviate ordering effects between trials while ensuring similar difficulty, we segmented the environment into three sections, and the arrangement of these sections were randomized, resulting in six variations of experimental environment (Fig. 7). Regarding the reading content, we randomly typed 35 letters for each UI instance (Fig. 8). Considering the performance of hologram displays can be highly affected by day light conditions, we conducted our user study in the same place and same time of day (5pm to 7pm) (Fig. 5-a).

5.4 Participants

24 naive individuals (9 female, mean age 32.29, SD 7.17, range 13-52), were recruited from the general population to take part in the study. All participants had normal or corrected-to-normal vision, little to no experience with AR and were able to ride bicycles. They were informed that they were allowed to quit the study at any time. We set the bicycle seat to a relatively low height and added provided protective gears (i.e. the helmet and joint pads as shown in Fig. 2) in order to protect the

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3333336

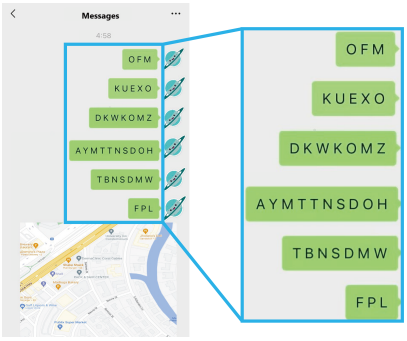

Fig. 8. An example of the content that participants had to read while cycling during the user study.

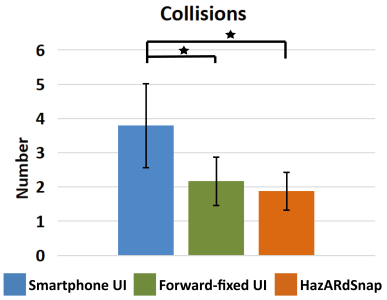

Fig. 9. Analysis of number of collisions showed that smartphone UI was associated with significantly more collisions than forward-fixed UI and HazARdSnap.

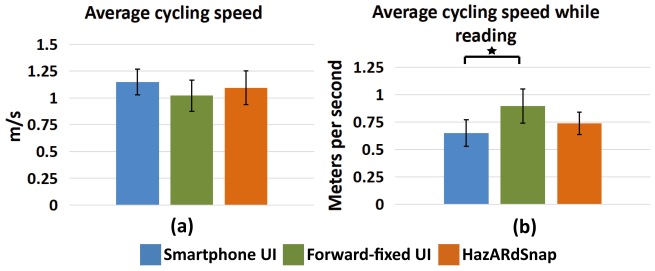

Fig. 10. Analysis of (a) speed and (b) average speed while reading. Although there was no significant difference within cycling speed, participants tend to ride faster while reading via forward-fixed UI.

participants in the unlikely event that they fell down. The experiment was conducted under approval of the Osaka University Cybermedia Center Institutional Review Board (IRB), number SA2022-02.

5.5 Dependent Measures

During the study, we logged the data necessary to calculate a comprehensive set of dependent measures as follows:

Number of collisions (i.e. hit an obstacle or ran out of boundaries): We collected collision data using a Unity collider that moves with the HMD, with its size set to match the actual size of the bicycle. This measure represents the times the participant collided with any hazard.

Average cycling speed in m/s: We divided the AR track length by completion time.

Average cycling speed while reading (i.e. while focusing on UI) in m/s: We took the average cycling speed when the participants were focusing on the interfaces.

Reading time (total time focused on the UI): We recorded the total time that the participants were focused on the interfaces. This measure was applied to test the amount of time users spent viewing the text with different methods.

Reading-to-completion time ratio: We divided the reading time by trial completion time. Considering that reading speed likely varies between individuals, we used this measure for further comparison of reading efficiency between methods.

Average angular speed of head rotations and *average angular speed of gaze* in °/s: We averaged the differences in angles between frames, measured in degrees per second. These measures were applied to determine how much a participant's gaze (both eye and head) moved during cycling, which is revealing of the number of attention switches between tasks.

Average angle difference between head direction and face-forward (i.e., head deviation from face-forward) and *average angle differences between gaze direction and face-forward* (i.e., eye gaze deviation from face-forward) in degrees: we first collected the head facing direction and gaze direction data from the HMD and then calculated the average angle between the direction data with a forward vector (0,0,1). This measure tells us whether participants were focused on the forward direction and thus better able to view hazards.

Post-trial subjective ratings: participants were asked to rate the following subjective preference questions with 7-scale ratings from 1 ("Strongly disagree") to 7 ("Strongly agree"):

- I had a clear perception of the pedestrians' and potholes' positions while looking at the UI.
- I felt safe while riding.
- Reading content on the interface disturbed me.
- I could read the content easily and clearly.

The quantitative measures listed above were recorded with the HMD's built-in eye tracker and sensors. In order to estimate time of focus data for the smartphone UI condition, we recorded the time periods when participant gaze was pointing the left-down direction with a degree to forward more than 20 degrees.

5.6 Procedure

As each participant joined our experiment, we briefly introduced the experiment, and we informed the participants about the tasks and requirements. Then, the participant put on the protective gear and the HMD with help from an experimenter. The experiment included three trials in total, with each trial representing a single condition (either Smartphone UI, Forward-fixed UI, or HazARdSnap). Before the first trial started, the participants were asked to complete an eye calibration using the calibration tool bundled with the HoloLens 2. Before each trial was conducted, we carefully checked the clearance of the physical experimental space. After each trial, the experimental environment was reset and the participant was asked to fill in a questionnaire. This was repeated until all three conditions were tested.

5.7 Results

Here, we describe the results of our experiment with respect to the three UI conditions, differences on accessibility, and subjective preference of the participants. For parametric data, we applied ANOVA tests and Tukey's HSD post-hoc tests. On the other hand, Friedman tests and Wilcoxon post-hoc tests were applied for non-parametric data. In addition, η^2, Cohen's d, Kendall's W, and r values are included in our results to define effect sizes. In the figures, the heights of the bars represent the means, with error bars indicating standard deviations. Significance was determined at a significance level of 0.05, denoted by '$*$'.

5.7.1 Task Performance

For the cycling task, regarding number of collisions, ANOVA test revealed significance between the three conditions ($F(2,69) = 8.26$, $p < 0.001$, $\eta^2 = 0.19$) (Fig. 9). Further differences were found in the smartphone UI vs. the forward-fixed UI ($p < 0.01$, $d = 0.81$) and the smartphone UI vs. the HazARdSnap ($p < 0.001$, $d = 1.00$), indicating that participants were more likely to collide while reading the smartphone UI. Regarding average cycling speed, no significant difference was found ($F(2,69) = 1.20$, $p = 0.31$, $\eta^2 = 0.03$) (Fig. 10-

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3333336

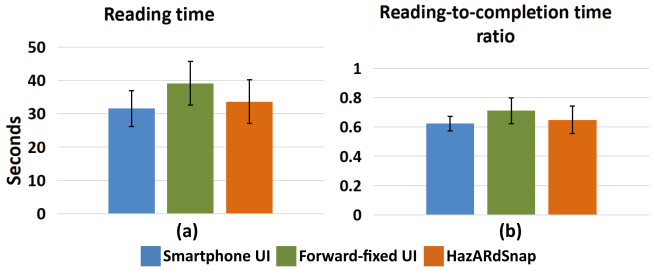

Fig. 11. Analysis of (a) reading time and (b) reading-clear time ratio. Regarding the UI reading task, these two metrics revealed that reading speed was not affected by conditions.

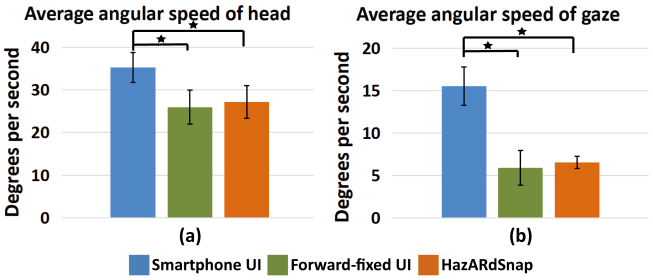

Fig. 12. Analysis of (a) head's average angular speed and (b) gaze's average angular speed. These showed that there are significant differences in the rotating frequency of both head and gaze.

a). Regarding average speed while reading, significance was found ($F(2,69) = 5.62$, $p < 0.01$, $\eta^2 = 0.14$) (Fig. 10-b) and post hoc tests show that there was a significant effect in the smartphone UI vs. the forward-fixed UI ($p < 0.01$, $d = 0.88$), which showing that participants rode faster with the forward-fixed UI.

For the UI reading task, regarding reading time, the data revealed no significance ($F(2,69) = 2.39$, $p = 0.10$, $\eta^2 = 0.06$) (Fig. 11-a). Regarding reading-clear time ratio, no significant difference appeared ($F(2,69) = 1.99$, $p = 0.15$, $\eta^2 = 0.05$) (Fig. 11-b).

5.7.2 Head and Gaze Behavior

To determine whether the augmented UI delivering methods affected participants' head and gaze rotation, ANOVA tests were applied to reveal that there were significant differences in both of the head's average angular speed data ($F(2,69) = 9.92$, $p < 0.001$, $\eta^2 = 0.24$) and the gaze's average angular speed data ($F(2,69) = 52.81$, $p < 0.001$, $\eta^2 = 0.60$). Further Tukey's HSD tests showed that for head's average angular speed (Fig. 12-a), there were significant differences in the smartphone UI vs. the forward-fixed UI ($p < 0.001$, $d = 1.24$) and the smartphone UI vs. the HazARdSnap ($p < 0.01$, $d = 1.10$) and for gaze's average angular speed (Fig. 12-b), there were significant differences in the smartphone UI vs. the forward-fixed UI ($p < 0.001$, $d = 2.23$) and the smartphone UI vs. the HazARdSnap ($p < 0.001$, $d = 2.67$), indicating participants, on average, moved their heads and gazes more frequently when using the smartphone UI as compared to both the Forward-fixed and HazARsnap UI. In addition, significant differences were also found in average angle of head facing direction to forward ($F(2,69) = 117$, $p < 0.001$, $\eta^2 = 0.77$) and average angle of gaze direction to forward ($F(2,69) = 91.64$, $p < 0.001$, $\eta^2 = 0.73$). For average angle of head facing direction to forward (Fig. 13-a), significant differences were located in the smartphone UI vs. the forward-fixed UI ($p < 0.001$, $d = 2.69$), the smartphone UI vs. the HazARdSnap ($p < 0.001$, $d = 3.64$) and the forward-fixed UI vs. the HazARdSnap ($p < 0.001$, $d = 2.17$). For average angle of gaze direction to forward (Fig. 13-b), significant differences were located in the smartphone UI vs. the forward-fixed UI ($p < 0.001$, $d = 2.50$), the smartphone UI vs. the HazARdSnap ($p < 0.001$, $d = 3.59$) and the forward-fixed UI vs. the HazARdSnap ($p < 0.01$, $d = 1.22$). These results revealed that participants averagely focused more on the forward direction when using the HazARdSnap.

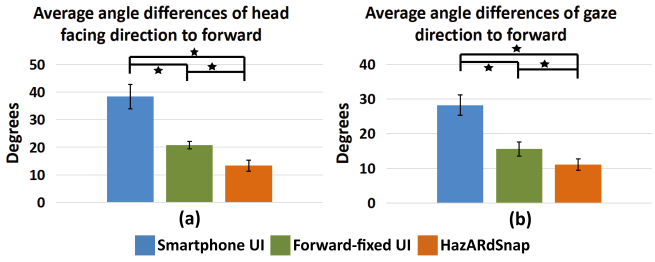

Fig. 13. Analysis of (a) the average angle difference between head direction and face-forward and (b) average angle differences between gaze direction and face-forward. These two metrics roughly represent the average time spent viewing the road.

5.7.3 Subjective Preference

We analyzed data from post-trial questionnaires with Friedman tests, where significant differences were found between the three UI conditions for "perception of hazards" ($\chi^2 = 12.61$, $p < 0.01$, $W = 0.55$), "safety" ($\chi^2 = 16.22$, $p < 0.001$, $W = 0.59$) and "disturbing" ($\chi^2 = 11.63$, $p < 0.01$, $W = 0.56$). Secondly, Wilcoxon tests were applied as paired tests: For "perception of hazards" (Fig. 14-a), significant differences were found in the smartphone UI vs. the forward-fixed UI ($V = 5$, $p < 0.01$, $r = 0.63$) and the smartphone UI vs. the HazARdSnap ($V = 5$, $p < 0.01$, $r = 0.63$), which means that participants subjectively perceived less hazards with the smartphone UI applied. For "safety" (Fig. 14-b), significant differences were located in the smartphone UI vs. the forward-fixed UI ($V = 30$, $p < 0.01$, $r = 0.58$) and the smartphone UI vs. the HazARdSnap ($V = 5$, $p < 0.001$, $r = 0.76$), indicating participants felt less safe when reading the smartphone UI. For "disturbing" (Fig. 14-c), significant differences were located in the smartphone UI vs. the forward-fixed UI ($V = 157$, $p < 0.01$, $r = 0.63$) and the smartphone UI vs. the HazARdSnap ($V = 140.5$, $p < 0.05$, $r = 0.54$), showing that the smartphone UI was considered more disturbing as compared with the forward-fixed UI and HazARdSnap. Besides, no significance was observed in the result of "readability" ($\chi^2 = 0.94$, $p = 0.63$, $W = 0.35$) (Fig. 14-d).

6 Discussion

Based on the collision data, participants encountered fewer collisions when using the forward-fixed UI (43%, $mean = 2.17, SD = 1.40$) and HazARdSnap (51%, $mean = 1.88, SD = 1.12$) as compared to the smartphone UI ($mean = 3.79, SD = 2.47$), suggesting that delivering UIs in a head-up fashion (e.g., via AR HMDs) is safer than using handlebar mounted smartphones. In addition, the head and gaze behavior revealed how AR based methods helped preventing dangers. Participants performed lower values on the head's and gaze's average angular speed data when using the forward-fixed UI and HazARdSnap. Which suggested that with the AR based methods, the participants were able to rotate their head and gaze in a less frequency and gain a longer period of stable view.

Furthermore, the average angle of head and gaze direction to forward data showed significant lower values when using the forward-fixed UI and HazARdSnap. This indicated that while reading UI content via an AR HMD, the participants were allowed to look forward more often and gain a higher chance to notice front hazards. It was interesting that both the forward-fixed UI and HazARdSnap performed similarly in general. It may be the case that users have a preference for the location of data in their field of view or that these methods may benefit from further context-dependent adaptation. Further refinement of both methods is planned as future work

On the other hand, the average speed of riding did not show any difference between the three conditions which indicates that interface

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3333336

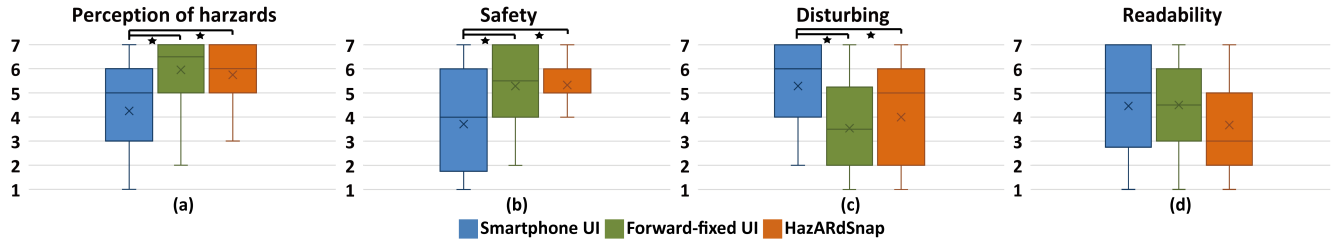

Fig. 14. Analysis of subjective ratings for (a) perception of hazards, (b) safety, (c) disturbing, and (d) readability. Participants were more likely to feel safer and have clearer views as forward-fixed UI and HazARdSnap were applied.

delivery methods are not likely to effect the user's preferred speed of riding. However, if the hazards were real objects, there may be differences since real crashes may cause the participants to be nervous and slow down. In conclusion, HazARdSnap did reduce the number of collisions in comparison with the phone-based UI.

Although significant differences did not exist for most metrics between the two AR-based UIs, HazARdSnap did help participants to outperform in both average angle of head facing direction and gaze direction to forward. Fixed UIs have a limitation that they can not be placed straight forward in the middle of the view, since they will block the view. According to the demonstration by Orlosky et. al. [26], they was also preferred to be placed at a forward-down position by users. Therefore, dangers may happen with this kind of methods when cycling toward to upper located hazards such as hang-up wires or lift-up arms of pedestrians. On the other hand, keeping the view straight forward allows cyclists to better gather information about signboards, traffic signals and routes, as shown by Orlosky et al. [27]. On the other hand, these results may also indicate that participants are offloading some of the cognitive demands of scanning the environment for hazards to the UI. However, while a reduction in gaze and head movements might suggest a reliance on our approach, it does not necessarily equate to a decrease in situational awareness. Alternatively, considering the results regarding the number of collisions, it could be interpreted as cyclists efficiently reallocating their attention, thereby enabling them to identify potential hazards more promptly. We recommend that future research delves deeper into the impact of UI delivery AR technologies on situational awareness and the balance between them.

Regarding enhancement of reading ability, our analysis of average cycling speed while reading data revealed that while reading the UI content, participants were able to ride faster using forward-fixed UI than the smartphone UI. This result was reasonable since looking down at the smartphone causes a huge loss of front view, therefore participants are likely to slow down in order to prevent a crash. On the other hand, the HazARdSnap did not show any significance with this data. We believe this was caused by the tendency of the UI to snap or "jump" from one object to another since several participants orally reported that they thought the snapping behavior made it difficult to find the position of the UI while cycling with HazARdSnap. Therefore when the UI was snapped, they were more likely to slow down and read. As this situation was noticed during the experiment phase, we applied a smoothing algorithm to the real implementation. Besides, the reading-to-completion time ratio data did not reveal any significance between the three conditions. Note that reading comprehension was not measured in the experiment, which is one limitation of the study.

In general, these results demonstrate that content delivered by AR HMDs would not likely have a negative effect on the time spent viewing the text, as long as the UI is delivered in a constant and readable scale. Moreover, the reading time data did not reveal any significance between the conditions. This suggests that different interface delivery approaches may not effect the participants' reading speed as long as the content has similar resolution. In addition, the subjective ratings of "Readability" also revealed no significance between the three conditions, which aligns with the quantitative results.

As one of our primary design considerations, we expected that HazARdSnap may positively affect the participants' subjective feelings of safety since our approach allows them to read while looking at a hazard at the same time. However, although both of the AR based methods were more preferred than the smartphone UI in the subjective rating results of "Perception of hazards" and "Safety", there was no significant difference found between them. We believe that this was because both the AR based methods did not require high frequency head rotations and the stability of their views were preserved. Furthermore, looking at the smartphone was considered as more disturbing by the participants. The AR based methods were more effective on helping the participants to gain a stronger feeling of stability which leads to a feeling of safety.

7 Limitations

In the experiment, we noticed that most of the participants, based on oral feedback, hesitated to use the HMD because of a lack of experience with AR and HMDs. This situation may slightly effect the results, especially since participants needed to read and ride at the same time. We suggest that future work focuses more on counter balancing this situation when conducting experiments.

Several practical considerations come into play when deploying AR technology, particularly the HoloLens 2, in real-world cycling scenarios. Firstly, the HoloLens 2 has a built-in thermal management system, and if it becomes too hot, it will shut down, potentially affecting the continuity of the AR experience. Secondly, the visibility of the display can be compromised when exposed to direct sunlight, making it challenging for cyclists to see crucial information. Thirdly, the field of view provided by AR glasses like the HoloLens 2 is limited, which may not be ideal for presenting a wide field of UI information simultaneously.

Furthermore, integrating AR glasses into cycling helmets raises concerns related to size and weight. The added bulk and heft could potentially increase the load on the cyclist's head, which may lead to discomfort or safety concerns if not adequately addressed. There is also the worry of the integrated device falling off during cycling.

These practical considerations highlight the need for further research and development to optimize the use of AR glasses in cycling scenarios and ensure a seamless and safe user experience.

8 Conclusions and Future Work

In this paper, we proposed a novel augmented reality-based information delivery system designed to help cyclists read UI content in a safe and easy manner. To accomplish this, we integrated real-time forward hazard detection and gaze tracking so that information can be snapped to a hazard that is in focus, which allows the cyclist to both view UI content and perceive the locations of hazards. We finally conducted an AR based within-subjects user study with 24 participants to study how HazARdSnap and a forward-fixed UI would perform in comparison to a traditional bike-mounted phone interface. Random text were delivered to participants via each UI, including a handlebar-mounted smartphone, a forward-fixed UI and our proposed HazARdSnap approach. Results from this study show that both the forward-fixed UI and our proposed approach effectively reduced the number of collisions with hazards in comparison to the smartphone UI. Moreover, the AR methods enabled users to maintain higher stability of head and gaze without a significant reduction of reading speed.

In addition, since the participants were not familiar with AR and HMDs, the systems may prove to be more effective when AR is more common and broadly accepted by users in the future. In contrast to the popular virtual reality based user tests on driving and cycling, our experimental setup is also a novel way to safely test AR based interfaces and provide more realistic but safe user tests. We presented and tested the system with relative non-urgent information just for demonstration. In practice, the UI content can be changed to more urgent information such as navigation popups, stopwatch and distance to the hazard.

One current limitation of our work is that multiple elements of information cannot yet be distributed amongst multiple hazards in the environment. One potential extension of this work is to allocate various spaces, such as the backs of multiple cyclists in a cycling team, for different types of information such as speed or biometrics. This might allow for easier readability without sacrificing the amount of information available. Our interface may also benefit from eye-based engagement or control to satisfy individual user preferences for information display. Another potential limitation of our work is that reading comprehension was not tested in the experiment. Comprehension of long meaningful sentences may require more attention and cause tunnel vision. Therefore, the outcome of the experiment may be different when testing those kinds of sentences.

In future studies, we plan to further explore wearable UI delivery approaches that can better help cyclists handle multiple tasks at the same time while preserving safety. We also plan to conduct experiments testing user interactions with UIs while cycling. Moreover, we would like to improve the proposed approach and investigate its effectiveness with more types of hazards such as vehicles, trees and/or roadblocks.

We hope that this work will encourage the development of other AR based UI delivery techniques for cyclists that can improve riding safety and maintain ideal readability as well as the exploration of augmented reality-based experimental designs that allow for safer testing of collision avoidance systems.

Acknowledgments

The authors would like to thank Prof. Denis Kalkofen from Graz University of Technology for his guidance. This research was funded in part by ONRG, grant N62909-18-1-2036.

References

[1] S. H. Berge, M. Hagenzieker, H. Farah, and J. de Winter. Do cyclists need hmis in future automated traffic? an interview study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 84:33–52, 2022.

[2] S. Chatterjee, K. Scheck, D. Küster, F. Putze, H. Moturu, J. Schering, J. M. Gómez, and T. 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*, pp. 426–432, 2020.

[3] J. Chen, P. S. Pyla, and D. A. Bowman. Testbed evaluation of navigation and text display techniques in an information-rich virtual environment. In *IEEE Virtual Reality 2004*, pp. 181–289. IEEE, 2004.

[4] S. Chong, R. Poulos, J. Olivier, W. L. Watson, and R. Grzebieta. Relative injury severity among vulnerable non-motorised road users: comparative analysis of injury arising from bicycle–motor vehicle and bicycle–pedestrian collisions. *Accident Analysis & Prevention*, 42(1):290–296, 2010.

[5] Chuanqi305. Caffe implementation of google mobilenet ssd detection network. `https://github.com/chuanqi305/MobileNet-SSD/`, 2017.

[6] A. Dancu, V. Vechev, A. A. Ünlüer, S. Nilson, O. Nygren, S. Eliasson, J.-E. Barjonet, J. Marshall, and M. Fjeld. Gesture bike: examining projection surfaces and turn signal systems for urban cycling. In *Proceedings of the 2015 international conference on interactive tabletops & surfaces*, pp. 151–159, 2015.

[7] D. De Waard, B. Lewis-Evans, B. Jelijs, O. Tucha, and K. Brookhuis. The effects of operating a touch screen smartphone and other common activities performed while bicycling on cycling behaviour. *Transportation research part F: traffic psychology and behaviour*, 22:196–206, 2014.

[8] D. Falanga, E. Mueggler, M. Faessler, and D. Scaramuzza. Aggressive quadrotor flight through narrow gaps with onboard sensing and computing using active vision. In *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 5774–5781. IEEE, 2017.

[9] D. M. Gavrila. Pedestrian detection from a moving vehicle. In *European conference on computer vision*, pp. 37–49. Springer, 2000.

[10] F.-T. Ghiurău, M. A. Baytaş, and C. Wickman. Arcar: On-road driving in mixed reality by volvo cars. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 62–64, 2020.

[11] E. Ginters. Augmented reality use for cycling quality improvement. *Procedia Computer Science*, 149:167–176, 2019.

[12] J. Hariyono, V.-D. Hoang, and K.-H. Jo. Moving object localization using optical flow for pedestrian detection from a moving vehicle. *The Scientific World Journal*, 2014, 2014.

[13] Kita-Kyushu City. Bicycle rules and etiquette many don't know about. `https://smakita.jp/rules/index_en.html`.

[14] Kyoto Police Office. Rules and penalties on riding bicycle. `https://www.pref.kyoto.jp/fukei/foreign/koki_k_t/jitensha/index.html`.

[15] R. Labayrade, D. Aubert, and J.-P. Tarel. Real time obstacle detection in stereovision on non flat road geometry through" v-disparity" representation. In *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, pp. 646–651. IEEE, 2002.

[16] W. Lan, J. Dang, Y. Wang, and S. Wang. Pedestrian detection based on yolo network model. In *2018 IEEE international conference on mechatronics and automation (ICMA)*, pp. 1547–1551. IEEE, 2018.

[17] S. Langlois and B. Soualmi. Augmented reality versus classical hud to take over from automated driving: An aid to smooth reactions and to anticipate maneuvers. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*, pp. 1571–1578. IEEE, 2016.

[18] League of American Cyclists. State bike laws. `https://bikeleague.org/StateBikeLaws`.

[19] Y. Li, F. Xue, X. Fan, Z. Qu, and G. Zhou. Pedestrian walking safety system based on smartphone built-in sensors. *IET Communications*, 12(6):751–758, 2018.

[20] X. Liang, T. Zhang, M. Xie, and X. Jia. Analyzing bicycle level of service using virtual reality and deep learning technologies. *Transportation research part A: policy and practice*, 153:115–129, 2021.

[21] P. Lindemann and G. Rigoll. Examining the impact of see-through cockpits on driving performance in a mixed reality prototype. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*, pp. 83–87, 2017.

[22] N. Matsunaga, R. Kimura, H. Ishiguro, and H. Okajima. Driving assistance of welfare vehicle with virtual platoon control method which has collision avoidance function using mixed reality. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1915–1920. IEEE, 2018.

[23] A. Matviienko, F. Müller, D. Schön, P. Seesemann, S. Günther, and M. Mühlhäuser. Bikear: Understanding cyclists' crossing decision-making at uncontrolled intersections using augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2022.

[24] J. Nataprawira, Y. Gu, K. Asami, and I. Goncharenko. Pedestrian detection in different lighting conditions using deep neural networks. In *IICST*, pp. 97–104, 2020.

[25] W. Oliveira, W. Gaisbauer, M. Tizuka, E. Clua, and H. Hlavacs. Virtual and real body experience comparison using mixed reality cycling environment. In *International Conference on Entertainment Computing*, pp. 52–63. Springer, 2018.

[26] J. Orlosky, K. Kiyokawa, and H. Takemura. Towards intelligent view management: A study of manual text placement tendencies in mobile environments using video see-through displays. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 281–282. IEEE, 2013.

[27] J. Orlosky, K. Kiyokawa, and H. Takemura. Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2):20–31, 2014.

[28] H. S. Park, M. W. Park, K. H. Won, K.-H. Kim, and S. K. Jung. In-vehicle ar-hud system to provide driving-safety information. *ETRI journal*, 35(6):1038–1047, 2013.

[29] T. Rateke, K. A. Justen, and A. Von Wangenheim. Road surface classification with images captured from low-cost camera-road traversing knowledge (rtk) dataset. *Revista de Informática Teórica e Aplicada*, 26(3):50–64, 2019.

[30] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv*

preprint arXiv:1804.02767, 2018.

[31] J. Ren, Y. Chen, F. Li, C. Xue, X. Yin, J. Peng, J. Liang, Q. Feng, and S. Wang. Road injuries associated with cellular phone use while walking or riding a bicycle or an electric bicycle: a case-crossover study. *American Journal of Epidemiology*, 190(1):37–43, 2021.

[32] D. Scaramuzza. Autonomous, agile micro drones: Perception, learning, and control. `https://roboticstoday.github.io/`, 2020.

[33] A. Shashua, Y. Gdalyahu, and G. Hayun. Pedestrian detection for driving assistance systems: Single-frame classification and system level performance. In *IEEE Intelligent Vehicles Symposium, 2004*, pp. 1–6. IEEE, 2004.

[34] K. Terzano. Bicycling safety and distracted behavior in the hague, the netherlands. *Accident Analysis & Prevention*, 57:87–90, 2013.

[35] T. Uchiya and R. Futamura. Consideration of presentation timing in bicycle navigation using smart glasses. In *International Conference on Intelligent Networking and Collaborative Systems*, pp. 209–217. Springer, 2021.

[36] A. R. A. Van der Horst, M. de Goede, S. de Hair-Buijssen, and R. Methorst. Traffic conflicts on bicycle paths: A systematic observation of behaviour from video. *Accident Analysis & Prevention*, 62:358–368, 2014.

[37] K. van Lopik, M. Schnieder, R. Sharpe, M. Sinclair, C. Hinde, P. Conway, A. West, and M. Maguire. Comparison of in-sight and handheld navigation devices toward supporting industry 4.0 supply chains: First and last mile deliveries at the human level. *Applied ergonomics*, 82:102928, 2020.

[38] P. Vansteenkiste, G. Cardon, E. D'Hondt, R. Philippaerts, and M. Lenoir. The visual control of bicycle steering: The effects of speed and path width. *Accident Analysis & Prevention*, 51:222–227, 2013.

[39] P. Vansteenkiste, G. Cardon, and M. Lenoir. Visual guidance during bicycle steering through narrow lanes: A study in children. *Accident Analysis & Prevention*, 78:8–13, 2015.

[40] T. von Sawitzky, P. Wintersberger, A. Löcken, A.-K. Frison, and A. Riener. Augmentation concepts with huds for cyclists to improve road safety in shared spaces. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2020.

[41] S. Wang, J. Cheng, H. Liu, F. Wang, and H. Zhou. Pedestrian detection via body part semantic and contextual information with dnn. *IEEE Transactions on Multimedia*, 20(11):3148–3159, 2018.

[42] T. Wang, G. Cardone, A. Corradi, L. Torresani, and A. T. Campbell. Walksafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads. In *Proceedings of the twelfth workshop on mobile computing systems & applications*, pp. 1–6, 2012.

[43] C. D. Wickens. Processing resources and attention. *Multiple-task performance*, 1991:3–34, 1991.

[44] D. Yeo, G. Kim, and S. Kim. Toward immersive self-driving simulations: Reports from a user study across six platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

Dr. Guanghan Zhao received his bachelors degree in Accounting from Donbei University of Finance and Economics, China, in 2015 and the M.F.A. degree in Interactive Media from University of Miami, US, in 2017. He then worked as a game developer for two years. He received his Ph.D. degree in information science and technology from Osaka University, Japan, in 2023. He is currently an invited researcher at Osaka University. His research interests include Computer Vision, Virtual Reality, Augmented Reality and Human-Computer Interaction.

Dr. Jason Orlosky received his bachelors degree in Computer Engineering from the Georgia Institute of Technology in 2006. He then worked in information technology for three years in the United States before returning to school in 2010 to study Japanese language and literature. In 2011, he studied abroad as a research student at Osaka University, where he graduated with his PhD in 2016. He is currently an Associate Professor at Augusta University with a dual-appointment at Osaka University.

Dr. Joseph L. Gabbard is director of the COGnitive Engineering for Novel Technologies (COGENT) Lab and Associate Professor of Human Factors at Virginia Tech's Grado Department of Industrial Systems Engineering. Dr. Gabbard is also an executive committee member of Virginia Tech's Center for Human-Computer Interaction; one of the largest, oldest and most diverse centers focused on HCI in the US. Dr. Gabbard received his PhD and MS in computer science from Virginia Tech; both his Master's thesis and Doctoral dissertation focused on usability of VR and AR systems respectively. Dr. Gabbard's research focuses on the connections between user interface design and human performance and specifically, the development of techniques to design and evaluate novel user interfaces for augmented reality technologies and virtual environments. Dr. Gabbard has been a pioneer in usability engineering with respect to applying to and creating methods for new interactive systems for close to 25 years. With funding from a variety of sources, he has developed several innovative methods for designing complex interactive systems and assessing their usability and impact on human performance and disseminated this work in over 100 publications.

Dr. Kiyoshi Kiyokawa received his MS and PhD degrees in information systems from the Nara Institute of Science and Technology in 1996 and 1998, respectively. He has been a professor with the Nara Institute of Science and Technology since 2017, where he leads the Cybernetics and Reality Engineering Laboratory. He worked for the Communications Research Laboratory from 1999 to 2002. He was a visiting researcher with Human Interface Technology Laboratory of University of Washington from 2001 to 2002. He was an associate professor with the Cybermedia Center at Osaka University from 2002 to 2017. His research interests include virtual reality, augmented reality, human augmentation, 3D user interfaces, CSCW, and context awareness. He is a member of IEEE and ACM, and a Fellow of the Virtual Reality Society of Japan (VRSJ). He is an Associate Editor-in-Chief of IEEE TVCG and an inductee of the IEEE VGTC Virtual Reality Academy (Inaugural Class).