HindSight: Enhancing Spatial Awareness by Sonifying Detected Objects in Real-Time 360-Degree Video

Eldon Schoop
UC Berkeley EECS
Berkeley, CA, USA
eschoop@berkeley.edu

James Smith
UC Berkeley EECS
Berkeley, CA, USA
james.smith@berkeley.edu

Bjoern Hartmann
UC Berkeley EECS
Berkeley, CA, USA
bjoern@eecs.berkeley.edu

ABSTRACT

Our perception of our surrounding environment is limited by the constraints of human biology. The field of augmented perception asks how our sensory capabilities can be usefully extended through computational means. We argue that spatial awareness can be enhanced by exploiting recent advances in computer vision which make high-accuracy, real-time object detection feasible in everyday settings. We introduce HindSight, a wearable system that increases spatial awareness by detecting relevant objects in live 360-degree video and sonifying their position and class through bone conduction headphones. HindSight uses a deep neural network to locate and attribute semantic information to objects surrounding a user through a head-worn panoramic camera. It then uses bone conduction headphones, which preserve natural auditory acuity, to transmit audio notifications for detected objects of interest. We develop an application using HindSight to warn cyclists of approaching vehicles outside their field of view and evaluate it in an exploratory study with 15 users. Participants reported increases in perceived safety and awareness of approaching vehicles when using HindSight.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
360-Degree Video; Computer Vision; Sonification; Augmented Perception

INTRODUCTION

The human visual system has both biological and cognitive constraints. Our vision spans a usable field of roughly 114 degrees [14], and our anatomy restricts our sharpest, foveal vision to a field of only 5.2 degrees [40]. Cognitively, as we become absorbed in a task, our "locus of attention" narrows [33]; i.e., we tune out external stimuli, increasing our focus but possibly drowning out important events such as alarms or environmental dangers. Interfaces which can redirect a user's attention to these overlooked stimuli have the potential to prevent serious accidents. Our research goal is to augment real-time spatial awareness for objects that are outside of a person's visual field.

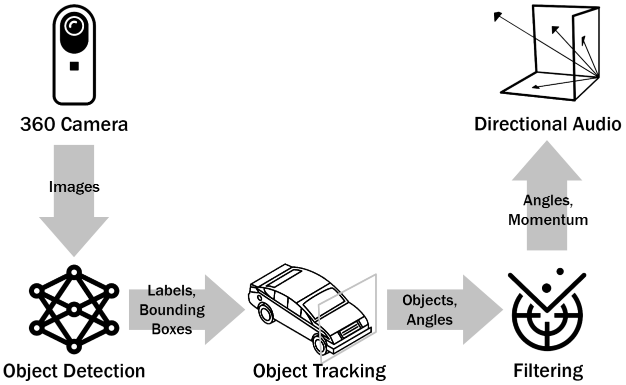

Figure 1. HindSight uses a neural network to detect objects from live, ego-centric 360-degree video, a filter bank to extract relevant ones, and a application-specific sonification program to convey results to a user.

Some approaches substitute all information for specific senses–e.g., by using a head-mounted display to show a LIDAR point cloud [25] or a live 360-degree video stream [2]. Repurposing the visual system is potentially powerful, but such systems are not easily integrated into daily activities because they require an adaptation period before use and can create hurdles in social interactions. We seek to develop a system that enhances spatial awareness by redirecting attention to objects outside a user's visual field without impeding natural senses.

We introduce HindSight, a wearable system that increases spatial awareness by detecting relevant objects in live, ego-centric 360-degree video and sonifying their location and properties through bone conduction headphones. Our approach draws upon advances in computer vision to identify points of interest in a user's surroundings, and work in delivering continuous feedback for physical tasks to notify the user when necessary to redirect their attention.

HindSight streams 360-degree video from a head-worn camera to a real-time object detection neural network running on a laptop worn in a backpack (Figure 2). The system filters the neural network's output and picks the most relevant objects

CHI'18, April 21âĂŞ26, 2018, Montreal, QC, Canada.

ACM ISBN 978-1-4503-5620-6/18/04.
https://doi.org/10.1145/3173574.3173717

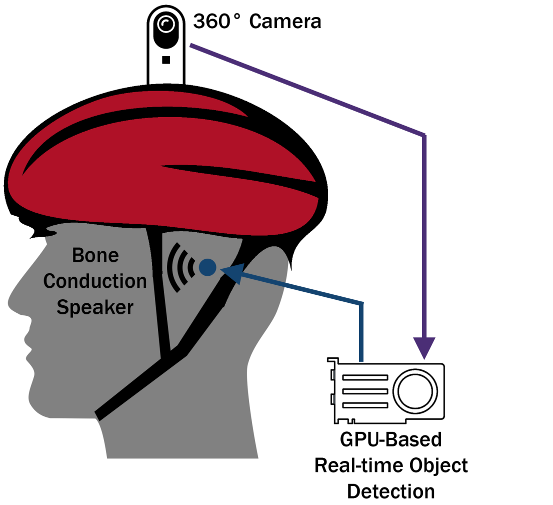

Figure 2. HindSight uses a spherical camera mounted to a bike helmet to capture a user's surroundings. Video is streamed to a laptop worn in a backpack.

for the application. These objects are sonified, conveying attributes such as their type, location, and velocity. The user hears audio through Bluetooth wireless bone conduction headphones, which transmit vibrations directly through the skull. The key benefit of using bone conduction is it leaves the ears unobstructed, enabling retention of normal auditory acuity.

We develop an application for HindSight: a program to augment cyclists' sensory ability by warning users of vehicles which are approaching in a potentially dangerous way. We calculate potential danger by attributing momentum and direction data to oncoming vehicles outside the cyclist's field of view. The momentum and direction data of vehicles approaching the cyclist are used to calculate a directional "danger" metric, which is sonified by modulating beeps using panning (to indicate direction) with tempo and pitch (to indicate danger). We provide a technical evaluation of our system to measure its precision and the window of usable time a cyclist has to react to its output. On average, we find HindSight detects potentially unsafe approaching vehicles 1.89 seconds ($\sigma = 0.40$) s before they would hit the bicycle.

We conduct an exploratory user evaluation with 15 participants to determine how users perform with our system. When comparing our system to the control condition, users reported perceived increases in safety ($\mu = 4.00, \sigma = 0.82$), time to react ($\mu = 3.73, \sigma = 0.57$), comfort ($\mu = 3.47, \sigma = 0.96$), and awareness ($\mu = 3.53, \sigma = 0.96$), on a 5-point Likert scale. Several participants noted our system detected potential dangers they would have otherwise missed: *"[HindSight could] sense the danger from the views that people not normally see"* (P4). Open-ended feedback revealed areas for potential improvement: *"If it could detect danger slightly (just slightly) sooner, that would be better"* (P14). On a 5-point Likert scale, users expressed they would use the system during their real commutes if it was available ($\mu = 3.87, \sigma = 0.96$).

In summary, we contribute: 1) The HindSight real-time computer vision system for detecting and sonifying objects of interest in 360-degree body-worn video; 2) an application of Hindsight for augmenting cyclists' spatial awareness; 3) A technical evaluation and exploratory evaluation of this application.

RELATED WORK

HindSight builds on prior work in three primary areas: devices that enhance spatial awareness, systems which provide real-time feedback for physical tasks, and tools for interpreting 360-degree video.

Enhancing Spatial Awareness

Devices which enhance spatial awareness ingest information from a user's surroundings, process it into a meaningful representation, and output it via visual, audio, or haptic displays.

Augmenting Field of View

We are inspired by systems such as FlyViz, which augments a user's field of view by displaying reprojected 360-degree panoramic video into a Head-Mounted Display (HMD) [2]. FlyViz effectively remaps a user's surroundings to their visual field, but requires an adaptation period before it can be used comfortably. The Skully motorcycle helmet [42] projects a rear-facing camera feed onto a transparent HMD. LiDARMAN takes this idea further by projecting a 3D point cloud from a head-mounted Lidar scanner into an HMD [25]. Closely related to our work is SpiderVision, which blends front and rear-facing video feeds into an HMD based on motion detected behind the user [10]. Rather than use motion to trigger additional visual input, HindSight identifies objects around a user and determines if they are important enough to redirect the user's attention. HMD-based solutions face multiple hurdles to real-world use. First, HMDs have limited resolution and field of view compared to natural human vision. Second, social acceptability of their continual use is not yet established. Finally, some users additionally experience *virtual reality sickness* when using VR headsets and HMDs [26]. In contrast with this prior work, while we us an HMD as an experimental apparatus in our exploratory study, HindSight exclusively uses audio for output during use.

Assistive Technology

Assistive devices for the visually impaired ingest visual or spatial data and encode this information into a different sensory output, such as audio or vibrotactile displays.

Systems to aid the visually impaired often employ sonification techniques to help users create a mental representation of their surroundings. As early as 1974, Sonar has been used to sonify obstacles in front of a user as a navigation aid [16]. Depth and color of objects in a scene can be sonified with rich audio, such as orchestral instruments, [11]. We draw upon this work to develop our sonification framework, but focus on objects *outside* of the user's visual field. HindSight is not designed to *replace* the visual system, but *augment* its capabilities.

Varying degrees of computational intelligence can be used to extract higher-level information from images, from relying on remote human assistance to on-device or cloud-based machine learning tools. VizWiz uses on-demand, crowdsourced support to answer visual questions for pictures taken from a smartphone [3]. Computer vision algorithms can help users discriminate objects [9], locate visual markers [46], and describe scenes with machine-generated natural language [44]. Depth cameras can be used to create an interactive map of the user's surroundings [19] and identify obstacles in real-time, such as unoccupied chairs and walls [45]. HindSight uses machine learning based object detection to identify objects outside of a user's visual field and alert them when necessary. The goal of HindSight is not to show *all* object information, but only *relevant* objects which require attention.

Enhancing Awareness in Traffic

Our cycling application builds off related work in increasing awareness of traffic situations. Projected AR displays have been used to alert other drivers of a cyclist [15, 5] and display a "safety envelope" where others may pass the bicycle [7]. Audio [35, 18] and haptic [8, 1] feedback can increase driver awareness of other vehicles. Sonification can increase detectability of approaching vehicles in environments with background noise [17, 22], especially in the case of quiet electric vehicles [24, 23]. Diedrichs and Parizet separately describe design principles for sonifying approaching vehicles, such as amplitude modulation, pitch, and rhythm [8, 31]. HindSight leverages these design principles to generate audio to alert the user of oncoming vehicles outside their visual field.

Real-Time Feedback for Physical Tasks

Actions taken in the physical world can entail a sense of *risk*, i.e., actions are often irreversible, and may potentially harm the user if performed incorrectly [20]. HindSight operates in the physical world, and our cycling application exhibits this type of risk. We are inspired by digital fabrication devices that provide real-time feedback to reduce or mitigate risk.

Devices can make users aware of variables that are relevant to a task but not readily perceivable by a person. Projected AR visualizations can reveal the otherwise invisible forces inside CNC machines [30] or warn users when they are drilling too far into a surface [37]. Haptic feedback can alert users to take corrective action when cutting a block of material if they are approaching the edges of a predetermined model [47]. HindSight draws upon the metaphor of using real-time feedback to display variables in the environment and suggest the user take corrective action. In particular, our cycling application provides audio feedback to redirect the user's attention to potentially dangerous situations, prompting the user to take corrective action if necessary.

Exploring and Interpreting 360-degree Video

360-degree video captures information from the camera's entire surrounding area, which can be explored by users manually or interpreted with computer vision algorithms. Research system have allowed users to annotate prerecorded 360-degree video [32] or explore streaming video in real-time from a head-worn camera array [28]. Computer vision techniques have been used to recognize the faces of speakers in 360-degree videos and generate a simulated "multi camera" output [36]. Pano2Vid generalizes this approach, simulating human motion of an artificial camera to track areas of interest in 360-degree video [41]. Deep neural networks have also been used to correct skew in 360-degree video [43]. HindSight utilizes computer vision techniques to detect objects in 360-degree video and requires a 360-degree camera to dynamically adjust the analyzed field of view when head orientation changes from travel direction, i.e., the user does not look straight ahead.

HindSight uses monocular, optical sensing to detect vehicles. This is one of several possible techniques that has been used in the literature [38, 27]. One distinguishing feature of our approach is that we use a panoramic camera which captures the relative angle of each pixel, yielding accurate direction information for detected vehicles.

HindSight DESIGN CONSIDERATIONS

At a high level, HindSight seeks to enhance the spatial awareness of users while preserving their ability to rely on unaugmented sensory input. Our technique was guided by several overarching guidelines:

Do not impede natural sensory input: For safety and social acceptability, we aim to preserve real-world sensory input. This precludes uses of opaque HMDs and suggests audio or haptic displays. However, audio stimuli that block out environmental sound are not appropriate. To satisfy these guidelines, we rely on delivering audio notifications through bone conduction headphones, which leave the ear canals unobstructed. It is possible for users to perform auditory and visual tasks at the same time [12], so we believe audio to be a proper interface for a warning system for bicycle users.

Importantly, cyclists may not always be able to rely on natural audio cues, e.g., in dense traffic or in urban areas where sound is reflected from multiple facades. In these situations, HindSight could provide additional, resolvable audio cues.

Provide real-time interpretation: Extracting higher-level information from a scene can provide more concise, semantically meaningful information to users. We use a computer vision pipeline to recognize objects in the environment and only sonify detected objects that are of critical importance.

Be conservative in information delivery: The system should only engage the user when important and necessary. The level of display should be proportional to the importance of the message, i.e., ramp up the level of warning with the level of danger. Our sound design is further informed the particularities of bone conduction headphones.

Designing Audio for Bone Conduction

Using bone conduction as our information display poses several design challenges over traditional headphones because audio does not enter the user's ear canal, but is instead transmitted through the user's skull through vibrations. The the primary benefit is that bone conduction headphones can be worn safely in situations where the users must still use their ears as an important channel for information.

The primary goal of our sound design for our cycling application is to provide a clear auditory message to the user of our system that there is a danger in their vicinity. We design the audio such that it can transmit three key dimensions of information to the user: direction, proximity, and type of danger. We use Hermann et al.'s sonification framework [12] together with principles from SAFERIDER [8] to inform our design decisions, as described below.

Parameterizing Information

Auditory displays fall into the broad categories of alarms, status indication, data exploration, and entertainment [12]. HindSight is an alarm system, because its primary purpose is to indicate the presence of a dangerous object. HindSight has properties of *safety auditory displays*, which prompt for corrective action, and *imminent auditory displays*, which alert time-critical corrective action is needed [8].

We map our three primary dimensions of information (direction, proximity, and type of danger) in the following ways:

Direction is mapped to directional audio. By mapping the direction of the dangerous object to directional audio, we aim to assist the user in localizing that object so that they can respond to it appropriately. Because of the limited effectiveness of using binaural audio with bone conduction headphones (described below), we use panning to convey spatial information.

Distance to the detected object is mapped to tempo and pitch. We take inspiration from parking assist and cross traffic alert systems in automobiles, which emit a sequence of beeps of increasing tempo as the car is approaching obstacles (or vice versa). We play the given sound at a faster rate and increase its pitch the closer an object is to the user. This is chosen to create a sense of urgency in the user as the object approaches, reflecting the need for time-critical corrective action [8].

Types of objects are mapped to different timbres. Categorical types of data should be represented by changing acoustic variables like timbre. This makes it easier for users to isolate different sounds and still resolve the direction of these sources. The pitches of the sounds we play range between 500Hz and 2000Hz, proportional to an object's danger level. This range transmits clearly using our bone conduction headphones, and the 2000Hz pitch has been recommended for urgent safety indicators [8].

Only Show Two Objects at Once Timbre and directional audio are two of the best ways to help users resolve multiple simultaneous sound sources [12]. However, we found it is easy to get overwhelmed with information when more than two moving objects are represented with audio. Therefore, we chose to limit our system to only display up to two most dangerous objects, to limit the amount of cognitive load our system places on a user.

We also design for several particularities of bone conduction:

Exploit the Boundary between Haptics and Audio Lower frequencies behave almost haptically on bone conduction headphones, providing vibrating sensations at the contact points with the user's head. We exploit this property by overlaying low frequency sounds in our audio to help direct the user's attention to the direction of the audio.

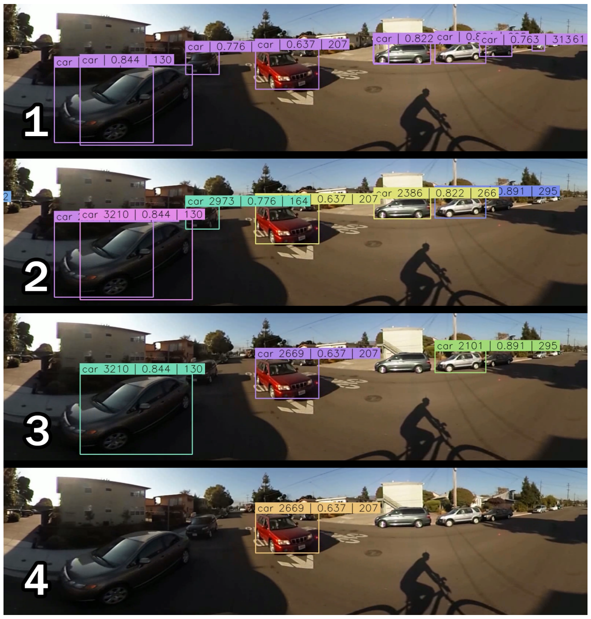

Figure 3. Detected objects at intermediate filtering stages of HindSight: (1) The neural network outputs bounding boxes of detected objects. (2) Objects are tracked frame-to-frame. (3) Only objects moving nearer to the user are kept. (4) Only object approaching the user are kept.

Use Panning Instead of Binaural Spatialization Standard audio spatialization algorithms do not work well for bone conduction and it is difficult for users to resolve the direction of spatialized audio. This is due to the fact that most audio spatialization software uses Head-Related Transfer Functions (HRTF) to determine how much audio should go to each ear. HRTFs are calculated based on a model of the user's ear and head size, and assume that audio is entering through the ear canal. In our application, audio is passed to the ear drum through the skull. Because humans skulls have different acoustic properties, existing HRTFs are not appropriate [6].

As a workaround to this limitation, we lower the dimension of data by instead panning the audio between the left and right channels of the bone conduction headphones. This provides reasonable directional feedback (users can still resolve the general direction of danger) at the cost of not allowing as many unique angles of direction.

SYSTEM ARCHITECTURE

Our system consists of a 360 degree video camera attached to a bicycle helmet, connected via USB to a laptop in the user's backpack. The laptop is connected to a pair of bone conduction headphones via bluetooth.

Image Acquisition

We acquire a stream of 1280 x 720 pixel equirectangular images at 15Hz using a Ricoh Theta S camera and process them using OpenCV.

The equirectangular image format projects a spherical image onto a rectangular image by mapping latitude coordinates of the spherical image directly to x pixel coordinates, and longitude values directly to y pixel coordinates [39]. A major downside to this format is that the image distorts near the poles, but minimizes distortion near the equator.

To obtain the best classification performance, we perform some processing on the frame of video before passing it to our object detector. For our application, the top and bottom 27° of the image generally contain no useful data (the user's helmet and the sky) and are removed. The rest of the image is cut into three parts to produce nearly square images, which minimizes aspect ratio distortion and increases performance with our object detector. These partitions overlap slightly to aid resolving objects which traverse their boundaries.

Object Detection

We use the YOLOv2 realtime object detection framework [34] because it provides accurate, low latency predictions. We additionally considered SSD [21], which achieved similar performance, but with higher latency on our particular hardware.

Using pretrained model weights, YOLOv2 is capable of classifying 80 labels, several of which are traffic related: car, truck, bus, bicycle, person, stop sign, traffic light. The output from this step is a list of bounding boxes, labels, and confidence values. The object detector is generalizable to multiple classes of objects through retraining on example images of desired classes. Classification takes about 50 ms when using the down sampled input images described earlier.

All classification is done in real time on a laptop carried in the user's backpack. The laptop is an Origin EON17-SLX with an Intel i7-6700K 4.0 GHz processor, 16GB DDR4 RAM, and a GeForce GTX 980 video card with 8GB DDR5 RAM running Windows 10. Software used are Python 3.6, Tensorflow with CUDA extensions enabled, and OpenCV 3.2. The high-performance GPU of the laptop is critical in order to run a deep neural network such as YOLOv2 in real time.

Object Tracking

Output from YOLOv2 provides us with no frame to frame coherence of objects. Frame to frame tracking is important because we wish to filter objects based on their behavior. We developed a simple and fast algorithm for approximating the most likely bounding box for a given object between two frames, with an acceptable amount of accuracy. Our algorithm greedily merges weighted object bounding boxes over a sequence of frames, "remembering" previous merges.

Object Filtering

Several filters are applied to the set of tracked objects to narrow down which objects the user might find the most important. The type of filtering applied depends heavily on the application that the system is used for. The following filters are used for the bicycle in traffic scenario. For each filter, we manually count falsely identified dangerous objects during a 21-second training video clip and report the number of false positives (Objects which are falsely tracked and reported as dangerous).

Only Accept Vehicles Outside the User's Visual Field

Our first filter eliminates objects irrelevant to cycling in traffic from consideration (e.g., toasters, airplanes, clocks). We also eliminate any objects that are in the front 110 degrees of the user's visual field (slightly less than human peripheral vision). This is trivially calculated because the camera position tracks the user's gaze, as a *head-stabilized* configuration [4].

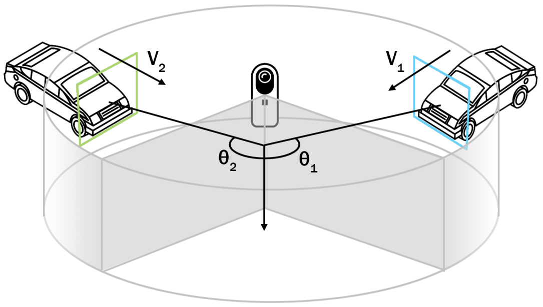

Figure 4. HindSight only notifies users of objects which are approaching and outside of their field of view. We approximate the human visual field to 110°.

Only Display Growing Objects

Objects whose bounding boxes are decreasing in size over time can be assumed to be moving away from the user, and likely pose no danger. We calculate the square root of the area of each bounding box over a time period of 10 frames (300 ms) of video and fit a linear function to approximate the growth of the bounding box. If the slope of this line is positive, then the area of the box is trending larger, and the object is coming closer to the user. Any object with a negative area-growth slope is removed from consideration. The growth filter reduces the number of false positives from 122 (vehicle filter only) to 46 in our sample data.

Orientation Filtering

For our application, objects that are approaching the user from behind pose the most risk, and any object that is moving away from the user in their direction of travel most likely passed by them. We thus reject objects which are not traveling in the direction of the user from behind them. The orientation filter reduces the number of false positives to 4 in our sample data.

We determine the latitude of an object by considering its center point and subtract 180 from it to determine its angle from the rear of the user. Then we take the absolute value of this to determine absolute angle from the rear of the user.

$$\sphericalangle_{user} = \text{abs}(\sphericalangle_{object} - 180)$$

We fit a linear function to these values over a window of 10 video frames, in parallel with the object growth filter (the orientation filter does not introduce additional delay). If the slope is positive, the object is most likely moving toward the user from behind. We filter out any object with a negative slope from consideration, leaving only objects moving in the direction that the user is looking. An IMU attached to the user's helmet can base this calculation on the direction of travel as opposed to the direction the user is facing by offsetting the center point (180°) by the head orientation value.

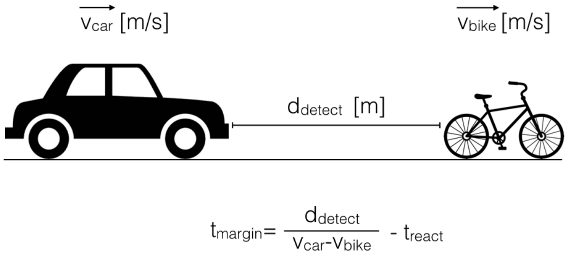
$\overrightarrow{v_{car}}$ [m/s]
$\overrightarrow{v_{bike}}$ [m/s]
d_{detect} [m]
$t_{margin} = \frac{d_{detect}}{v_{car} - v_{bike}} - t_{react}$

Figure 5. We calculate the system detection time t_{detect} using the relative velocity of the car and bicycle $\vec{v}_{car} - \vec{v}_{bike}$ and by finding the average detection distance d_{detect} of our system. Margin of safety t_{margin} is calculated using $t_{react} = 1.6$ s from Olson and Sivak [29].

Removing Additional False Positives

We apply a final filter that requires an object to have made it through the previous filters for at least 3 frames. This minimizes briefly appearing misclassfied objects, as well as any object that erroneously passed through the set of filters. This provides less distractions for the user, so they don't need to divert attention towards these false positives. The averaging filter reduces the number of false positives to only 1 (a parked car) in our training video sample.

After all filters are applied, we calculate a danger metric that is roughly proportional to each object's momentum. This is equal to an approximation of the object's mass times the rate at which the bounding box is growing.

$$D_i = \underset{x}{\operatorname{argmax}}(M_{x,i} V_{x,i})$$

Where D_i is the most dangerous object for frame i, $M_{x,i}$ is the approximate mass, and $V_{x,i}$ is the object's bounding box growth rate in frame i.

Audio Output

Output from the filtering process goes into an audio system that synthesizes and spatializes sounds based on which objects are considered the most dangerous. Audio is sent over Bluetooth to AfterShokz Trekz Titanium bone conduction headphones.

Sounds played to the user were authored in FL Studio 12, a professional digital audio workstation. A virtual loopback MIDI interface, loopMIDI, was loaded onto the laptop to allow our software to communicate with FL Studio. Custom MIDI control messages were specified to allow our software to start, stop, and spatialize various sounds. FL Studio listens for these messages and controls audio playback accordingly. The benefits of this approach are the robustness it provides when trying different user interfaces. Any software that can listen to MIDI can respond to our system, providing a many ways our system can connect to various actuators.

TECHNICAL EVALUATION

We perform a technical evaluation to characterize the precision of our system and the margin of safety it provides to users with the cycling application. For test data, we run our system on 8 sample videos from a cyclist's point of view in traffic situations. There are two main classes of these videos: 5 of them have a car approaching roughly 15 mph (24 km/h) relative to the bicycle, the other 3 have vehicles moving the same speed as the bicycle.

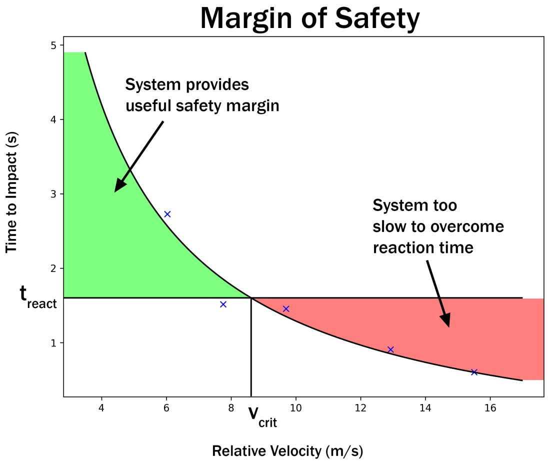

Figure 6. HindSight provides adequate time to react when vehicles approach the user at or under $v_{crit} = 8.62 \pm 1.24$ m/s (green fill, top left), assuming a baseline reaction time of $t_{react} = 1.6$ s. The points × are detection distances measured from system use. The isodistance curve is fitted from the average of the distances.

We use a pretrained model for our object detector, which has been characterized by its creators to have a Mean Average Precision (mAP) 78.6 [34]. Once objects have been tracked and filtered, they remain detected by the system with a confidence of 89.7% per frame over our training data.

We define the "margin of safety" t_{margin} of our system as the difference in time between when HindSight detects a potentially dangerous vehicle and a baseline reaction time t_{react} to avoid accidents in traffic. We compute t_{margin} by assuming a constant relative velocity between the bicycle and an approaching vehicle $v_{car} - v_{bike}$ and determining the average distance d_{detect} at which HindSight detects objects (Figure 5).

As intended, the cars moving the same speed as the bicycle in the 3 videos are not detected by our system because of the bounding box growth filter. For the remaining 5 videos, we select the first frame where our system detects a dangerous car and visually determine how far from the bicycle the detected car is. To determine the average detection distance d_{detect}, we assume an average car length of 4.7 meters and constant relative velocity $v_{car} - v_{bike}$ of 6.7 m/s (15 mph).

On average, the system detects the car 1.89 seconds ($\sigma = 0.40$ s) before the car would hit the bicycle. These values were determined by observing the videos and counting the number of frames from the time that the dangerous object is detected to the time it would hit the user.

We determine the margin of safety for our system by plotting detection time values against approximate relative velocity (Figure 6). Relative velocity is calculated by dividing the distance the vehicle needs to travel to hit the bicycle by the amount of time it takes the vehicle to reach that point. An inverse function is fit to these points to generate an isodistance curve that represents the average distance our system detects a dangerous vehicle.

$$t(v) = -0.65\ \text{s} + \frac{19.4 \pm 2.8\ \text{m}}{v\ \text{m/s}}$$

Assuming a maximum[1] baseline reaction time of 1.6s to an unexpected roadway obstacle t_{crit} [29], we can determine the maximum speed that a car can be moving relative to the user for our system to provide enough time to react, v_{crit}.

$$v_{crit} = v(t_{react}) \quad = \frac{19.4 \pm 2.8\ \text{m}}{(1.6) + 0.65\ \text{s}} = 8.62 \pm 1.2\ \text{m/s}$$

Therefore, our system can operate safely in situations where nearby vehicles are traveling at most 8.62 m/s (19.28 mph, 31.03 km/h) relative to the bicycle. Assuming an average bicycle speed of 10 mph (16 km/h) means HindSight can currently handle situations where cars travel around 25 mph, a common city speed limit, but that it may need earlier detection to handle speed limits 35 mph (55 km/h) or above.

EXPLORATORY EVALUATION

To determine whether HindSight's cycling application can increase users' awareness of vehicles approaching in a potentially unsafe way, we conducted an exploratory evaluation. We recruited 16 participants (11 male, 5 female) using university mailing lists, all graduate students with experience riding a bicycle. 13 participants had ridden a bicycle in traffic, with most participants reporting only occasionally doing so ($\mu = 2.7, \sigma = 1.4$ on a 5-point Likert scale where 1 is "I have never ridden a bicycle in traffic" and 5 is "I commute on a bicycle 5+ times per week").

VR-Based Simulation

Because of the safety concerns of using a prototype system in a live traffic situation, we developed a simulator to approximate the experience of riding a bicycle in light traffic while using HindSight. In our simulator, participants watch 360-degree videos recorded from the point of view of a bicyclist via a head-mounted VR display. Videos shown were not stereoscopic. Video data is fed into HindSight to generate sounds from prerecorded objects during trials.

Videos of live traffic situations were recorded by two researchers in a suburban location with minimal vehicular and pedestrian traffic. One researcher rode a bicycle with our system running and capturing 360-degree video, while the other drove a car to simulate various traffic situations. A studio-quality stereo audio recorder was attached to the bicycle to collect environmental sound with approximate spatial cues.

In the evaluation apparatus (Figure 7), a Unity[2] application plays the 360-degree videos and recorded audio back to an Oculus DK2 VR display and in-ear headphones. Users can look around as the video plays using head orientation data from the DK2's IMU. This data is also used in calculations for the HindSight Orientation Filter and for logging metrics for the user evaluation. Audio cues for objects are played through Trekz bone conduction headphones which are positioned in front of the regular headphones on the participants' skull.

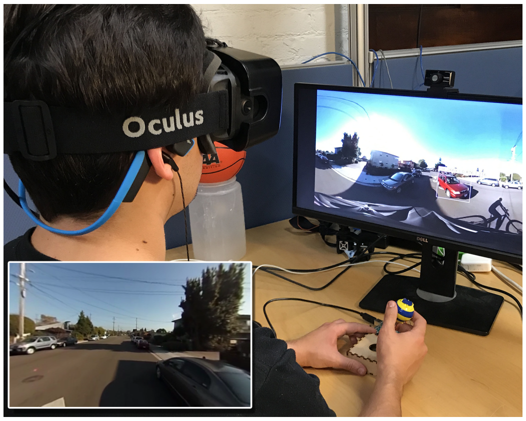

Figure 7. Users wear an Oculus DK2 head-mounted display which plays back 360-degree videos and manipulate a joystick to indicate areas with potential danger. Bottom left: users see images of the scene through a "virtual camera"

A minor technical difference between using our system live and in the simulator is that our 360-degree camera is capable of recording video at 30 Hz, but only capable of streaming video at 15 Hz. We expect this impact on results to be small.

Method

Users were instructed to sit in a kneeling chair to emulate riding a bicycle and were fitted with our evaluation apparatus. Users were then shown 9 distinct videos. The first, consistent across all trials, was played twice–without and with the HindSight system activated–to familiarize users with our experimental apparatus and system. The remaining 8 videos were shown to the users in random order, and with HindSight randomly enabled for each. This allowed us to obtain a fair distribution of results for each video with a roughly even number of users trying each video with and without HindSight . All sessions lasted under 30 minutes, and each participant successfully completed the evaluation. One participant's results were omitted due to a technical error which caused their data to not be logged.

During each video, participants were instructed to point a provided joystick towards the area where they considered the most potential danger was in the scene, if it existed. At the end of the evaluation, users were asked to fill out an exit survey. Questions were included open-ended answers and 5-point Likert scales (1 = "Strongly Disagree", 5 = "Strongly Agree"). Likert scale questions were phrased as follows: *(Awareness) The system identified dangerous situations I would NOT have noticed without it, (Comfort) I felt more comfortable when the system was activated, compared to when it was not, (Safety) I felt safer when the system was activated, compared to when it was not, (Stress) I felt LESS stressed when the system was*

[1] Olson and Sivak suggest once drivers are *alerted* of an upcoming obstacle beforehand, 95th percentile perception-response time for the same population drops to about 1.1 seconds

[2] https://unity3d.com/

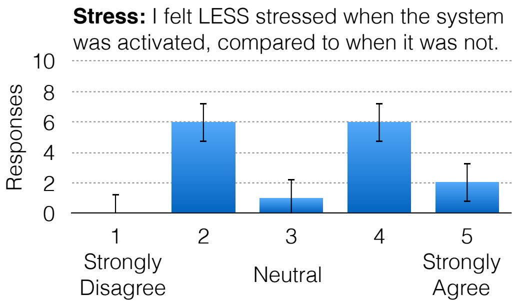

Figure 8. Participants were split on whether HindSight increased or decreased their stress level.

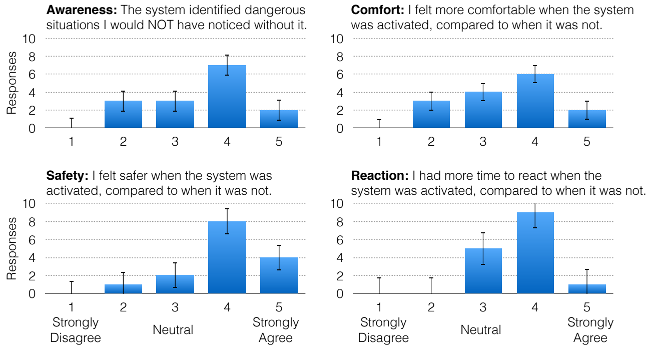

Figure 9. In the exit survey, participants generally gave positive subjective ratings to HindSight on Likert scales asking about awareness, comfort, safety and reaction time.

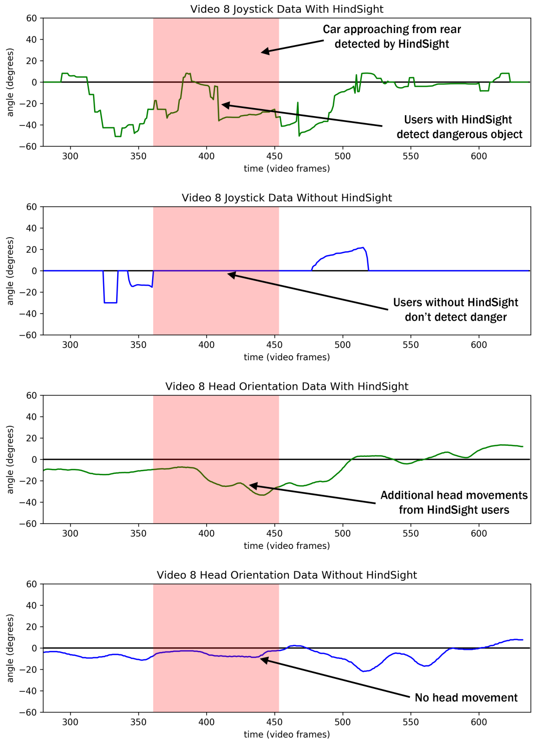

Figure 10. Top two: averaged joystick angle for the "right turn" video for users using and not using HindSight. Bottom two: averaged head orientation for the "right turn" video. The red region shows when an unexpected car appears behind the users. 7 out of 11 participants who used HindSight reacted to the unexpected passing vehicle, while 0 out of 4 who *did not* use HindSight reacted.

activated, compared to when it was not, (Reaction Time) I had more time to react to situations when the system was activated, compared to when it was not.

RESULTS AND DISCUSSION

Increased Perceived Awareness, Safety, and Reaction

In the exit survey, participants generally expressed positive reactions to using our system (Figure 9). Participants reported a perceived increase in awareness, ($\mu = 3.53, \sigma = 0.96$), defined as the ability to identify dangerous situations they otherwise would not have noticed. Some participants explicitly commented on this aspect in open responses: *"It identified passing cars before I could hear them passing by."* (P9); and *"[it did well on] Notifying cyclers of unseen approaching objects, especially if they did not hear/anticipate it."* Participants also reported a perceived increase in time to react ($\mu = 3.73, \sigma = 0.57$), comfort ($\mu = 3.47, \sigma = 0.96$), and safety ($\mu = 4, \sigma = 0.82$) when using HindSight. Full distributions are shown in Figure 9.

Bimodal Response on Perceived Stress

Interestingly, we see a bimodal distribution for perceived stress (Figure 8); for some HindSight increased stress, while for others it decreased their perception of stress. P8, P9, and P12 reported that our system generated some false positives, which may have led to stress: *"Too many false positives. False positives might stress out the user"* (P9). On the other hand, P11 expressed the desire for richer feedback: *"the system gave me a decent view of everything, but I still felt like I wasn't getting the full view even when there was no danger."* More training would let users become more familiar with the system and could eliminate increased stress for some. As P1 notes, (*"I definitely became more accustomed to the system as time went on"*) and P5 remarks (*"My lack of comfort or increased stress with the system might've just been because I'm not used to it. I recognize that it sometimes alerted me to things sooner than I would've noticed them but it also felt a little distracting. I would guess that this would get better with time"*).

Quantitative Results are Inconclusive

Quantitative data from joystick and head movement did not differ significantly between conditions. Using joystick movement as a proxy for when users react to potentially dangerously approaching cars, we found users who viewed videos without our system reacted to a potentially unsafely approaching car in ($\mu = 1.01$ s$, \sigma = 0.48$ s), compared to users who used HindSight ($\mu = 1.04$ s$\sigma = 0.74$ s). The 30 millisecond average difference corresponds to the duration of a single video frame.

One possible explanation for the inconclusive results is that the interpretations of "potential danger" was too subjective. Some users moved the joystick towards parked cars to mark them as "dangerous", while others did not. Future work would need to determine a more objective measure that can be interpreted consistently by participants.

HindSight May Effectively Redirect Attention When Users are Distracted

Although our quantitative results were inconclusive across the set of videos, one datum of interest emerged for a video with a distractor. Near the end of this clip, the bicycle slows down to a stop at an intersection as a truck quickly stops and clears the intersection. In the meantime, a car out of view quickly stops alongside the bicycle from behind. Without HindSight , 0 out of 4 turned towards the approaching car from behind, whereas 7 out of 11 users using HindSight noticed the car, as determined by head orientation data (Figure 10). This suggests that our system may be especially effective at redirecting user attention when they are distracted by other stimuli. This effect should be further investigated in the future.

Feedback for Improving HindSight

Participants also offered suggestions for how they would improve HindSight for use during their real commutes. As is, participants rated the system favorably when asked if they would use it during their real commutes on a 5-point Likert scale ($\mu = 3.87, \sigma = 0.96$). However, comments regarding the audio output and occurrence of false positives suggest avenues for further work.

Explore a Broader Space of Audio Cues

Many users remarked they would like to see revised audio cues: *"I'd want to see some more granularity in the alarm response depending on the seriousness of the danger"* (P11), *"maybe it could use a more distinct effect to denote the severity/distance of the danger"* (P5), *"It could also be a measure of how fast or how big the vehicle approaching is"* (P2). Our current design uses beeps which change in tempo and volume. Investing additional resources in sound experience design could enrich the experience of using our system.

Reduce Incidences of False Detections

Although we designed our filtering stages to reduce our system's false detections, users felt the remaining false positives still impacted usability. *"I heard some false positives from parked cars receeding away from the bicycle"* (P12), *"Sometimes does send misleading beeps (got a few when no car was immediately approaching)"* (P7). Additional signal processing will be needed to further reduce the number of false positives. One proposal for future work is to incorporate the *trajectory* of approaching objects with a dynamics model.

LIMITATIONS

As a prototype system, HindSight has limitations from engineering constraints and the availability of technology. Our exploratory study design also limits the types of claims and generalizations we can currently make.

The resolution of our panoramic camera is relatively low. Output is at 1280x720 at 15 fps streamed live. A rough calculation shows using a 4K panoramic camera could provide twice the detection distance of our 720p camera, increasing users' limited time to react.

Our system requires a 10 lb laptop to be worn. Our laptop was chosen as a solution to balance portability and a high-end GPU. Although it can be comfortably worn in a backpack, it is not an ideal form factor. Developments in low-power, small-footprint hardware designed for neural network computations[3] and considering mobile-optimized neural network architectures [13] will likely address this limitation.

The Orientation Filter can reduce sensitivity to objects approaching from directly behind. The Orientation Filter works effectively in practice because cars commonly approach the bicycle at an offset from the rear. However, objects which are approaching directly from behind may be detected later because their tracked x value does not change. Engineering a dynamics model which estimates the *trajectory* of directly approaching objects could resolve this limitation.

Object tracking does not merge bounding boxes. Our frame to frame tracking algorithm could be improved by adding a step where we merge bounding boxes if items are likely the same at seams of image partitions.

Exploratory study has limited realism. A primary limitation of the study is that participants had no agency to change their trajectory or speed, as they were watching pre-recorded videos. The most externally valid study design would be real-world deployment in live traffic situations. However, before exposing participants to potential risk in traffic, we believe that a simulation study in virtual reality, where use of can be evaluated safely, would be an appropriate next step.

CONCLUSION

We introduced HindSight, a wearable system that increases spatial awareness by detecting relevant objects in live, egocentric 360-degree video and sonifying their attributes through bone conduction headphones. HindSight draws upon advances in computer vision and work in delivering continuous feedback for physical tasks to identify points of interest in a user's surroundings and notify the user when necessary to redirect their attention.

Our analysis suggests that at current detection performance, bicyclists can be notified in time to react to dangers when vehicles travel up 8.6 m/s faster than the cyclists. This margin may be sufficient for many, but not all urban cycling situations. Progress in camera technology and object classification can further improve on this threshold.

In our exploratory study, we find HindSight increased users' reported comfort, awareness, reaction time, and safety, and identify potential avenues for future work, such as reducing the recall rate of object detection and designing broader audio experiences for users.

While our prototype is somewhat limited by the need to wear a laptop with a powerful GPU, multiple hardware companies are currently developing chips can run deep neural networks in real time, which would make a truly portable solution feasible.

Beyond the domain of cycling, we believe that combining an enhanced awareness of visual periphery with the rich semantic understanding of objects and scenes from computer vision techniques has the potential to enable an entire new class of applications that improve on unaided human capabilities.

[3] https://developer.movidius.com/

ACKNOWLEDGMENTS

This work was supported in part by the VeHICaL project (NSF grant #1545126) and by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

REFERENCES

1. Kiomars Anvari. 2017. Helmet with wireless sensor using intelligent main shoulder pad. US Patent No. US9596901. (Mar 2017).

2. Jérôme Ardouin, Anatole Lécuyer, Maud Marchal, Clément Riant, and Eric Marchand. 2012. FlyVIZ: A Novel Display Device to Provide Humans with 360 Vision by Coupling Catadioptric Camera with Hmd. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology (VRST '12)*. ACM, New York, NY, USA, 41–44. DOI: http://dx.doi.org/10.1145/2407336.2407344

3. Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 333–342. DOI: http://dx.doi.org/10.1145/1866029.1866080

4. M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett. 1998. An evaluation of wearable information spaces. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*. 20–27. DOI:http://dx.doi.org/10.1109/VRAIS.1998.658418

5. Blaze. 2017. Blaze: Innovative products for urban cycling. https://blaze.cc/. (2017).

6. Z. Cai, D. G. Richards, M. L. Lenhardt, and A. G. Madsen. 2002. Response of human skull to bone-conducted sound in the audiometric-ultrasonic range. *Int Tinnitus J* 8, 1 (2002), 3–8.

7. Alexandru Dancu, Velko Vechev, Adviye Ayça Ünlüer, Simon Nilson, Oscar Nygren, Simon Eliasson, Jean-Elie Barjonet, Joe Marshall, and Morten Fjeld. 2015. Gesture Bike: Examining Projection Surfaces and Turn Signal Systems for Urban Cycling. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 151–159. DOI:http://dx.doi.org/10.1145/2817721.2817748

8. Frederik Diederichs and Benjamin Fischle. *Advanced Telematics for Enhancing the Safety and Comfort of Motorcycle Riders: HMI Concepts and Strategies*. http://www.saferider-eu.org/assets/docs/deliverables/SAFERIDER_D5_1_HMI_Concepts_and_Strategies.pdf

9. M R Everingham, B T Thomas, and T Troscianko. 1999. Head-mounted mobility aid for low vision using scene classification techniques. *International Journal of Human-Computer Interaction* 15, 2 (1999), 231–244.

10. Kevin Fan, Jochen Huber, Suranga Nanayakkara, and Masahiko Inami. 2014. SpiderVision: Extending the Human Field of View for Augmented Awareness. In *Proceedings of the 5th Augmented Human International Conference (AH '14)*. ACM, New York, NY, USA, Article 49, 8 pages. DOI: http://dx.doi.org/10.1145/2582051.2582100

11. Juan Diego Gomez, Guido Bologna, and Thierry Pun. 2012. Spatial Awareness and Intelligibility for the Blind: Audio-touch Interfaces. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 1529–1534. DOI: http://dx.doi.org/10.1145/2212776.2223667

12. Thomas Hermann, Andy Hunt, and John G. Neuhoff. 2011. *The Sonification Handbook*. Logos Verlag, Berlin.

13. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR* abs/1704.04861 (2017). http://arxiv.org/abs/1704.04861

14. Ian P. Howard and Brian J. Rogers. 1995. *Binocular Vision and Stereopsis*. Oxford University Press, New York.

15. IPPINKA. 2013. Xfire: On-Demand Laser Bike Lane. https://www.ippinka.com/blog/xfire-on-demand-laser-bike-lane/. (2013).

16. L. Kay. 1974. A sonar aid to enhance spatial perception of the blind: engineering design and evaluation. *Radio and Electronic Engineer* 44, 11 (November 1974), 605–627. DOI:http://dx.doi.org/10.1049/ree.1974.0148

17. S. Kerber and H. Fastl. 2008. Prediction of perceptibility of vehicle exterior noise in background noise. In *Tagungsband Fortschritte der Akustik (DAGA '08)*, U. Jekosch and R. Hoffmann (Eds.). DEGA, 623–624.

18. Dagmar Kern and Albrecht Schmidt. 2009. Design Space for Driver-based Automotive User Interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '09)*. ACM, New York, NY, USA, 3–10. DOI: http://dx.doi.org/10.1145/1620509.1620511

19. Vinitha Khambadkar and Eelke Folmer. 2013. GIST: A Gestural Interface for Remote Nonvisual Spatial Perception. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 301–310. DOI: http://dx.doi.org/10.1145/2501988.2502047

20. Scott R. Klemmer, Björn Hartmann, and Leila Takayama. 2006. How Bodies Matter: Five Themes for Interaction Design. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. ACM, New York, NY, USA, 140–149. DOI: http://dx.doi.org/10.1145/1142405.1142429

21. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. 2015. SSD: Single Shot MultiBox Detector. *CoRR* abs/1512.02325 (2015). http://arxiv.org/abs/1512.02325

22. D. Menzel, K. Yamauchi, F. Völk, and F. Fastl. 2011. Psychoacoustic experiments on feasible sound levels of possible warning signals for quiet vehicles. In *Tagungsband Fortschritte der Akustik (DAGA '11)*. DEGA.

23. Nicolas Misdariis and Andrea Cera. 2013. Sound signature of Quiet Vehicles: state of the art and experience feedbacks. In *Inter-Noise*. Innsbruck, Austria. https://hal.archives-ouvertes.fr/hal-01106897

24. Nicolas Misdariis, Andrea Cera, Eugenie Levallois, and Christophe Locqueteau. 2012. Do electric cars have to make noise? An emblematic opportunity for designing sounds and soundscapes. In *Acoustics 2012*, Société Française d'Acoustique (Ed.). Nantes, France. https://hal.archives-ouvertes.fr/hal-00810920

25. Takashi Miyaki and Jun Rekimoto. 2016. LiDARMAN: Reprogramming Reality with Egocentric Laser Depth Scanning. In *ACM SIGGRAPH 2016 Emerging Technologies (SIGGRAPH '16)*. ACM, New York, NY, USA, Article 15, 2 pages. DOI: http://dx.doi.org/10.1145/2929464.2929481

26. M. Mon-Williams, J. P. Wann, and S. Rushton. 1995. Design factors in stereoscopic virtual-reality displays. *Journal of the Society for Information Display* 3, 4 (1995), 207–210. DOI:http://dx.doi.org/10.1889/1.1984970

27. A. Mukhtar, L. Xia, and T. B. Tang. 2015. Vehicle Detection Techniques for Collision Avoidance Systems: A Review. *IEEE Transactions on Intelligent Transportation Systems* 16, 5 (Oct 2015), 2318–2338. DOI:http://dx.doi.org/10.1109/TITS.2015.2409109

28. Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 113–116. DOI: http://dx.doi.org/10.1145/2677199.2680549

29. Paul L. Olson and Michael Sivak. 1986. Perception-Response Time to Unexpected Roadway Hazards. *Human Factors* 28, 1 (1986), 91–96. DOI: http://dx.doi.org/10.1177/001872088602800110

30. Alex Olwal, Jonny Gustafsson, and Christoffer Lindfors. 2008. Spatial augmented reality on industrial CNC-machines. *Proc.SPIE* 6804 (2008), 6804 – 6804 – 9. DOI:http://dx.doi.org/10.1117/12.760960

31. Etienne Parizet, Wolfgang Ellermeier, and Ryan Robart. 2014. Auditory warnings for electric vehicles: Detectability in normal-vision and visually-impaired listeners. *Applied Acoustics* 86, Supplement C (2014), 50–58. DOI: http://dx.doi.org/10.1016/j.apacoust.2014.05.006

32. R. Pea and R. Lindgren. 2008. Video Collaboratories for Research and Education: An Analysis of Collaboration Design Patterns. *IEEE Transactions on Learning Technologies* 1, 4 (Oct 2008), 235–247. DOI: http://dx.doi.org/10.1109/TLT.2009.5

33. Jef Raskin. 2000. *The Humane Interface: New Directions for Designing Interactive Systems*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA.

34. Joseph Redmon and Ali Farhadi. 2016. YOLO9000: Better, Faster, Stronger. *CoRR* abs/1612.08242 (2016). http://arxiv.org/abs/1612.08242

35. Andreas Riener, Myounghoon Jeon, Ignacio Alvarez, and Anna K. Frison. 2017. *Driver in the Loop: Best Practices in Automotive Sensing and Feedback Mechanisms*. Springer International Publishing, Cham, 295–323. DOI: http://dx.doi.org/10.1007/978-3-319-49448-7_11

36. Yong Rui, Anoop Gupta, and J. J. Cadiz. 2001. Viewing Meeting Captured by an Omni-directional Camera. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 450–457. DOI: http://dx.doi.org/10.1145/365024.365311

37. Eldon Schoop, Michelle Nguyen, Daniel Lim, Valkyrie Savage, Sean Follmer, and Björn Hartmann. 2016. Drill Sergeant: Supporting Physical Construction Projects Through an Ecosystem of Augmented Tools. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1607–1614. DOI: http://dx.doi.org/10.1145/2851581.2892429

38. S. Sivaraman and M. M. Trivedi. 2013. Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis. *IEEE Transactions on Intelligent Transportation Systems* 14, 4 (Dec 2013), 1773–1795. DOI: http://dx.doi.org/10.1109/TITS.2013.2266661

39. John P. Snyder. 1987. *Map projections: A working manual*. Technical Report. Washington, D.C. http://pubs.er.usgs.gov/publication/pp1395

40. Hans Strasburger, Ingo Rentschler, and Martin Jüttner. 2011. Peripheral vision and pattern recognition: A review. *Journal of Vision* 11, 5 (2011), 13. DOI: http://dx.doi.org/10.1167/11.5.13

41. Yu-Chuan Su, Dinesh Jayaraman, and Kristen Grauman. 2016. Pano2Vid: Automatic Cinematography for Watching 360° Videos. *CoRR* abs/1612.02335 (2016). http://arxiv.org/abs/1612.02335

42. SKULLY Systems. 2013. SKULLY AR-1 The World's Smartest Motorcycle Helmet. https://www.indiegogo.com/projects/skully-ar-1-the-world-s-smartest-motorcycle-helmet. (2013).

43. Matt Uyttendaele. 2017. Optimizing 360 photos at scale. (Aug 2017). https://code.facebook.com/posts/129055711052260/optimizing-360-photos-at-scale/

44. Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2014. Show and Tell: A Neural Image Caption Generator. *CoRR* abs/1411.4555 (2014). http://arxiv.org/abs/1411.4555

45. H. C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré, and D. Rus. 2017. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 6533–6540. DOI: http://dx.doi.org/10.1109/ICRA.2017.7989772

46. Michael Zöllner, Stephan Huber, Hans-Christian Jetter, and Harald Reiterer. 2011. *NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect*. Springer Berlin Heidelberg, Berlin, Heidelberg, 584–587. DOI: http://dx.doi.org/10.1007/978-3-642-23768-3_88

47. Amit Zoran and Joseph A. Paradiso. 2013. FreeD: A Freehand Digital Sculpting Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2613–2616. DOI: http://dx.doi.org/10.1145/2470654.2481361