ARcoustic: A Mobile Augmented Reality System for Seeing Out-of-View Traffic

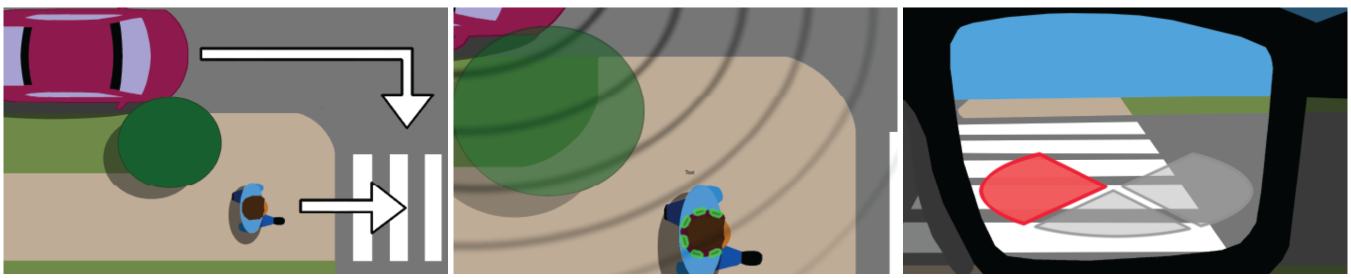
Xuesong Zhang
Department of Computer Science, KU Leuven
Leuven, Belgium
xuesong.zhang@kuleuven.be

Xian Wu
Mecha(tro)nische SysteemDynamica (LMSD), Department of Mechanical Engineering, KU Leuven
Leuven, Belgium
xian.wu1@kuleuven.be

Robbe Cools
Department of Computer Science, KU Leuven
Leuven, Belgium
robbe.cools@kuleuven.be

Adalberto L. Simeone
Department of Computer Science, KU Leuven
Leuven, Belgium
adalberto.simeone@kuleuven.be

Uwe Gruenefeld
Human-Computer Interaction Group, University of Duisburg-Essen
Essen, Germany
uwe.gruenefeld@uni-due.de



(a) Scenario of traffic encounter. **(b) Microphone array for localisation.** **(c) Visualisation of out-of-view vehicle.**

Figure 1: The proposed *ARcoustic* system. The system supports pedestrians in everyday traffic encounters, helping them to avoid potential collisions with other vehicles (a). It utilises a microphone array to localise nearby out-of-view vehicles (b). The localised nearby vehicles are then visualised using a novel visualisation technique (c). *Best seen in colour.*

ABSTRACT

Locating out-of-view vehicles can help pedestrians to avoid critical traffic encounters. Some previous approaches focused solely on visualising out-of-view objects, neglecting their localisation and limitations. Other methods rely on continuous camera-based localisation, raising privacy concerns. Hence, we propose the *ARcoustic* system, which utilises a microphone array for nearby moving vehicle localisation and visualises nearby out-of-view vehicles to support pedestrians. First, we present the implementation of our sonic-based localisation and discuss the current technical limitations. Next, we present a user study (n = 18) in which we compared two state-of-the-art visualisation techniques (Radar3D, CompassbAR) to a baseline without any visualisation. Results show that both techniques present too much information, resulting in below-average user experience and longer response times. Therefore, we introduce a novel visualisation technique that aligns with the technical localisation limitations and meets pedestrians' preferences for effective visualisation, as demonstrated in the second user study (n = 16). Lastly, we conduct a small field study (n = 8) testing our *ARcoustic* system under realistic conditions. Our work shows that out-of-view object visualisations must align with the underlying localisation technology and fit the concrete application scenario.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Virtual reality**; **Visualization techniques**; **User studies**.

KEYWORDS

out-of-view, visualisation, AR, VR, sonic-based localisation

ACM Reference Format:
Xuesong Zhang, Xian Wu, Robbe Cools, Adalberto L. Simeone, and Uwe Gruenefeld. 2023. ARcoustic: A Mobile Augmented Reality System for Seeing Out-of-View Traffic. In *15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '23), September 18–22, 2023, Ingolstadt, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3580585.3606461

AutomotiveUI '23, September 18–22, 2023, Ingolstadt, Germany

ACM ISBN 979-8-4007-0105-4/23/09…$15.00
https://doi.org/10.1145/3580585.3606461

1 INTRODUCTION

Awareness of objects in the environment is important for everyday life. Lack of information about the locations of surrounding objects may lead to severe consequences. A prominent example of this is overlooking other road users when navigating traffic, which can result in collisions. A previous study found that more than half of drivers do not look before turning, which highlights the elevated risk of collisions [37]. Since pedestrians are usually more at risk in such situations, they are commonly referred to as vulnerable road users [51]. Consequently, pedestrians need to be on high alert when navigating traffic to avoid close encounters with vehicles. In this paper, we introduce the *ARcoustic* system, which is designed to help pedestrians in critical traffic encounters by visually highlighting moving vehicles.

A natural way for humans to perceive objects in their surrounding environment is their peripheral vision. However, the human field of view is restricted to about 180 degrees horizontally and 90 degrees vertically [50]. This means that objects outside the field of view cannot be perceived visually. In addition, being distracted (e.g. by a smartphone [52]) further amplifies the problem of not perceiving relevant objects, which ultimately increases the risk of an accident [66]. Besides visual perception, our ability to hear relevant objects approaching can be impacted as well. For example, people with hearing impairment cannot perceive all acoustic environmental noises; thus, they rely more on their visual perception. Another problem is that pedestrians often wear headphones while navigating traffic, which may prevent them from hearing important sounds (e.g. a car honking) [67].

While researchers have proposed various methods to visualise potential risks for pedestrians in traffic (e.g. [29]), the detection connected to these risks (e.g. a nearby vehicle) is often assumed to exist. Thus, no consideration is given to technical limitations resulting from tracking (e.g. tracking inaccuracies). If tracking solutions are proposed, they focus on very specific scenarios, such as a vehicle approaching from behind while a user plays an Augmented Reality (AR) game [34] or encountering fixed obstacles on the sidewalk [53, 71]. Current solutions to track nearby vehicles mostly rely on optical sensors [1, 35]. However, cameras in public spaces are perceived as invasive of privacy [13, 39]. Despite being a promising alternative, sonic-based tracking has not been frequently researched in the traffic context. Furthermore, on the visualisation side, previous work has primarily investigated the visualisation of static out-of-view objects [7, 42], neglecting moving objects. Out-of-view objects moving adds another dimension, increasing the overall complexity of the problem [25]. The degree to which existing solutions scale towards these problems with greater complexity remains unclear. It is likely that the additional complexity results in a greater mental load, making such solutions difficult to use in everyday situations [25]. Finally, it is unclear what information pedestrians actually need to navigate traffic safely.

The primary research objective of this work is to develop a sonic-based solution for the detection of out-of-view traffic in pedestrian environments. Meanwhile, this work aims to explore the potential applicability of existing out-of-view visualisation techniques in conveying this pertinent information to users. In the event that the existing techniques are unsuitable, we design and develop a new visualisation technique to address this requirement.

In our paper, we propose the *ARcoustic* system. It utilises a 360-degree microphone array and a machine-learning approach to track nearby vehicles, estimating their directions and distances relative to the user. The vehicles detected as out-of-view are then visualised for the user to increase their situational awareness. In the early design stage, we employed Virtual Reality (VR) as a test bed. To understand how to visualise the out-of-view road users that are potentially relevant, we first compared different visualisation techniques (n = 18) from previous work (i.e. CompassbAR and Radar3D). As our findings showed that these techniques present too much information to the user, we then designed a novel visualisation technique that takes into account both the user's mental requirements and the technical limitations of our tracking approach. We continued with a second lab study in which we investigated our *ARcoustic* system (n = 16) and a field study (n = 8) in which we tested the system under realistic conditions.

Contributions. The contributions of our paper are constructive and empirical: 1) We contribute the *ARcoustic* system, which detects and visualises out-of-view moving vehicles with great accuracy. 2) We present two lab studies conducted in VR and use them to compare existing out-of-view object visualisation techniques, finding that existing techniques are not well-suited to the task. We then compare our *ARcoustic* to no visualisation, showing that it reduces task load. 3) We perform a field study in which we test the system under realistic conditions using AR glasses.

2 RELATED WORK

In the following, we take a closer look at the related work that informed the design of our *ARcoustic* system.

2.1 Locating Objects in Outdoor Environments

Commonly used technologies for locating and tracking objects are inertial-, optical-, and sonic-based tracking. While inertial-based tracking is limited to relative movements and needs to be applied to the tracked object itself [74], optical- and sonic-based tracking appear more feasible for tracking nearby road users. Their advantage is that they enable tracking from the user's position without the need for communication with the tracked road users (e.g. to communicate locations tracked individually by GPS). Optical tracking is often implemented using an RGB camera in combination with a tracking algorithm (e.g. YOLO [59], which allows real-time tracking from individual picture frames). In previous work, researchers have applied optical tracking in traffic for collision avoidance with fixed structures [36] and nearby vehicles [34]. Here, one camera is sufficient for position estimation [1], and capturing multiple frames allows estimation of velocity [35]. However, optical tracking is perceived as privacy-invasive by bystanders; hence, it is less feasible for public contexts [13, 39]. A possible solution to privacy-invasive optical tracking can be sonic-based tracking. In previous work, researchers have deployed microphones to monitor traffic [45]. While a microphone could capture sensitive information or identify people [12], doing so would require bystanders to speak in close proximity to it, making it a good alternative to a continuously recording camera [46]. In our work, we track nearby vehicles using a microphone array; thereby, we estimate their locations relative to the user.

Researchers have explored the use of microphone arrays for estimating the positions of vehicles in outdoor environments. Approaches include passive acoustic perception using time-difference-of-arrival (TDOA) techniques, as proposed by Schulz et al. [63], and active acoustic perception using emitted sound waves and measured echoes [33]. More recently, data-driven approaches have also been proposed for object detection and tracking using multichannel acoustic signals. Gan et al. proposed a self-supervised system to track moving vehicles using stereo sound [21]. Valverde et al. went further by incorporating a microphone array for estimating vehicle positions in outdoor environments, as well as using depth and thermal imaging in addition to RGB for training their system [69]. These papers collectively demonstrate the potential of microphone arrays for vehicle position estimation in outdoor environments. Hence, we follow this approach in our work.

2.2 Visual Guidance in Traffic

To help people navigate everyday traffic, researchers have proposed various approaches that either help road users find a target (i.e. navigation) or avoid a target (i.e. collision avoidance). For navigation, previous work has investigated navigation with AR to a single target (e.g. finding the correct bus stop among many others [49]) or to multiple targets (e.g. showing multiple points of interest during sightseeing [62]). Nevertheless, in traffic encounters, road users often face situations in which they need to avoid a target instead of navigating toward it. Here, researchers focused on car drivers and either provided them with a visualisation of surrounding objects [27] or directly guided their attention to critical objects [14]. For autonomous vehicles, previous work has also proposed the use of external interfaces on the vehicles to reassure pedestrians that they have not been overlooked [9, 15, 16]. However, only a few papers have investigated systems that can directly support pedestrians in avoiding collisions. For example, Wang et al. proposed a pedestrian safety app for mobile phone users who walk and talk while crossing roads [71]. Moreover, Gruenefeld et al. suggested peripheral LEDs to shift the attention of smartphone users to potential traffic collisions [27]. While these systems can support pedestrians in avoiding collisions, they put a stronger emphasis on users distracted by their smartphones. More importantly, they directly shift the attention of users to hazardous objects, which means that they must be able to tell which objects are actually hazardous. Since they are not yet able to do so, they either shift the user's attention too frequently or they miss relevant objects (with potentially fatal consequences). Therefore, in our work, we want to visualise relevant objects outside the user's field of view and empower the user to identify hazardous objects on their own (similar to [34]).

2.3 Visualising Moving Out-of-View Objects

One obvious solution to empower humans to see out-of-view objects is to extend their vision to 360 degrees, removing the limit of the human field of view (e.g. by using 360-degree cameras [41]). However, researchers have shown that this leads to disorientation and can overstimulate the human brain [41]. Thus, previous work has initially focused on encoding only the directions toward out-of-view objects. A common approach uses LEDs mounted on glasses to encode direction [29, 47, 55]. This idea has also been explored with VR glasses [28, 75]. As most of the LED-based approaches distribute the LEDs around the user's eyes, they do not directly encode the direction of the object; instead, they encode the direction of the head movement necessary to bring the object into view. However, as most relevant objects in a traffic scenario are placed on a 2D ground plane, only LEDs on the left and right sides of the eyes can be used to encode out-of-view objects. Thus, multiple objects are more difficult to encode and tend to overlap with each other (similar to the corner-density problem of off-screen visualisation techniques such as Halo [5]). A better approach to encode direction is to use on-screen AR visualisations [18, 23]. For example, Gruenefeld et al. transferred well-known off-screen visualisation techniques such as Halo and Wedge to AR [23]. However, the large visual cues are difficult to show on small-screen devices. CompassbAR overcomes this limitation by using a 2D bar positioned at the top of the screen, where each position on the 2D bar encodes a direction (from left -180 degrees to right 180 degrees) [18].

For our work, we selected CompassbAR to encode the locations of relevant out-of-view road users. Besides direction, there are visualisation techniques that encode not only directions but also distances to out-of-view objects [7, 24–26, 42]. For example, EyeSee360 uses a radar-like visualisation in which every point encodes a direction and the color of every point encodes distance [24]. Previous work has compared EyeSee360 to non-visual approaches and found that it results in faster search times than audio and haptic cues [42]. Moreover, others have compared EyeSee360 to Radar3D, an alternative radar-like visualisation that encodes direction and distance together. It was found that the cognitive load is higher for Radar3D [30], while both techniques perform equally well in terms of several other metrics (e.g. search time) [7]. Nevertheless, it should be noted that most of these studies involved out-of-view objects in static positions. When using moving objects, Gruenefeld et al. found that Radar3D actually results in better performance and understanding of object movement [25]. Thus, for our work, we selected Radar3D as the second visualisation technique to be tested for visualising out-of-view objects in our traffic context.

3 GENERAL APPROACH

In this paper, we propose the *ARcoustic* system, which allows both the localisation of nearby moving vehicles and their visualisation on AR glasses. The system aims to support pedestrians during critical traffic encounters in which a nearby vehicle approaches outside the pedestrian's field of view, potentially resulting in a collision. So far, a few papers have investigated solutions outside the lab that explore localisation and visualisation together (e.g. [34]). However, each localisation technology introduces technical limitations that need to be considered when designing the visualisation technique. Furthermore, traffic encounters are rather complex situations that involve several moving objects and require decisions to be made quickly, while previous work has primarily investigated support for static objects (e.g. [25]).

To design our *ARcoustic* system, we followed user-centred design (UCD) principles [22], enabling us to iteratively improve our design. The UCD process initiates with a concept derived from an understanding of the usage context; followed by the development of one or more prototypes, which undergo iterative refinement through successive evaluations [6]. Moreover, in the early design stages (first and second study), we employed VR as a test bed to examine the design concept [78]. Rebelo et al. argued that VR enables one to develop realistic virtual environments that come with greater control of the experimental conditions compared to a lab setting [58]. In addition, VR enables researchers to evaluate systems in different contexts [2, 65], including hard-to-replicate or even dangerous contexts, such as pedestrian safety [11, 64]. As illustrated in Figure 2, we first developed a localisation system that utilises a 6-channel microphone array with Raspberry Pi. The device is worn on the user's head, as detailed in Section 4. Then, we compared existing visualisation techniques for displaying out-of-view objects in a lab study (n = 18; see Section 5). As we found that existing techniques do not work well in our selected traffic scenario, we designed a novel visualisation technique (see Section 6). Thereafter, we conducted two user studies: a lab study (n = 16) to demonstrate the usefulness of our novel visualisation technique (see Section 7) and a field study (n = 8) to test the *ARcoustic* system under realistic conditions (see Section 8).

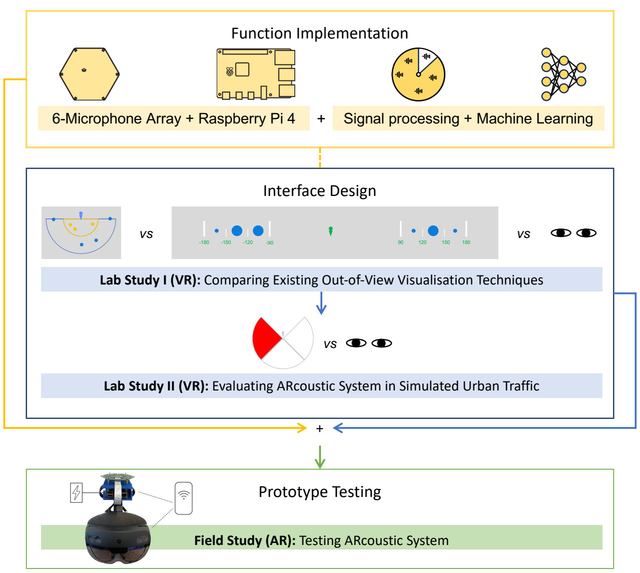


Figure 2: Simplified development process schematic.

4 LOCALISING OBJECTS WITH A MICROPHONE ARRAY

In recent years, sound source localisation has been extensively used as a standard tool for localising sound sources in various applications, including robotics [68], surveillance [61], hearing aids [73] and smart home systems [3]. The primary goal of this process is to determine the location of a sound source, such as a moving vehicle on the road, in real-time and in real-world environments, based on measurements acquired from an array of microphones. The challenges in achieving this objective include dealing with environmental noise, reflections, and varying distances between the sound source and the microphone array [61]. Previous studies have demonstrated that by leveraging microphone arrays, it is possible to develop advanced systems that can accurately identify and locate various sound sources, even under such challenging conditions.

4.1 Concept and Implementation

To address the challenges of real-time sound source localisation and object classification, this work proposes a tailored system that integrates a compact and cost-effective hardware setup with signal processing and machine learning algorithms. The hardware setup, comprising a Raspberry Pi 4 and a ReSpeaker 6-Mic Circular Array kit [60], forms a lightweight and affordable platform with limited computational power, that is suitable for practical deployment. The ReSpeaker 6-Mic Circular Array kit is an extension board for Raspberry Pi. It contains a circular microphone array with six individual microphones to capture multichannel audio signals.

As illustrated in Figure 3, we employ a hybrid approach that combines delay and sum beamforming with a Multi-Layer Perceptron (MLP) neural network. Beamforming is a widely used technique in which the parameters of each element in a phased array are adjusted to enhance signals coming from specific angles while suppressing signals from other angles. This enables the output signal to be steered towards a desired direction, effectively forming a 'beam' [10]. In this work, we use the frequency domain delay and sum beamforming method in the preprocessing stage, a popular foundation for many advanced algorithms [44]. The preprocessing involves transforming multichannel raw signals into the frequency domain using Fast Fourier Transform (FFT) with a sampling rate of 16 kHz and an FFT length of 2048. This results in a frame of sound with a duration of 128 ms. The frequency range of 1.5 kHz to 4 kHz is then selected, as it captures the air-pumping noise from the tire-road noise [77], which is suitable for detecting both Internal Combustion Engine (ICE) vehicles and Electrical Vehicles (EV) and provides an acceptable lobe width given the array geometry. Beamforming computes every 10 degrees from 0 to 360 degrees, resulting in 36 directions. Each direction contains 384 FFT bins, which make one frame with 384 data points. These points are normalised by their mean value.

In addition to the beamforming technique, MLP plays a crucial role in the proposed system by performing sound event classification. In this work, the MLP architecture is designed with a hierarchical structure consisting of an input layer with 384 neurons, followed by three hidden layers containing 240, 120, and 60 neurons, respectively, and finally an output layer with two neurons. This configuration takes the 384 FFT bins of the 1.5 kHz to 4 kHz input frequency range from the output of the beamforming process and classifies the measured noise from a certain direction into two classes: *car* and *other*.

In order to further refine the classification results, a cascade threshold is introduced based on the empirical observation of the mean energy values of the detected sound sources. Specifically, when the output class is identified as a car, the system examines the

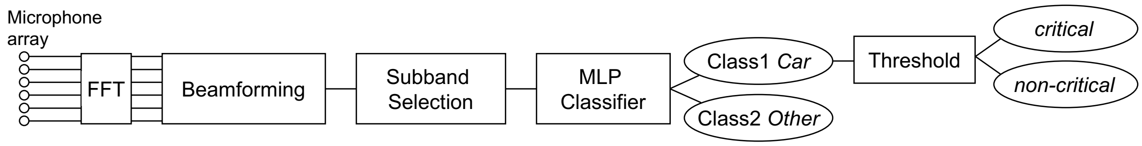


Figure 3: Schematic representation of the overall system

mean energy value of the signal. This threshold, set at 1.0, corresponds to the mean energy value observed for a car located approximately 40 m away. This cascade threshold is grounded through Fu et al.'s pedestrian safety model [19]. The threshold value originates from the braking distance equation, which considers a 50 km/h test scenario, a maximum reaction time of two seconds, and a friction coefficient of 0.8 for a dry asphalt road. The pedestrian safety distance is determined to be 40 m, on which the cascade threshold is based, allowing the system to differentiate between near and far cars and categorise car noise as safety *critical* or *non-critical*.

4.2 Data Collection and Training

The data collection process was carried out in a real-world environment in order to capture audio recordings representative of various acoustic scenarios. The dataset consists of two distinct classes: *car* and *other*. The car class includes sound recordings of vehicles driving at approximately 50 km/h, as well as those with lower speeds when cars stopped at traffic lights. The *other* class contains audio samples of human voices, ventilation noise from indoor office spaces, and outdoor noise in the absence of cars.

To collect the audio data, the researchers positioned themselves near a road where cars travelled at various speeds, including the target speed of 50 km/h. The recordings were initiated manually upon visual identification of an approaching car and terminated once the car was no longer audible. This method, although intuitive and flexible, may introduce some inaccuracies due to its reliance on human judgement and timing. Nevertheless, we believe that the collected dataset provides a reasonably accurate representation of the two designated classes for the purposes of training the MLP neural network. In total, the dataset comprises approximately 10 minutes of *car* recordings and 20 minutes of *other* recordings.

For preprocessing, the Welch method with a Hanning window and 50% overlap was applied to the multi-channel data to compute spectrograms [72]. Then, beamforming was applied to obtain 36-direction signals. For the *car* class, the top six directions with the highest mean energy were assumed to be within the lobe. For the *other* class, the top three directions were considered sufficient due to the lower level of activity. This procedure resulted in a dataset comprising 50k frames for each class.

Once the data were preprocessed, they were stratified and partitioned into training, validation, and test sets, with proportions of 70%, 20%, and 10%, respectively. The training process leveraged the previously introduced MLP neural network. The model was trained using the Generalised Cross-Entropy Loss as the loss function, with an initial learning rate of 0.01 and a weight decay of 1e-5, for a total of 100 epochs.

4.3 Technical Evaluation

The MLP model's performance was evaluated after 100 epochs, as summarised in Table 1. The test F1 score achieved by the model was 0.7806, suggesting that the model's performance is satisfactory, though not optimal. The precision and recall values offer further insight into the performance of the model. In the *other* class, the precision and recall are 0.80 and 0.73, respectively, indicating that the model is more adept at identifying non-car sounds, albeit with some misclassification as car sounds. Conversely, the precision is 0.77 and the recall is 0.83 for the *car* class, indicating that the model is better at identifying car sounds but might classify some non-car sounds as cars.

Table 1: Classification performance of the MLP model

Class	Precision	Recall	F1 score	Support
Car	0.77	0.83	0.80	5260
Other	0.80	0.73	0.76	4789

5 LAB STUDY I: COMPARING EXISTING OUT-OF-VIEW VISUALISATION TECHNIQUES

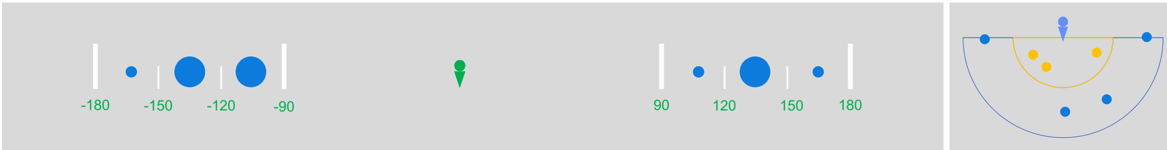
In the following, we report on our first user study in which we compare existing out-of-view visualisation techniques for our explored traffic scenario.

5.1 Study Design

To compare the different visualisation techniques for out-of-view objects, we conducted a within-subjects controlled laboratory study in VR with the Meta Quest 2. Our independent variable was the localisation *technique* with three levels (*Baseline* [no visualisation] vs. *Radar3D* [25] vs. *CompassbAR* [18]; see Figure 4).

The 180-degree rear view is divided into seven directions in both systems, each spaced 30 degrees apart and capable of detecting the closest car within its range. The systems distinguish between two levels of proximity: *critical* (less than 40 m) and *non-critical* (between 40 m and 80 m). *CompassbAR* employs a half linear compass bar to display out-of-view vehicle information, with *non-critical* level cars appearing smaller and *critical* cars appearing larger. *Radar3D* utilises blue and yellow dots to indicate the position of the nearest car in each direction. A yellow dot on the inner ring indicates *critical* proximity, while a blue dot on the outer ring signifies *non-critical* proximity. According to the original design, the CompassbAR is attached to the top of the field of view (FoV), while the Radar3D is placed in the middle of the FoV.

Each technique was tested in a block consisting of four measured trials with the participants crossing the road in each trial, resulting



(a) CompassbAR. **(b) Radar3D.**

Figure 4: Existing out-of-view visualisation techniques we used in Study 1. (*Note: The figure background is grey here for clarity, but it is transparent in the studies.*)

in a total of 216 trials (3 *conditions* × 4 *trials* × 18 *participants*). All blocks, each consisting of one technique, were counterbalanced using a Latin-square design. We used quantitative methods to evaluate user performance. We included time to cross (TTC) as well as subjective measures such as task load (NASA Raw-TLX [31]), user experience (short User Experience Questionnaire [UEQ] [40]), usability (System Usability Scale [SUS] [8]), and individual Likert-items as our dependent variables. For this study, we asked: **(RQ) To what extent can existing out-of-view visualisation techniques (*CompassbAR, Radar3D*) support users in crossing scenarios compared to no visualisation?**

- H_1 For the time to completion, we expect the *Baseline* condition to result in short times because we hypothesise that the existing out-of-view visualisation techniques contain too much information.
- H_2 For the same reason, we expect the *Baseline* condition to result in a lower task load.

5.2 Demographics

A total of 18 volunteers (nine female, nine male) with an average age of 26.17 years (SD = 2.94 years) were recruited for the study through internal mailing lists, word-of-mouth and social media. The study was conducted in our 5.5 m × 9 m VR Lab and was approved by the Ethical Review Board of the institution. Only two participants had no prior experience with VR. There were no colour-blind volunteers. Eleven participants reported having used headphones/earphones while crossing the street, while all participants reported having observed others using headphones/earphones crossing the road.

5.3 Apparatus

This VR project was implemented with Unity 3D (2021.3.16f1). A mesh collider was used to detect the position and distance of each vehicle. We updated the visual information every 0.3 seconds to simulate the physical prototype performance. There is no auditory feedback in the simulation to emulate the use of headphones in a real-world scenario.

5.4 Procedure

The procedure for the study began with obtaining informed consent from each participant and having them fill out a demographic questionnaire. Participants were then briefly informed about the procedure and asked to put on a headset. They were given the opportunity to become familiar with the first user interface before being asked to navigate to six different points, making four times safe crossings with the help of the visualisation (if exists). They walked a straight line distance of 6.30 m each time. Once the task was completed, participants were asked to remove the headset and fill out the UEQ (*CompassbAR*, *Radar3D* only), SUS (*CompassbAR*, *Radar3D* only), NASA Raw-TLX and custom questionnaires. The same task was then repeated using the other two conditions. Finally, an interview was conducted with each participant to gather information on their preferences and overall experience with the two visualisations. The order in which the participants tested the user interfaces was counterbalanced.

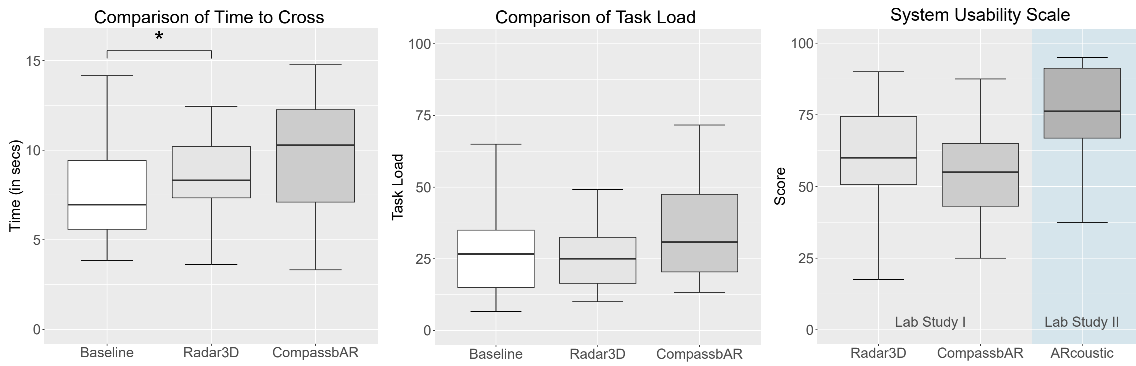
5.5 Results

Time to Cross. For all conditions, participants were asked to cross the road four times. Here, we used the median time of all four crossings. The median (inter-quartile range) times to cross for each condition are (in ascending order): *Baseline* = 7.10 *s* (IQR = 5.17 *s*), *Radar3D* = 9.29 *s* (IQR = 4.49 *s*), and *CompassbAR* = 11.14 *s* (IQR = 5.91 *s*). All times are compared in Figure 5a. A Shapiro-Wilk Test showed that the data is not normally distributed ($p < 0.001$), so we applied non-parametric tests. A posthoc test using Wilcoxon Signed-rank with Bonferroni-Holm correction showed significant differences between *Baseline* and *Radar3D* ($W = 17$, $Z = -2.44$, $p = 0.037$, $r = 0.41$). We can conclude that participants were significantly faster in the *Baseline* condition than with *Radar3D*.

Task Load. For all three conditions, participants were asked to answer the NASA Raw-TLX, for which a lower score indicates a lower task load. The scores are compared in Figure 5b. The median scores are 26.67 (IQR = 20.00) for *Baseline*, 25.00 (IQR = 16.04) for *Radar3D*, and 30.83 (IQR = 27.08) for *CompassbAR*. A Friedman test did not reveal any significant differences ($\chi^2(2) = 4.33$, $p = 0.115$, $N = 18$).

Usability. To assess the usability of both visualisations, we asked participants to complete the SUS. All scores are compared in Figure 5c. The median scores are 60.00 (IQR = 23.75) for *Radar3D* and 55.00 (IQR = 21.88) for *CompassbAR*, indicating below average but okay usability [4].

User Experience. For both conditions with visualisation, we asked participants to fill out the short UEQ. For pragmatic quality, the median scores are 0.75 (IQR = 1.69) for *Radar3D* and 0.50 (IQR = 1.69) for *CompassbAR*. For hedonic quality, the median scores are 0.88 (IQR = 1.38) for *Radar3D* and 1.38 (IQR = 1.62) for *CompassbAR*. Overall, the user experience was rated with 0.69 (IQR = 1.44) for *Radar3D* and 0.75 (IQR = 1.38) for *CompassbAR*. According to Laugwitz et al. [40], the responses for both visualisations indicate below-average user experience. A Wilcoxon signed-rank



(a) Time to cross. **(b) NASA Raw-TLX.** **(c) SUS across lab studies.**

Figure 5: Results from the first user study in which existing out-of-view visualisation techniques (*CompassbAR, Radar3D*) are compared to a baseline without any visualisation support. In subfigure (c), we compare the usability across Studies 1 and 2.

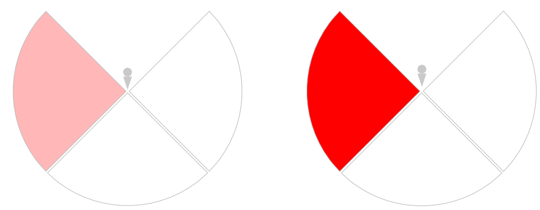
test did not reveal a significant difference between the conditions ($W = 63$, $Z = -0.64$, $p = 0.538$).

Individual Likert-Items. After each condition, we presented participants with various statements and asked them to rate them on a 5-point Likert scale (1=strongly disagree, 3=neutral, 5=strongly agree). The first statement was *Q1: 'I felt safe during the condition.'* Overall, participants agreed for *Baseline* and *Radar3D* ($Md = 4$, $IQR = 1$), while they were neutral for *CompassbAR* ($Md = 2.5$, $IQR = 2$). The second statement was *Q2: 'I felt the overview of the traffic is good.'* Here, participants agreed for *Baseline* ($Md = 4$, $IQR = 1.75$), *Radar3D* ($Md = 4$, $IQR = 2$), and *CompassbAR* ($Md = 4$, $IQR = 1$). Friedman tests did not reveal any significant differences (Q1: $\chi^2(2) = 2.05$, $p = 0.358$, $N = 18$; Q2:$\chi^2(2) = 1.42$, $p = 0.491$, $N = 18$).

Qualitative Feedback. In the semi-structured interview, ten participants preferred the *Radar3D*, as this visualisation is easy to understand and effective at offering a traffic overview, while eight preferred the *CompassbAR*, as it is reminiscent of games. Seven participants stated that they interpreted the information from *CompassbAR* as left and right, enabling them to make quick decisions. Overall, participants wished for a larger field of view (FoV) and suggested that the system should only display information that requires immediate attention instead of also presenting non-critical traffic information.

6 DESIGNING THE ARCOUSTIC VISUALISATION TECHNIQUE

To enhance the visualisation of out-of-view information, we incorporated the participants' feedback in the design process. Our goal was to minimise the user's workload in acquiring information, leading to the development of the new *ARcoustic* visualisation shown in Figure 9. We expanded the out-of-view range to include the left front and right front, each at a 45-degree angle, while the three quarter-turns represent the user's left, back, and right sides, respectively. *ARcoustic* is specifically designed to display only the most critical direction to the user at any given moment, reducing information overload and improving the user's ability to process important information.



(a) Example: *non-critical* car on the left side of the user. **(b) Example: *critical* car on the left side of the user.**

Figure 6: Simple schematic demonstrating different states of the *ARcoustic* visualisation

We used red in the visualisation to indicate approaching cars. We chose this colour because red is associated with concepts such as *potential hazard* and *danger* [48, 56]; thus, it communicates that the user should treat the nearby moving car as such. While green is associated with it being 'safe' to cross, we opted not to use it according to knowledge from information visualisation that indicated it could lower the usability of the system for colour-blind users [43]. Rather, we opted to use different saturation levels of red: 0 saturation for no car, 0.5 for a *non-critical* car (Figure 6a) and 1 for a car that is *critical* (Figure 6b).

7 LAB STUDY II: EVALUATING ARCOUSTIC IN SIMULATED URBAN TRAFFIC

We report on our second user study in which we evaluated the new out-of-view *ARcoustic* visualisation technique. We used the same apparatus as in the first study (Section 5).

7.1 Study Design

We designed and conducted a within-subjects user study. We aimed to investigate whether the presence of a visualisation system enhances the quality of decision-making compared to a condition where no such system is present. Our independent variable was the localisation *technique* with two levels (*Baseline* [no visualisation] vs. *ARcoustic*).

Each technique was tested in a block consisting of 12 measured trials. In each trial, participants were asked to make a decision about whether it was safe to cross the road. Overall, we had 384 trials (2 $conditions \times 12\ trials \times 16\ participants$). We used quantitative methods to evaluate user performance. We took the number of safe crossing decisions as well as subjective measures such as confidence in the decision, task load (NASA Raw-TLX), user experience (short UEQ), usability (SUS), and individual Likert-items as our dependent variables. For this study, we asked: **(RQ) Can the new visualisation technique (*ARcoustic*) support users in crossing scenarios compared to no visualisation?**

- H_3 We expect the *ARcoustic* condition to result in more correct decisions because we hypothesise that the new out-of-view visualisation technique will help users to detect dangers quickly and easily.
- H_4 For the same reason, we expect the *ARcoustic* condition to result in a reduced task load.

7.2 Demographics

Sixteen new volunteers (four female, 12 male) with an average age of 28.38 years (SD = 3.03 years) were recruited in this study. Of these, five had no prior experience with VR. The recruitment procedures were consistent with those of the first study (Section 5). Overall, 15 participants used headphones/earphones while crossing the street, and 12 used smartphones/smartwatches while doing so. All participants reported having seen others using headphones/earphones or smartphones/smartwatches when crossing the road. None of the participants had colour blindness.

7.3 Procedure

After giving informed consent and filling out a demographic questionnaire, participants were briefly informed about the procedure and asked to put on a headset. They first familiarised themselves with the VR visualisation system as well as the messages they could receive during the observation. During the study, participants were informed that there are different traffic situations and were tasked with observing them on the roadside from the beginning. Between the 3 and 8-second intervals, messages were displayed in front of the users. After an additional 1.5 seconds of observation, they were required to make a decision regarding the safety of crossing the road and to reply to a question about the message that they had previously received. Then they observed the traffic situations once with the visualisation and once without in a counterbalanced order. There are 12 different traffic situations, in which car movements and states vary. These include left or right-hand drive in one or two directions as well as scenarios with parked cars that may or may not be movable.

Once the task was completed, participants were asked to remove the headset and fill out the UEQ (*ARcoustic* only), SUS (*ARcoustic* only), NASA-TLX and custom questionnaires. The same task was then repeated with another condition. Finally, we interviewed each participant to gather information on their opinion of and overall experience with the new user interface. The order in which the participant tested the user interfaces was counterbalanced.

7.4 Results

Decision to Cross. For each condition, we conducted 12 trials. Within each trial, we first exposed participants to a traffic situation and then asked them if they thought it was safe to cross the road and how confident they were in their answers (5-point Likert scale). If they answered with the wrong decision, we inverted the confidence score, using it as a negative value. Thereafter, we calculated the weighted decisions from the mean of all confidence scores for that condition. The mean (IQR) weighted decisions are 1.25 (IQR = 3.67) for *Baseline* and 1.29 (IQR = 3.65) for *ARcoustic*. A Wilcoxon test did not reveal a significant difference between the conditions ($W = 58$, $Z = 0.35$, $p = 0.749$). In terms of the number of correct decisions, when traffic conditions were unsafe, participants using *ARcoustic* made more correct decisions ($MEAN = 5.09$, $SD = 1.87$) (not crossing the road) than those using *Baseline* ($MEAN = 4.18$, $SD = 2.32$), as tested by a Wilcoxon test ($W = 2.5$, $Z = -1.98$, $p = 0.047$).

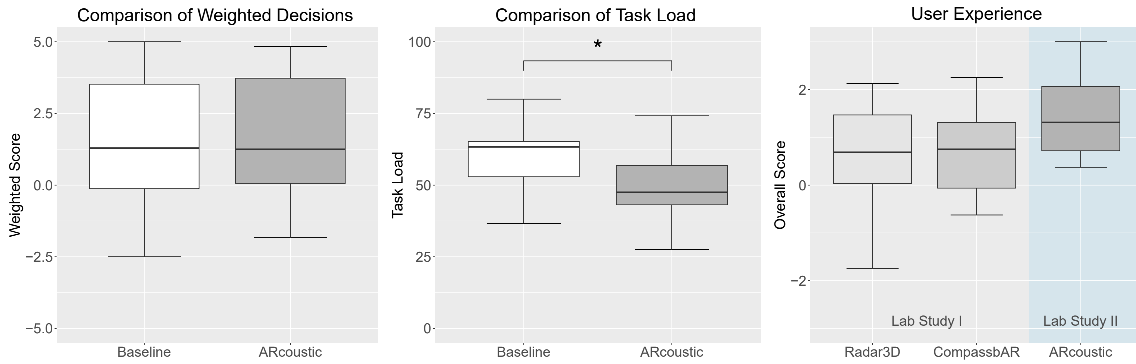
Task Load. For both conditions, participants were asked to answer the NASA Raw-TLX. The resulting score for *ARcoustic* ($Md = 47.50$, $IQR = 13.75$) is significantly lower than for the *Baseline* condition without a visualisation ($Md = 63.33$, $IQR = 12.29$) ($W = 97.5$, $Z = 2.13$, $p = 0.032$, $r = 0.38$).

Usability. To assess the usability of *ARcoustic*, we asked participants to complete the SUS after using *ARcoustic*. The median resulting score is 76.25 (IQR = 24.38), which indicates good usability according to Bangor et al. [4].

User Experience. We asked participants to fill out the short UEQ after using *ARcoustic*. For pragmatic quality, participants rated *ARcoustic* with a median score of 1.63 (IQR = 1.94) for pragmatic quality; and a median score of 1.50 (IQR = 2.06) for hedonic quality. Overall, the user experience was rated with a median score of 1.31 ($IQR = 1.34$). According to Laugwitz et al. [40], the participants' responses for *ARcoustic* indicate good user experience.

Individual Likert-Items. After each condition, we presented participants with two statements and asked them to rate them on a 5-point Likert scale , as in the first Study (Section 5). The responses can be seen in Figure 8. Overall, participants agreed with the first statement for *ARcoustic* ($Md = 4.0$, $IQR = 0.5$), while they were neutral for *Baseline* ($Md = 3.0, IQR = 2.0$). A Wilcoxon signed-rank test revealed a significant difference between the conditions with a medium effect size ($W = 6$, $Z = -2.44$, $p = 0.016$, $r = 0.43$). We can conclude that participants felt significantly safer in the *ARcoustic* condition than in the *Baseline*. Participants were neutral with the second statement for the *Baseline* condition ($Md = 3.0$, $IQR = 2.0$), while they agreed for the *ARcoustic* condition ($Md = 4.0$, $IQR = 2.0$); however, the difference between the conditions is not significant ($W = 13.5$, $Z = -1.75$, $p = 0.094$).

Qualitative Feedback. Overall, the participants described ARcoustic as having a *simple* and *intuitive* user interface that is *easy to understand* (as mentioned by nine participants). With ARcoustic, users feel safer and are able to obtain a brief summary of their surroundings even when they are focused on a secondary task. This increased awareness can provide users with more confidence when making decisions and help them to rotate their heads less frequently. Four participants expressed a particular liking for the use of red in the middle of FoV to indicate danger, while others



(a) Weighted decision to cross. (b) NASA Raw-TLX. (c) UEQ across lab studies.

Figure 7: Results from the second lab study in which the novel *ARcoustic* visualisation is compared to the baseline without visualisation. In subfigure (c), we plot the user experience across the first and second studies.

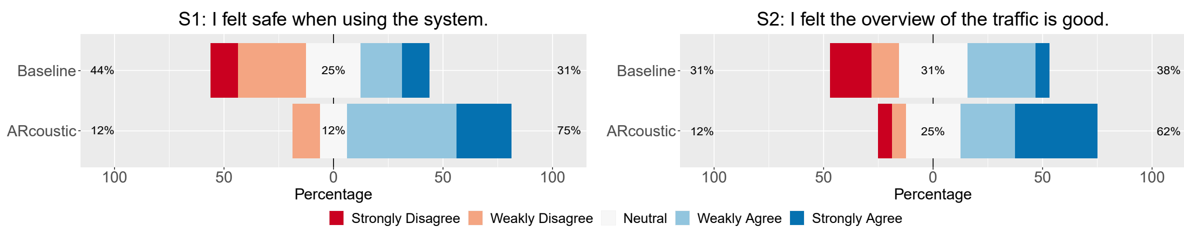


Figure 8: After each condition, participants were asked to rate two statements on a 5-point Likert scale.

felt that the visual feedback could be distracting and suggested incorporating peripheral indicators. Furthermore, five participants expressed the desire to receive information about the movement direction of cars in their vicinity, in addition to simply being alerted to their presence.

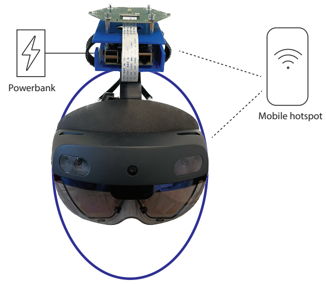
8 FIELD STUDY OF THE ARCOUSTIC SYSTEM

To validate the feasibility of the ARcoustic system, we conducted a field study and asked participants to wear it mounted on their heads. The system was implemented using Unity 3D (2021.3.2f1) with the Mixed Reality Toolkit package. A HoloLens 2 was used to project information to the user. A Raspberry Pi 4 was connected to a mobile hot spot to transmit the data via TCP/IP socket and was powered by an additional power bank, as shown in Figure 9a. To correctly display the colour with the desired visual effect, we adjusted it manually.

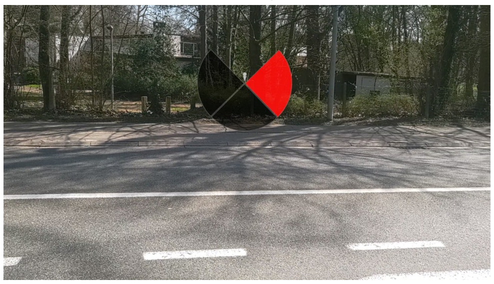
Participants. Eight new volunteers (three female, five male) with an average age of 26 years (SD = 2.51 years) participated in this field study. Everyone had had experience with mixed reality. One participant had never used headphones/earphones or a smartphone/smartwatch while crossing the road. All participants had observed others using them while crossing. None of them are colour-blind. The recruitment procedures were consistent with those of the first study (Section 5). The study was carried out on the side of Celestijnenlaan in Leuven, Belgium.

Procedure. After obtaining consent and collecting demographic information, we directed the participants to the starting point on the pavement. They were then instructed to proceed slowly along the 100 m pavement towards a bus stop, a journey that typically takes around two minutes. Participants then crossed a bike lane and remained at the bus station for a minimum of three minutes while at least five cars passed by. Subsequently, participants returned to the starting point. Throughout the study, we used the Think-Aloud Protocol, which required participants to verbalise their thoughts and observations continuously [17]. At the end of the study, participants were asked to complete the UEQ, NASA Raw-TLX, and a custom questionnaire, followed by a semi-structured interview.

Results. Participants evaluated the system as positive (overall score of 1.23) with the UEQ questionnaire. In general, observing the traffic with ARcoustic has a medium task load ($MEAN$ = 23.54, SD = 9.76). Participants' feedback regarding the visualisation in this study was consistent with that of the second study. They found the interface to be *easy*, *simple*, and *intuitive*, allowing them to quickly identify the directions of moving vehicles. Additionally, participants noted that the ARcoustic device was lighter and more portable than they had anticipated. They also expressed confidence in the system's accuracy, as they were able to observe that the corresponding quadrant changed colour when a car passed by. However, it is important to note that the colour red may not always be easily seen in bright outdoor environments, which could impact the system's sensitivity for some users. The participants expressed distinct preferences regarding how the information pertaining to their position should be updated. One participant preferred that out-of-view information be updated based on eye movement, while



(a) Hardware setting of *ARcoustic*.



(b) The view with the HoloLens 2.

Figure 9: Hardware setting and the screenshot of *ARcoustic* on the HoloLens 2. (*Note: black will appear as transparent in b).*

the other participants preferred that the information be based on body position rather than head movement. In terms of hardware performance, we noticed that strong winds tend to cause false positive *non-critical* occurrences.

9 GENERAL DISCUSSION

In this section, we discuss the result we obtain from the technical evaluation and the three studies we conducted.

Object Localisation. From a technical evaluation perspective, the relatively high F1 score for the *car* class suggests that the model is well-suited for safety-critical applications, as it can effectively distinguish between car and non-car sounds. Still, there is potential for enhancing the model's performance. Future work could investigate more advanced models, such as 1D CNNs [38] or Transformers [70], to improve performance. Considering hardware constraints and real-time processing requirements, the current MLP model is deemed adequate. The processing time for a single frame, including beamforming and classification across 36 directions, is 50 ms on a Raspberry Pi 4 using a single CPU thread, which enables the implementation of real-time applications. Moreover, future research could explore more fine-grained labelling or employ self-supervised techniques such as cross-modality training [21, 69]. Utilising such techniques could potentially leverage the co-occurrence of visual and audio streams in unlabelled videos, without the need to collect ground truth annotations.

Existing Out-of-View Visualisations. The result partially supported H_1 and H_2. Existing out-of-view visualisation techniques (CompassbAR, Radar3D) are not perfectly suitable for supporting users in crossing scenarios compared to no visualisation. This was mainly due to information overload and user distrust of the system.

Information Overload. While most participants acknowledged the usefulness of the system in certain situations, they also noted that the existing out-of-view visualisations provide an overwhelming amount of information. Both systems we tested require time to learn and interpret data, which is a critical issue when crossing a road. Participants who are familiar with shooting games are accustomed to the CompassbAR. Other participants may require more time to learn how to map the linear compass bar to the world coordinate system. Consequently, many users would opt not to use the system in these situations. In general, detailed information for all detected cars is not necessary for this scenario. Too much out-of-view information can distract users and interrupt their focus. While the interruptions serve to alert users, they also increase the TCC, since users check the alert information and then return to what they were focusing on before.

User Distrust of the System. Despite the system's ability to accurately compute and display real-time information, users often encounter a discrepancy when attempting to locate a car's current position. As pedestrians, participants perceived the car to be moving at a relatively high speed. However, by the time they checked, the car may have moved from its previously displayed location, eroding user trust in the system. Furthermore, based on the hardware performance of the Raspberry Pi 4, it is currently only possible to present a moving car as a discrete dot without a continuous display. Continuous tracking visualisation of the same vehicle is not feasible with the current technological implementation, which differs from existing optical tracking systems used in autonomous driving (e.g. overlaying graphics and data onto a live video feed from the vehicle's cameras). This is not the same as they experienced or expected, which further decreases the system's usability. As a result, participants did not fully trust the system. However, the ability to track a car is not an essential function of our system.

ARcoustic Visualisation. The new *ARcoustic* visualisation is rated as having good usability as well as good user experience with improved and simplified visualisation. The results partially supported H_3. A technical limitation of the current system is that it only provides the direction of the moving car and does not indicate whether the car is approaching or moving away from the user. In the following four street conditions, participants may misunderstand the indication, since a leaving car can also trigger a danger indicator: 1) left-steering vehicles approaching each other, 2) right-steering vehicles approaching each other, and 3-4) previously stopped vehicles, both left-steering or right-steering, moving in opposite directions. Moreover, unlike in the first study, participants did not have enough time to observe the traffic after reading the received message or map. We observed that participants employed different strategies when the system detected a *critical* level of danger: those who exhibited trust in the system promptly refrained from crossing upon receiving a danger message, while those who preferred to verify the system's information before making a decision chose not to cross until they had turned their head to check for danger. In such cases, they rated their decisions with low confidence. This behaviour is in line with daily experience, as rotating one's head to check for traffic before crossing without a traffic signal is common [57]. In the field study, participants expressed fewer concerns about understanding the moving direction of the car. Traffic flow in real-world conditions is less intense than in the virtual environment, allowing participants to better assess the safety of their surroundings.

The result from the user study supports H_4. *ARcoustic* summarises the out-of-view information so that users can perceive potential dangers directly. By incorporating an additional danger detector, users experienced a significant increase in their sense of safety compared to the baseline. *ARcoustic* reduces task load by providing information in the middle of the user's FoV, so it can be quickly and easily perceived with just a glance. This eliminates the need for users to rotate their heads to check their surroundings, enabling them to identify potential dangers more efficiently. However, due to the high difficulty level of the tasks themselves, the NASA Raw-TLX values for ARcoustic are still classified as *somewhat high*.

During the field study, participants reported difficulty in perceiving colours with high levels of brightness outdoors. To address this issue, we recommend that future systems enable the AR device to detect ambient brightness and automatically adjust the brightness of the AR projection and the colours accordingly to improve visibility and user experience.

Using VR to Design AR Systems in the early design stage. VR offers a controlled and secure test bed for the evaluation of potentially hazardous scenarios [2, 58, 65]. In our study, we employed VR to simulate an AR system within a virtual urban environment featuring traffic. Through the implementation of experiments within the VR environment, we were able to directly introduce users to the system concept, gather their feedback, and promptly refine the design, ultimately resulting in the development of an improved visualisation. Notably, during the real-world AR field study, users provided feedback that closely aligned with the observations made in the VR simulation. While VR studies provide a valuable representation of the acoustic interface, it is crucial to acknowledge certain limitations inherent to this approach.

As we transitioned this system from VR to real-world AR, we noted that the colour patterns designed for VR could not be directly applied to AR. Consequently, we adjusted the colour and depth of the UI manually. It is worth emphasising that various factors, including the surrounding environment [20, 32] and display technology [76], can cause VR and AR to present the same colour with varying visual effects. Moreover, due to differences in depth perception between AR and VR [54], it is necessary to reevaluate the desired depth of the user interface to ensure optimal usability and user experience.

Limitations. We used sonic-based technology in our system, which is capable of detecting the sound of a moving vehicle behind a blind corner [63]. However, we did not evaluate such scenarios in the VR study and field tests.

10 CONCLUSION AND IMPLICATIONS

In this work, we proposed the *ARcoustic* system, which offers a solution for visualising out-of-view information for pedestrians. The proposed system employs a machine learning-based approach and utilises a Raspberry Pi 4 and a ReSpeaker 6-Mic Circular Array kit for localisation. The traffic information is projected onto a HoloLens 2, and the visualisation is first designed in virtual reality and then tested in real-world settings. The findings of this study demonstrate that this new visualisation technique effectively reduces information overload, assists users in identifying potential hazards with ease and efficiency, and facilitates accurate decision-making in unsafe situations. The participants' feedback also reveals a favourable attitude towards the system.

ACKNOWLEDGMENTS

This research is supported by *Internal Funds KU Leuven* (HFG-D8312-C14/20/078), and the European Commission through the *H2020 ETN ECO DRIVE project (Marie Skłodowska-Curie program)* (GA 858018).

REFERENCES

[1] Abduladhem Abdulkareem Ali and Hussein Alaa Hussein. 2016. Distance estimation and vehicle position detection based on monocular camera. In *2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*. 1–4. https://doi.org/10.1109/AIC-MITCSA.2016.7759904

[2] Tatsuya Amano, Shugo Kajita, Hirozumi Yamaguchi, Teruo Higashino, and Mineo Takai. 2018. Smartphone Applications Testbed Using Virtual Reality. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (New York, NY, USA) *(MobiQuitous '18)*. Association for Computing Machinery, New York, NY, USA, 422–431. https://doi.org/10.1145/3286978.3287028

[3] Mingsian R. Bai, Shih-Syuan Lan, Jong-Yi Huang, Yi-Cheng Hsu, and Hing-Cheung So. 2020. Audio Enhancement and Intelligent Classification of Household Sound Events Using a Sparsely Deployed Array. *The Journal of the Acoustical Society of America* 147, 1 (2020), 11–24. https://doi.org/10.1121/10.0000492

[4] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[5] Patrick Baudisch and Ruth Rosenholtz. 2003. Halo: A Technique for Visualizing off-Screen Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 481–488. https://doi.org/10.1145/642611.642695

[6] D. Benyon. 2019. *Designing User Experience*. Pearson Educación.

[7] Felix Bork, Christian Schnelzer, Ulrich Eck, and Nassir Navab. 2018. Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2983–2992. https://doi.org/10.1109/TVCG.2018.2868584

[8] John Brooke. 1996. SUS: A 'quick and dirty' usability scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, Bruce Thomas, Bernhard A. Weerdmeester, and Ian L. McClelland (Eds.). CRC Press, London, UK, Chapter 21, 189–194.

[9] Chia-Ming Chang, Koki Toda, Xinyue Gui, Stela H. Seo, and Takeo Igarashi. 2022. Can Eyes on a Car Reduce Traffic Accidents?. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 349–359. https://doi.org/10.1145/3543174.3546841

[10] Paolo Chiariotti, Milena Martarelli, and Paolo Castellini. 2019. Acoustic beamforming for noise source localization–Reviews, methodology and applications. *Mechanical Systems and Signal Processing* 120 (2019), 422–448.

[11] Shuchisnigdha Deb, Daniel W. Carruth, Richard Sween, Lesley Strawderman, and Teena M. Garrison. 2017. Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics* 65 (2017), 449 – 460. https://doi.org/10.1016/j.apergo.2017.03.007

[12] Najim Dehak, Patrick J. Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet. 2011. Front-End Factor Analysis for Speaker Verification. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 4 (2011), 788–798. https://doi.org/10.1109/TASL.2010.2064307

[13] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-Mediating Technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2377–2386. https://doi.org/10.1145/2556288.2557352

[14] Henrik Detjen, Sarah Faltaous, Jonas Keppel, Marvin Prochazka, Uwe Gruenefeld, Shadan Sadeghian, and Stefan Schneegass. 2022. Investigating the Influence of Gaze- and Context-Adaptive Head-up Displays on Take-Over Requests. In *Proceedings of the 14th International Conference on Automotive User Interfaces and*

Interactive Vehicular Applications (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 108–118. https://doi.org/10.1145/3543174.3546089

[15] Debargha Dey, Azra Habibovic, Melanie Berger, Devanshi Bansal, Raymond H. Cuijpers, and Marieke Martens. 2022. Investigating the Need for Explicit Communication of Non-Yielding Intent through a Slow-Pulsing Light Band (SPLB) EHMI in AV-Pedestrian Interaction. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 307–318. https://doi.org/10.1145/3543174.3546086

[16] Debargha Dey, Kai Holländer, Melanie Berger, Berry Eggen, Marieke Martens, Bastian Pfleging, and Jacques Terken. 2020. Distance-Dependent EHMIs for the Interaction Between Automated Vehicles and Pedestrians. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 192–204. https://doi.org/10.1145/3409120.3410642

[17] K Anders Ericsson and Herbert A Simon. 1984. *Protocol analysis: Verbal reports as data.* the MIT Press.

[18] Alessandro Evangelista, Vito Modesto Manghisi, Francesco Laera, Michele Gattullo, Antonio Emmanuele Uva, and Michele Fiorentino. 2022. CompassbAR: A Technique for Visualizing Out-of-View Objects in a Mixed Reality Environment. In *Design Tools and Methods in Industrial Engineering II*, Caterina Rizzi, Francesca Campana, Michele Bici, Francesco Gherardini, Tommaso Ingrassia, and Paolo Cicconi (Eds.). Springer International Publishing, Cham, 141–148.

[19] Ting Fu, Luis Miranda-Moreno, and Nicolas Saunier. 2018. A novel framework to evaluate pedestrian safety at non-signalized locations. *Accident Analysis & Prevention* 111 (2018), 23–33.

[20] Joseph L Gabbard, J Edward Swan, and Adam Zarger. 2013. Color blending in outdoor optical see-through AR: The effect of real-world backgrounds on user interface color. In *2013 IEEE Virtual Reality (VR)*. IEEE, 157–158.

[21] Chuang Gan, Hang Zhao, Peihao Chen, David Cox, and Antonio Torralba. 2019. Self-Supervised Moving Vehicle Tracking with Stereo Sound. arXiv:arXiv:1910.11760

[22] Jeff Gothelf. 2013. *Lean UX: Applying lean principles to improve user experience.* " O'Reilly Media, Inc.".

[23] Uwe Gruenefeld, Abdallah El Ali, Susanne Boll, and Wilko Heuten. 2018. Beyond Halo and Wedge: Visualizing out-of-View Objects on Head-Mounted Virtual and Augmented Reality Devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 40, 11 pages. https://doi.org/10.1145/3229434.3229438

[24] Uwe Gruenefeld, Dag Ennenga, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2017. EyeSee360: Designing a Visualization Technique for out-of-View Objects in Head-Mounted Augmented Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 109–118. https://doi.org/10.1145/3131277.3132175

[25] Uwe Gruenefeld, Ilja Koethe, Daniel Lange, Sebastian Weiß, and Wilko Heuten. 2019. Comparing Techniques for Visualizing Moving Out-of-View Objects in Head-mounted Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 742–746. https://doi.org/10.1109/VR.2019.8797725

[26] Uwe Gruenefeld, Daniel Lange, Lasse Hammer, Susanne Boll, and Wilko Heuten. 2018. FlyingARrow: Pointing Towards Out-of-View Objects on Augmented Reality Devices. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays* (Munich, Germany) *(PerDis '18)*. Association for Computing Machinery, New York, NY, USA, Article 20, 6 pages. https://doi.org/10.1145/3205873.3205881

[27] Uwe Gruenefeld, Andreas Löcken, Yvonne Brueck, Susanne Boll, and Wilko Heuten. 2018. Where to Look: Exploring Peripheral Cues for Shifting Attention to Spatially Distributed Out-of-View Objects. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) *(AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 221–228. https://doi.org/10.1145/3239060.3239080

[28] Uwe Gruenefeld, Tim Claudius Stratmann, Abdallah El Ali, Susanne Boll, and Wilko Heuten. 2018. RadialLight: Exploring Radial Peripheral LEDs for Directional Cues in Head-Mounted Displays. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 39, 6 pages. https://doi.org/10.1145/3229434.3229437

[29] Uwe Gruenefeld, Tim Claudius Stratmann, Jinki Jung, Hyeopwoo Lee, Jeehye Choi, Abhilasha Nanda, and Wilko Heuten. 2018. Guiding Smombies: Augmenting Peripheral Vision with Low-Cost Glasses to Shift the Attention of Smartphone Users. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 127–131. https://doi.org/10.1109/ISMAR-Adjunct.2018.00050

[30] Yuki Harada and Junji Ohyama. 2022. Quantitative evaluation of visual guidance effects for 360-degree directions. *Virtual Reality* 26, 2 (2022), 759–770.

[31] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[32] Nargess Hassani and Michael J Murdoch. 2016. Color appearance modeling in augmented reality. In *Proceedings of the ACM Symposium on Applied Perception*. 132–132.

[33] Alberto Izquierdo, Lara del Val, and Juan J. Villacorta. 2021. Feasibility of Using a MEMS Microphone Array for Pedestrian Detection in an Autonomous Emergency Braking System. *Sensors* 21, 12 (2021), 4162. https://doi.org/10.3390/s21124162

[34] Jinki Jung, Hyeopwoo Lee, Jeehye Choi, Abhilasha Nanda, Uwe Gruenefeld, Tim Stratmann, and Wilko Heuten. 2018. Ensuring Safety in Augmented Reality from Trade-off Between Immersion and Situation Awareness. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 70–79. https://doi.org/10.1109/ISMAR.2018.00032

[35] Moritz Kampelmühler, Michael G Müller, and Christoph Feichtenhofer. 2018. Camera-based vehicle velocity estimation from monocular video. *arXiv preprint arXiv:1802.07094* (2018).

[36] HyeongYeop Kang, Geonsun Lee, and JungHyun Han. 2019. Obstacle Detection and Alert System for Smartphone AR Users. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 11 pages. https://doi.org/10.1145/3359996.3364256

[37] Nazli Eser Kaya, Suzan Ayas, Canmanie Teresa Ponnambalam, and Birsen Donmez. 2018. Visual attention failures during turns at intersections: An on-road study. *Victoria* 1 (2018).

[38] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J Inman. 2021. 1D convolutional neural networks and applications: A survey. *Mechanical systems and signal processing* 151 (2021), 107398.

[39] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don't Look at Me That Way! Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 362–372. https://doi.org/10.1145/2785830.2785842

[40] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, Springer, Berlin, Germany, 63–76.

[41] Feng Liang, Stevanus Kevin, Holger Baldauf, Kai Kunze, and Yun Suen Pai. 2020. OmniView: An Exploratory Study of 360 Degree Vision Using Dynamic Distortion Based on Direction of Interest. In *Proceedings of the Augmented Humans International Conference* (Kaiserslautern, Germany) *(AHs '20)*. Association for Computing Machinery, New York, NY, USA, Article 5, 10 pages. https://doi.org/10.1145/3384657.3384796

[42] Alexander Marquardt, Christina Trepkowski, Tom David Eibich, Jens Maiero, Ernst Kruijff, and Johannes Schöning. 2020. Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3389–3401. https://doi.org/10.1109/TVCG.2020.3023605

[43] Riccardo Mazza. 2009. *Introduction to Information Visualization*. Springer, London. https://doi.org/10.1007/978-1-84800-219-7

[44] Roberto Merino-Martínez, Pieter Sijtsma, Mirjam Snellen, Thomas Ahlefeldt, Jérôme Antoni, Christopher J. Bahr, Daniel Blacodon, Daniel Ernst, Arthur Finez, Stefan Funke, Thomas F. Geyer, Stefan Haxter, Gert Herold, X. Huang, William M. Humphreys, Quentin Leclère, A.M.N. Malgoezar, Ulf Michel, Thomas Padois, A. Pereira, Christophe Picard, Ennes Sarradj, Henri A. Siller, Dick G. Simons, and Carsten Spehr. 2019. A review of acoustic imaging methods using phased microphone arrays. *CEAS Aeronautical Journal* 10 (2019), 197–230.

[45] Prashanth Mohan, Venkata N. Padmanabhan, and Ramachandran Ramjee. 2008. Nericell: Rich Monitoring of Road and Traffic Conditions Using Mobile Smartphones. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems* (Raleigh, NC, USA) *(SenSys '08)*. Association for Computing Machinery, New York, NY, USA, 323–336. https://doi.org/10.1145/1460412.1460444

[46] Vivian Genaro Motti and Kelly Caine. 2015. Users' Privacy Concerns About Wearables. In *Financial Cryptography and Data Security*, Michael Brenner, Nicolas Christin, Benjamin Johnson, and Kurt Rohloff (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 231–244.

[47] Takuro Nakao, Masashi Nakatani, Liwei Chan, and Kai Kunze. 2016. Smart Glasses with a Peripheral Vision Display. In *Proceedings of the 2016 Virtual Reality International Conference* (Laval, France) *(VRIC '16)*. Association for Computing Machinery, New York, NY, USA, Article 16, 3 pages. https://doi.org/10.1145/2927929.2927938

[48] Annie W.Y. Ng and Alan H.S. Chan. 2018. Color associations among designers and non-designers for common warning and operation concepts. *Applied Ergonomics* 70 (2018), 18–25. https://doi.org/10.1016/j.apergo.2018.02.004

[49] Antti OI Nurminen and Konsta Mikael Sirvio. 2021. Bus Stop Spotting: A Field Experiment Comparing 2D Maps, Augmented Reality and 3D Maps. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse and Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 37, 14 pages. https://doi.org/10.1145/

3447526.3472051
[50] James F Parker Jr and Vita R West. 1973. Bioastronautics Data Book: NASA SP-3006. *NASA Special Publication* 3006 (1973).
[51] Margie Peden, Richard Scurfield, David Sleet, C Mathers, Eva Jarawan, AA Hyder, D Mohan, AA Hyder, and E Jarawan. 2004. *World report on road traffic injury prevention*. World Health Organization.
[52] Dalibor Pešić, Boris Antić, Draženko Glavić, and Marina Milenković. 2016. The effects of mobile phone use on pedestrian crossing behaviour at unsignalized intersections–Models for predicting unsafe pedestrians behaviour. *Safety science* 82 (2016), 1–8.
[53] Max Pfeiffer, Tim Dünte, Stefan Schneegass, Florian Alt, and Michael Rohs. 2015. Cruise Control for Pedestrians: Controlling Walking Direction Using Electrical Muscle Stimulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2505–2514. https://doi.org/10.1145/2702123.2702190
[54] Jiamin Ping, Yue Liu, and Dongdong Weng. 2019. Comparison in depth perception between virtual reality and augmented reality systems. In *2019 IEEE conference on virtual reality and 3d user interfaces (VR)*. IEEE, 1124–1125.
[55] Benjamin Poppinga, Niels Henze, Jutta Fortmann, Wilko Heuten, and Susanne Boll. 2012. AmbiGlasses-Information in the Periphery of the Visual Field.. In *Mensch & Computer*, Vol. 12. 153–162.
[56] Karyn Pravossoudovitch, Francois Cury, Steve G Young, and Andrew J Elliot. 2014. Is red the colour of danger? Testing an implicit red–danger association. *Ergonomics* 57, 4 (2014), 503–510.
[57] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles* 3, 1 (2017), 61–70.
[58] Francisco Rebelo, Paulo Noriega, Emília Duarte, and Marcelo Soares. 2012. Using Virtual Reality to Assess User Experience. *Human Factors* 54, 6 (2012), 964–982. https://doi.org/10.1177/0018720812465006
[59] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 779–788. https://doi.org/10.1109/CVPR.2016.91
[60] ReSpeaker. 2023. *ReSpeaker 6-Mic Circular Array Kit for Raspberry Pi | Seeed Studio Wiki*. https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/ (Accessed on 04/19/2023).
[61] Simone Scardapane, Michele Scarpiniti, Marta Bucciarelli, Fabiola Colone, Marcello Vincenzo Mansueto, and Raffaele Parisi. 2015. Microphone Array Based Classification for Security Monitoring in Unstructured Environments. *AEU - International Journal of Electronics and Communications* 69, 11 (2015), 1715–1723. https://doi.org/10.1016/j.aeue.2015.08.007
[62] Torben Schinke, Niels Henze, and Susanne Boll. 2010. Visualization of Off-Screen Objects in Mobile Augmented Reality. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 313–316. https://doi.org/10.1145/1851600.1851655
[63] Yannick Schulz, Avinash Kini Mattar, Thomas M. Hehn, and Julian F. P. Kooij. 2021. Hearing What You Cannot See: Acoustic Vehicle Detection Around Corners. *IEEE Robotics and Automation Letters* 6, 2 (2021), 2587–2594. https://doi.org/10.1109/LRA.2021.3062254 arXiv:2007.15739
[64] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings Annual Symposium on Computer-Human Interaction in Play* (Amsterdam, The Netherlands) *(CHIPlay '17)*. ACM, New York, NY, USA, 507–515. https://doi.org/10.1145/3116595.3116596
[65] Adalberto L. Simeone, Robbe Cools, Stan Depuydt, João Maria Gomes, Piet Goris, Joseph Grocott, Augusto Esteves, and Kathrin Gerling. 2022. Immersive Speculative Enactments: Bringing Future Scenarios and Technology to Life Using Virtual Reality. In *Proceedings of the SIGCHI Conference on Human Factors (CHI)*. ACM, 1–20. https://doi.org/10.1145/3491102.3517492
[66] Despina Stavrinos, Katherine W. Byington, and David C. Schwebel. 2011. Distracted walking: Cell phones increase injury risk for college pedestrians. *Journal of Safety Research* 42, 2 (2011), 101–107. https://doi.org/10.1016/j.jsr.2011.01.004
[67] Leah L Thompson, Frederick P Rivara, Rajiv C Ayyagari, and Beth E Ebel. 2013. Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. *Injury prevention* 19, 4 (2013), 232–237.
[68] J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau. 2003. Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, Vol. 2. IEEE, Las Vegas, NV, USA, 1228–1233. https://doi.org/10.1109/IROS.2003.1248813
[69] Francisco Rivera Valverde, Juana Valeria Hurtado, and Abhinav Valada. 2021. There Is More than Meets the Eye: Self-Supervised Multi-Object Detection and Tracking with Sound by Distilling Multimodal Knowledge. arXiv:arXiv:2103.01353
[70] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
[71] Tianyu Wang, Giuseppe Cardone, Antonio Corradi, Lorenzo Torresani, and Andrew T. Campbell. 2012. 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* (San Diego, California) *(HotMobile '12)*. Association for Computing Machinery, New York, NY, USA, Article 5, 6 pages. https://doi.org/10.1145/2162081.2162089
[72] Peter Welch. 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15, 2 (1967), 70–73.
[73] B. Widrow. 2001. A Microphone Array for Hearing Aids. *IEEE Circuits and Systems Magazine* 1, 2 (2001), 26–32. https://doi.org/10.1109/7384.938976
[74] Oliver J. Woodman. 2007. *An introduction to inertial navigation*. Technical Report UCAM-CL-TR-696. University of Cambridge, Computer Laboratory. https://doi.org/10.48456/tr-696
[75] Robert Xiao and Hrvoje Benko. 2016. Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1221–1232. https://doi.org/10.1145/2858036.2858212
[76] Jianghao Xiong, En-Lin Hsiang, Ziqian He, Tao Zhan, and Shin-Tson Wu. 2021. Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light: Science & Applications* 10, 1 (2021), 216.
[77] Peter Zeller et al. 2009. *Handbuch Fahrzeugakustik*. Vol. 2. Springer.
[78] Xuesong Zhang and Adalberto L Simeone. 2022. Using the Think Aloud Protocol in an Immersive Virtual Reality Evaluation of a Virtual Twin. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*. 1–8.