The Effect of Visual and Auditory Modality Mismatching between Distraction and Warning on Pedestrian Street Crossing Behavior

Renjie Wu*
The University of Adelaide

Hsiang-Ting Chen†
The University of Adelaide

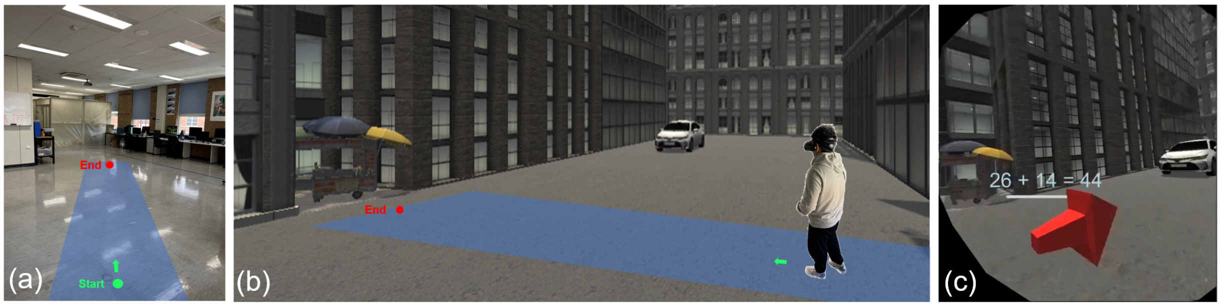


Figure 1: Our experiment investigated the effect of modality mismatching between distraction and warning on pedestrian street crossing behaviour. (a) Physical environment of our VR experiment. (b) The virtual environment of street crossing scenario. (c) The first-person view (right eye) showing a condition with the visual distraction (math equation) and visual warning (red arrow).

Abstract

Augmented reality (AR) headsets could provide useful information to users, but they may also be a source of distraction. Previous works have explored using AR to enhance pedestrian safety by providing real-time warnings, but there has been little research on the impact of modality matching between distractions and warnings on pedestrian street crossing behaviour. We conducted a VR experiment using a within-subjects 2-by-2 design (N = 24) with four conditions: (auditory distraction, visual distraction) $\times$ (auditory warning, visual warning). When experienced conditions with mismatched modalities, participants exhibited more cautious street crossing behaviours, such as reduced walking speed, and increased scan range after receiving the warning, and significantly faster reaction times to the incoming vehicle. The participants also expressed a preference for warnings to be presented in a modality different from the distraction. Our findings suggest that in the context of utilizing AR for pedestrian road safety, future AR interfaces should incorporate a warning modality that differs from the one causing distraction.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 Introduction

In Australia, between 2009 and 2018, 1,711 pedestrian fatalities and over 30,000 hospitalizations due to injuries sustained in road accidents were recorded [11]. These alarming statistics underline the growing concern over pedestrian safety in the era of mobile technology. The issue of pedestrian distraction may be further exacerbated by the rapid advancement of head-mounted displays (HMDs) and extended reality technologies. As HMDs become more convenient and portable, they may replace smartphones as the primary device for social interactions and media consumption. However, like smartphones, HMDs also compete for users' limited attentional resources, which are divided between virtual and real environments. Studies have shown that AR application usage increases the risk of pedestrian hazards, such as colliding with obstacles or being unaware of approaching vehicles [51].

Mitigating the dangers of pedestrian distraction caused by mobile phones and other digital devices in urban settings has emerged as an important research topic for the human-computer interaction community. Previous studies have explored the possibilities of using various sensory channels, such as visual [27], audio [6], haptic [14], galvanic vestibular stimulation [37], and electric muscle stimulation [45], to alert users to potential hazards in their surroundings or provide feedback for navigation. Beyond pedestrians, previous research has also investigated the use of warnings for different groups such as pedestrians, drivers, and cyclists [30, 32, 39, 40]. However, there remains a knowledge gap in the literature regarding the relationship between the modality of distraction and the modality of warning. For instance, if a user is engaged in an auditory activity, such as listening to a podcast, while crossing the street, it remains unclear whether visual or auditory warnings would be more effective in capturing their attention. Note that incorporating multiple sensory channels simultaneously in warning systems might not always be an optimal solution, as it could lead to sensory overload, further diminishing the user's ability to process and react to critical information in a timely manner [64].

In this paper, we evaluated whether the matching of modalities between distractions and warnings affected street-crossing behavior. As the first attempt in this research topic, we chose the two most commonly used modalities, i.e., visual and audio, as well as simple warning interfaces, i.e., an arrow and a beeping sound. We conducted a Virtual Reality (VR) experiment with 24 participants (one dropped out halfway due to severe motion illness), using a within-subject 2-by-2 design with two independent variables: *distraction modality* and *warning modality*. The experiment had four conditions: *AA* (auditory distraction and auditory warning), *AV* (auditory distraction and visual warning), *VV*, (visual distraction and visual warning), *VA* (visual distraction and auditory warning). The scenario was designed to simulate a pedestrian wearing an AR headset walking across a street while engaged with virtual content displayed visually

*e-mail: renjie.wu@adelaide.edu.au
†e-mail: tim.chen@adelaide.edu.au

(visual distraction) or absorbed in auditory content (auditory distraction). In each trial, the participant walked across a street (Fig. 1) while completing a secondary distraction task of answering a series of mathematical questions (in visual or auditory form). During the crossing, a virtual vehicle approached from either left or right (Fig. 1b), and the participant received either a visual or auditory warning (Fig. 1c). The trial ended when the participant noticed the oncoming vehicle (by pressing the trigger on the controller), was hit by the virtual vehicle or reached the opposite side of the street (the virtual vehicle does not appear in the controlled trials). The experimental results demonstrated that conditions with mismatched distraction and warning modalities (*AV* and *VA*) resulted in shorter reaction times, reduced movement speeds and an increased scan range (larger degree of head rotation) following the warning, indicating that participants exhibited more cautious behavior after receiving a warning of an oncoming vehicle. Additionally, the questionnaire and post-hoc interviews suggested that participants favoured warnings in a different modality from the distraction, as this approach led to less confusion and facilitated simpler reactions.

Contribution statement: This paper presents the first VR-based study to systematically investigate the effect of visual and auditory modality matching between distraction and warning on pedestrian street crossing behaviour. The results indicate that using different modalities between warning and distraction leads to shorter reaction times, user's slowing down after warning occurred and larger head rotation range. The participants also preferred using different modalities and found them more effective. The findings suggest that future AR interfaces should use warning modalities that are different from those causing distraction when alerting users about hazards.

2 Related Work

We briefly review the factors that contribute to pedestrian distraction and techniques for improving pedestrian safety.

2.1 Distracted Pedestrians

Before the widespread adoption of smartphones, the majority of road safety research focused on the dangers posed by distracted drivers [32, 43, 58]. However, in recent years, there has been a shift in research emphasis towards the risks associated with distracted pedestrians [59]. This change in focus can be attributed to the increasing prevalence of mobile devices and the evolution of their capabilities, which has led to a significant rise in pedestrian engagement with digital content while on the move.

Some studies [4,33,44] found that pedestrians would change their gait when walking and texting, which may lead to potential safety hazards. Another research [48] found that pedestrians texting or reading while walking resulted in slower speeds and a larger range of head rotation. Schwebel et al. [50] found that pedestrians were more likely to encounter road accidents while texting or listening. Further, Alsaleh et al. [2] showed that pedestrians distracted by texting or listening tended to slow down and control their walking speed by adjusting stride length or cadence. Also, Pedestrians distracted by texting had significantly shorter stride lengths and less stability while walking. Numerous studies have found that engaging in activities such as texting, talking, browsing content, playing AR games on a phone, or being under stress can lead to unsafe street crossing behaviours, such as delayed initiation of street crossing, stepping onto the street before vehicles come to a complete stop, and slower walking pace while crossing the street [5, 51, 69, 70]. These research also suggests that pedestrians may underestimate the risks associated with their distracted behaviour, further exacerbating the problem.

2.2 Warning to Pedestrians

The HCI community has explored several methods to guide users' attention towards critical information and enhance situational awareness, with the goal of improving road safety. Previous works have utilized the smartphone camera [62] and additional custom camera components [67] to detect potential hazards in the surrounding environment. More recently, Hollander et al. [27] developed an app, *SmomDe*, which communicates with vehicles to assist pedestrians crossing the street. Previous studies have investigated the use of AR interfaces as a means to enhance the situational awareness [16]. For a comprehensive systematic review on this topic, we refer readers to Woodward et al. [66]. In the specific context of using AR to improve pedestrian safety, Gruenefeld et al. [21] adapted classical Halo [3] and Wedge [23] out-of-screen visualization approaches for AR HMDs, EyeSee360 [22] designed a radar-like visualization of out-of-view vehicle locations and used colour to represent the distance of approaching vehicles, and Jung et al. [29] integrated a vehicle position estimation system (based on an additional RGB camera) with EyeSee360 and EdgeRadar [24] interface to provide potential collision warnings. Tong et al. [61] also proposed an AR interface that provided information about the direction and potential collision time of oncoming vehicles.

In addition to the AR display, HCI researchers have also explored other modalities for providing warnings to pedestrians, such as haptic feedback from gloves [46], belts [17], and shoes [19,28], and emerging technologies such as electrical muscle stimulation (EMS) for assisting with obstacle avoidance [45]. External Human-Machine-Interfaces (eHMIs) have also received significant attention in recent years, with a series of works exploring the design of eHMIs and the factors associated with pedestrians and street environments [9, 10, 12, 42]. Moreover, studies [7, 35] have found that on-body tactile displays and instructions of the eHMI's rationale can help pedestrians enhance situational awareness and make better decisions. Soret et al. [55] investigated attentional orienting in VR using both endogenous and exogenous cues in auditory and visual modalities, they found that vocal instructions and object highlighting could result in more specific and localized attention. The same team [56] also showed that attentional orienting can improve visual information processing in a 360° immersive environment, where information could be presented in the rear space. Building upon the existing body of research, our work investigated the impact of modality matching between distractions and warnings on pedestrian behaviour during the street crossing.

2.3 Multi-modal Warnings

Previous research also looked into multimodal warnings, as demonstrated in various studies [30, 32, 38, 47]. Marquardt et al. [38] discovered that blending several cues resulted in better outcomes than uni-modal cues. Martin et al. proposed an audio-visual AR approach [47] for warning users of hazardous situations while working with collaborative robots. Jylhä et al. [30] developed a glove with auditory and tactile feedback but did not observe any significant differences compared to a glove with only tactile feedback. Lee et al. [32] employed graded and single-stage haptic-visual and auditory-visual warnings to help drivers avoid sudden braking collisions with the car ahead. Their findings indicated that haptic-visual warnings were preferred, and graded warnings led to larger safe distances and reduced discomfort response rates.

While multimodal warnings have been proven advantageous in some contexts, simultaneously presenting warnings on multiple sensory channels might not always be optimal. Poorly designed warning systems can contribute to sensory overload, overwhelming users with excessive, simultaneous, or conflicting information. For example, Cortez et al. [60] found that exposing VR users to see-through video when they lose their balance, which was thought to help users find better footing, actually impairs the user's ability to regain balance. Haas and van Erp [25] demonstrated that while multimodal displays can improve safety and facilitate risk communication, they may also increase workload and incur extra costs associated with attention switching. When multiple warnings are presented across different

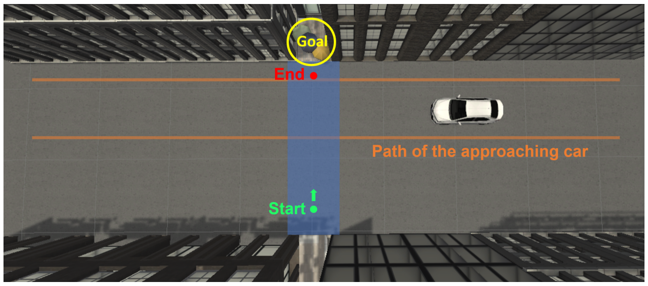


Figure 2: The top view of the virtual environment. The participant crosses the street from the starting point (green dot) toward a street stand (yellow circle). A virtual car would appear between two orange lines, at a random distance between 20 to 45 units away from the participant. The orange lines were not visible to the participants, while the blue stripe was only visible at the beginning of the experiment to remind the participant to walk straight forward.

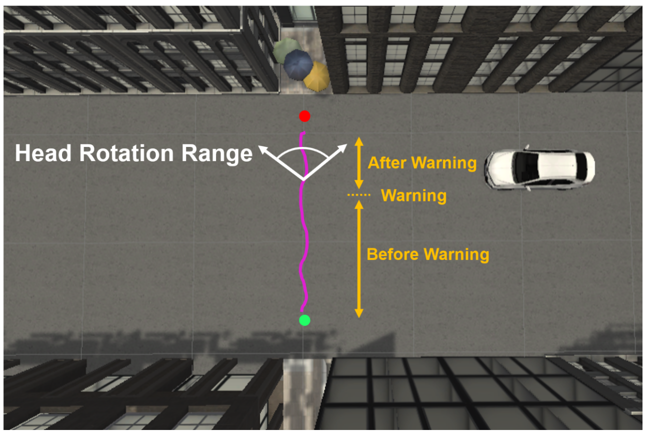


Figure 3: The visualisation of objective measurements. Pink line is the walking path and the yellow line indicates presentation of warning, which splits the trial into before and after warning.

modalities, the user may struggle to process and prioritize them, resulting in reduced attention and cognitive capacity [63, 64]. In sum, despite the valuable insights provided by these previous works, they did not specifically address the effect of modality differences between distraction and warning. Investigating the relationship between the modality of distraction and the modality of warning is crucial for understanding how to design more effective warning systems, as it could reveal the optimal ways to capture users' attention and improve safety. We believe the result presented in this paper would fill the knowledge gap on the impact of modality matching or mismatching on user performance and preferences in the context of pedestrian safety during AR usage.

3 Experiment

We designed and conducted a VR experiment with two independent variables, *distraction modality* and *warning modality*, each with two levels (*auditory* and *visual*) to evaluate whether the difference in sensory channels between distractions and warnings affected the street-crossing behaviour of a user wearing an HMD. More specifically, the experiment used a within-subject 2-by-2 design with four conditions: *AA* (auditory distraction and auditory warning), *AV* (auditory distraction and visual warning), *VV* (visual distraction and visual warning), *VA* (visual distraction and auditory warning).

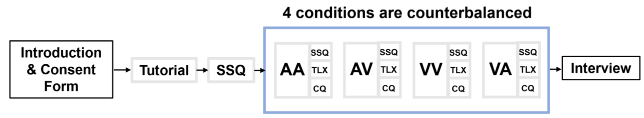


Figure 4: The experiment is divided into four parts that took 20, 15, 40, and 15 minutes each. There are 8 trials in each condition.

3.1 Participants

We recruited 25 participants from the university via social networks and personal contact (7 females and 18 males). Nine participants have used VR headsets before, but none of them used VR headsets frequently. Their ages were between 20 to 30 years (Mean = 23.67 years; SD = 2.71). One participant withdrew from the experiment due to cybersickness and was excluded from the data analysis later. The institute's ethics committee has approved this experiment.

3.2 Apparatus

The VR environment was developed using Unity 2020.3.37f1 and was run on a laptop with an RTX 3080Ti GPU. An HP Reverb G2 Omnicept Edition VR headset was used to experience the environment, which has a Field of View (FoV) of 114 degrees with a resolution of 2160×2160 pixels per eye. The headset weighs 727g and is wired to the VR laptop. The experiment was conducted in an indoor lab that measured 14×7.5 meters, as shown in Fig. 1a.

3.3 Scenario Design

The virtual city scene used in the experiment was shown in Fig. 2. The scenario is designed to mimic a pedestrian wearing an AR headset walking across a street while reading virtual content (visual distraction) or listening to a podcast (auditory distraction). In each trial, the participant walks across the street toward the street stand on the other side while completing a distraction task of evaluating a series of mathematical equations in visual (Fig. 1c) or auditory form. When participants reach an invisible trigger in the middle of the road, a virtual vehicle would appear on either side of the road at a random distance between 25 and 45 meters (Fig. 2). The vehicle moves at a standard speed of 50 km/h and would reach the participant after a random time between 1.8 and 3.2 seconds. The participant would then receive either a visual (Fig. 1c) or an auditory warning when the vehicle appears. Participants were instructed to press the trigger when they noticed the oncoming car. The trial ended when the participant was hit by the car, observed the car and pressed the trigger or reached the opposite side of the street.

3.4 Distraction and Warning Designs

Distraction: We used 2-digit addition and subtraction as the distraction task. After seeing or hearing the mathematical equation, the participant answered whether the equation was correct or incorrect. Fig. 1c shows an example visual mathematical equation of $26 + 14 = 44$. In this example, the participant should press the B button on the controller to indicate the equation is incorrect (button A for the correct). The distraction persisted throughout the entire trial, and the participants were instructed to keep reacting to the secondary distraction tasks. In the *visual distraction* conditions (*VA* and *VV*), the equation was randomly displayed in one of the six cells of the 2-by-3 grid. A progress bar with a duration of 3 seconds was shown beneath the equation to encourage the participant to solve the problem within the allotted time. If the participant provided a correct response, the equation turned light blue (Fig. 1c); if the answer was incorrect, the equation turned red. Each visual distraction persists for roughly 3 seconds. The *auditory distraction* conditions (*AA* and *AV*) involved presenting the math question as an audio clip using the Microsoft Azure text-to-speech API. The length of the audio clip was approximately 2 seconds, after which the participant had a 3-second window to respond. The participants were instructed to engage in a continuous stream of mathematical equations throughout the trial. There have been various types of secondary tasks used to investigate attentional mechanisms through dual-task methodology [41], such as simple reaction time tasks using either visual or auditory stimuli, the Stroop task, verbal memory tasks, n-back task, and math calculation tasks. In our study, we opted for a mathematical calculation task because it has suitable visual and auditory forms, and the user can evaluate the equation via a simple button click.

Warning: The experiment has two warning types: visual warning and auditory warning. The visual warning consisted of a red 3D arrow pointing towards the virtual vehicle, as depicted in Fig. 1c. In the condition of visual distraction and visual warning, the 3D arrow appears below the distraction and does not overlap with the visual distraction. For the auditory warning, we opted for a single beep sound that played in either the left or right ear, depending on the direction of the virtual vehicle. We deliberately avoided using sustained beeping sounds to prevent potential interference with the auditory mathematical equations that we presented to our participants. The volume of the auditory warning is higher than the auditory distraction. Throughout all conditions, we did not stop the distraction when the warning was presented. This design decision was to better simulate real-world situations where visual or auditory content is often continuously played. Additional analysis of the measurements before and after the presentation of the warning stimuli is discussed in the later sections.

3.5 Measurements

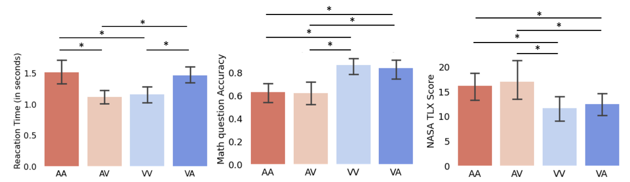
Objective measurements: The objective measurements for analysing participants' crossing street behaviour under four conditions include:

- The *success rate* represents whether the participant successfully notices the virtual vehicle or across the street.
- The *reaction time* is the time between the vehicle's appearance and the time the participant presses the trigger on the controller.
- The *accuracy* is the percentage where the participant correctly evaluates the mathematical equation in the distraction task.
- The *speed* is calculated from the time the participant finishes answering the first math question to the end of the trial. The reason for using this method is to exclude the initial period of the trial when the participant has just started walking.
- The *head rotation range* is calculated based on the orientation angle of the VR headset in relation to the walking direction.

When performing analysis, we further split speed and head rotation range data into *before warning* and *after warning* to understand the participant's street crossing behaviour (Fig. 3).

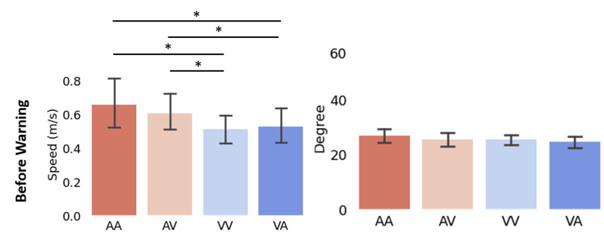
Subjective measurements: The participants were asked to report their level of VR sickness and workload using the Simulator Sickness Questionnaire (SSQ) [31] and NASA TLX [26] scales, respectively, after completing each of the four conditions. They also completed a post-condition questionnaire rating the effectiveness of the warning ("I felt the warning was effective in the current condition.") and the level of distraction ("I was distracted by the mathematical equations in the current condition.") on a 7-point Likert scale (See Appendix for the two figures). At the end of the study, the participants were given a semi-structured interview with three post-study questions:

- **Q1:** Which do you find more distracting - the visual or auditory mathematical equation? Why?
- **Q2:** Which do you find more effective - the visual or auditory warning? Why?

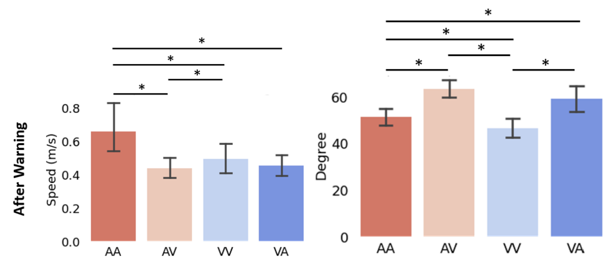


(a) Reaction time (b) Math Accuracy (c) NASA TLX Score

Figure 5: Four conditions of reaction time, math question accuracy and NASA TLX score.



(a) Speed (b) Head Rotation Range



(c) Speed (d) Head Rotation Range

Figure 6: Pairwise comparison about walking speed and head rotation range before and after warning

- **Q3:** Which do you prefer? Having both the distraction and warning happening in the same sensory channel (both auditory or both visual) or in a different sensory channel (one in visual while another auditory)? Why?

3.6 Procedure

The procedure for the experiment is illustrated in Fig. 4. First, the participant was introduced to the study and given a brief overview of the equipment and procedures, then was asked to sign a consent form and complete a demographic questionnaire. Next, the participant received a tutorial on navigating the VR environment using natural walking and instructions on completing their tasks. After the tutorial, the participant completed the first SSQ. In each trial, the participant walked toward a street stand across the road while answering distracting questions and reacting to oncoming vehicles. During the main experiment session, the participant completed all four conditions, each with eight trials. Each condition included two control trials out of eight with only distraction and no oncoming vehicles. The intention was to increase the unpredictability and minimize learning effects. After each condition, the participant completed the SSQ, the NASA TLX questionnaire, and the post-condition questionnaire (two questions based on a 7-point Likert scale). Finally, a post-hoc interview was conducted where the participant completed two questionnaires and answered interview questions. The total duration of the experiment was approximately 1.5 hours. The introduction and tutorial sessions took approximately 20 and 15 minutes, respectively, while the main experiment lasted for 40 minutes. The post-hoc interview was conducted for about 15 minutes.

4 RESULTS

Based on G*Power [18], to achieve a power of 0.90, with an alpha level of 0.05, 24 participants should result in an anticipated medium effect size (0.25) in a within-factors repeated measure ANOVA. In our study, 25 participants were initially recruited, with one dropping out due to cybersickness. Data from the remaining 24 participants was analyzed, all of whom completed all 32 trials across four conditions. We compared measurements among these conditions, as well as between modality matching (*AA* and *VV*) and mismatching (*AV* and *VA*) conditions. We use the Tests of Normality table output from IBM SPSS, which includes Shapiro-Wilk tests, Kolmogorov tests, and Q-Q plots, to test for normality. Due to the smaller sample size in this experiment, we primarily rely on the Shapiro-Wilk tests to determine the normality of the data. For normally distributed data, two-way repeated-measures ANOVA and paired samples t-test were performed with IBM SPSS. For non-parametric data, we used rank-aligned repeated-measures ANOVA and post-hoc analysis with ART-C [15,65]. Post-hoc pairwise comparisons were conducted with Bonferroni corrections. For simplicity, we call both RMANOVA in the result below.

None of the 24 participants experienced cybersickness, as indicated by their SSQ scores. All participants successfully completed the trials either by reaching the street stand or successfully noticing and reacting to the oncoming virtual car. None of the participants was hit by the virtual vehicle, i.e. the success rates for all four conditions are 100%. On average, the participant answered four questions in each trial. There was no significant difference in the number of questions answered between conditions. None of the participants reported difficulties in handling the controllers. Due to the page limits, please refer to appendix for tables and figures related to main effects of the measurements.

4.1 Results for Control Trials vs. Normal Trials

Accuracy of Math Question Answering: The math question accuracy under normal trials are the data before the warning occurred. During auditory distraction, an average of 0.64 (SD=0.21) accuracy was obtained on control trials and an average of 0.62 (SD=0.23) accuracy on normal trials. Control trials resulted in an average accuracy of 0.86 (SD=0.17), while normal trials showed an average accuracy of 0.85 (SD=0.20) during visual distraction.

Speed: The speed under normal trials are the data before the warning appeared. When auditory distraction was present, control trials yielded an average speed of 0.63 m/s (SD=0.30), whereas normal trials achieved an average speed of 0.63 m/s (SD=0.34). During the presence of visual distraction, control trials demonstrated an average speed of 0.53 m/s (SD=0.23), while normal trials exhibited an average speed of 0.52 m/s (SD=0.25).

Head Rotation Range: The head rotation range recorded during normal trials represents the data collected prior to the warning occurred. In the context of auditory distraction, control trials recorded an average head rotation range of 26.02° (SD=5.50), whereas normal trials registered an average head rotation range of 25.97° (SD=6.37). The average head rotation range during visual distraction was 25.06° (SD=4.56) for control trials and 24.82° (SD=4.94) for normal trials.

4.2 Results for Each Condition

Reaction Time: RMANOVA found no main effect of *distraction modality* on RT (F(1,23)=3.11, p=0.091, η^2=0.119). However, a significant main effect of *warning modality* on RT was found (F(1,23)=22.50, p<0.001, η^2=0.494), where conditions with visual warnings (M=1.14 seconds, SD=0.31) had a significantly shorter RT than those with auditory warnings (M=1.45 seconds, SD=0.44). The RMANOVA found no significant interaction between distraction modalities and warning modalities on reaction time (RT) (F(1,23)=2.03, p=0.168, η^2=0.081). The post-hoc tests showed that the RT of *AV* (M=1.12 seconds,SD=0.29) was significantly shorter than *AA* (t(23)=4.02, p=0.001) (M=1.51 seconds,SD=0.51) and *VA* (t(23)=-4.84, p<0.001) (M=1.38 seconds,SD=0.35). Also, the RT of *VV* (M=1.15 seconds,SD=0.32) was significantly shorter than *AA* (t(23)=3.85, p=0.001) (M=1.51 seconds,SD=0.51) and *VA* (t(23)=-3.38, p=0.003) (M=1.38 seconds,SD=0.35) (Fig. 5a)

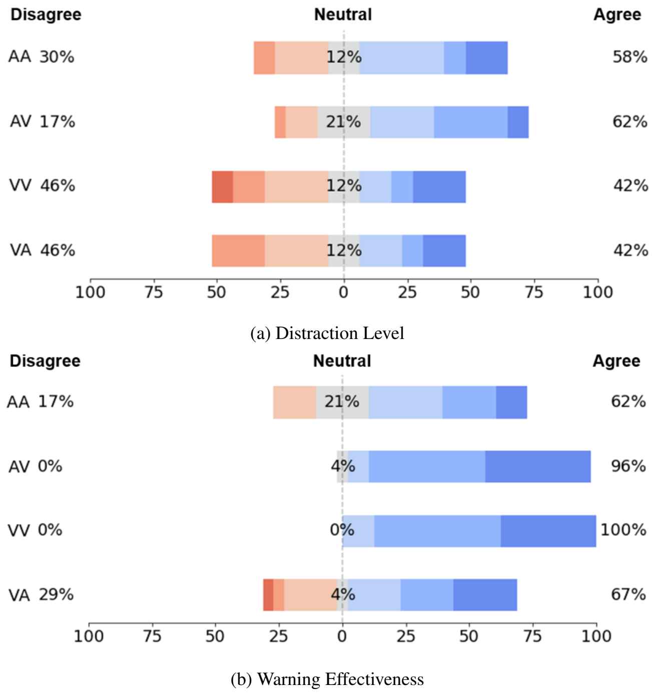
Accuracy of Math Question Answering: The RMANOVA found a significant main effect of the *distraction modality* on the accuracy of evaluating mathematical questions (F(1,23)=49.21, p<0.001, η^2=0.681). The accuracy of conditions with visual distraction (M=0.85, SD=0.20) was significantly higher than those with auditory distraction (M=0.62, SD=0.23). No significant main effect was found on *warning modality* (F(1,23)=0.144, p=0.708, η^2=0.006). The RMANOVA found no significant interaction between distraction and warning modalities on math question accuracy (F(1,23)=0.48, p=0.496, η^2=0.020). The post-hoc tests revealed that *AA* (M=0.63,SD=0.20) led to significantly lower accuracy in mathematical equations evaluation than *VA* (t(23)=-5.48, p<0.001) (M=0.84,SD=0.21) and *VV* (t(23)=-7.76, p<0.001) (M=0.87,SD=0.18). Additionally, *AV* (M=0.62,SD=0.25) had a significantly lower accuracy than *VV* (t(23)=5.90, p<0.001) (M=0.87,SD=0.18) and *VA* (t(23)=-4.10, p<0.001) (M=0.84,SD=0.21). (Fig. 5b)

TLX Score: A significant main effect of *distraction modality* on NASA TLX score was found by RMANOVA (F(1,23)=9.60, p=0.005, η^2=0.295). Conditions with auditory distraction (M=16.49, SD=8.50) resulted in significantly higher TLX scores than those with visual distraction (M=12.00, SD=6.14). There was no main effect of *warning modality* on TLX score (F(1,23)=0.82, p=0.374, η^2=0.035). There was no significant interaction between distraction and warning modality on TLX score (F(1,23)=0.002, p=0.962, η^2<0.001). Pairwise post-hoc tests found that the TLX score of *AA* (M=16.07,SD=6.82) were significantly higher than *VV* (t(23)=2.94, p=0.007) (M=11.64,SD=12.38) and *VA* (t(23)=2.78, p=0.011) (M=12.38,SD=6.17). The TLX score of *AV* (M=16.90,SD=10.03) were found significant higher than *VV* (t(23)=-2.55, p=0.018) (M=11.64,SD=12.38) and *VA* (t(23)=2.33, p=0.029) (M=12.38,SD=6.17). (Fig. 5c)

Perceived Level of Distraction and Effectiveness: The RMANOVA found a significant main effect of *warning modality* on the effectiveness of the warning (F(1,23)=21.67, p<0.001, η^2=0.485). The conditions with auditory warning (M=4.94, SD=1.55) were perceived as having significantly lower effectiveness than those with visual warning (M=6.25, SD=0.73). There was no significant main effect of *distraction modality* on the perceived level of being distracted by the secondary distraction task (F(1,23)=2.43, p=0.133, η^2=0.095). There was no significant interaction between distraction and warning modality on perceived level of distraction (F(1,23)=0.814, p=0.376, η^2=0.034) nor on effectiveness of warning found (F(1,23)=1.90, p=0.182, η^2=0.076). Post-hoc tests showed that participants considered *VV* (M=6.25,SD=0.68) more effective than *VA* (t(23)=2.961, p=0.007) (M=4.96,SD=1.81) and *AA* (t(23)=-4.09, p<0.001) (M=4.92,SD=1.28). In addition, *AV* (M=6.25,SD=0.79) was more effective than *AA* (t(23)=-5.49, p<0.001) (M=4.92,SD=1.28) and *VA* (t(23)=3.84, p=0.001) (M=4.96,SD=1.81). (Fig. 7)

Before Warning

Speed: Before the warning occurred, the RMANOVA found a significant main effect of *distraction modality* on participants' walking speed (F(1,23)=877.85, p<0.001, η^2=0.974). Participants with auditory distraction (M=0.63 m/s, SD=0.34) had a significantly higher walking speed than visual distraction (M=0.52 m/s, SD=0.25). There was no significant interaction between distraction and warning modality on walking speed before warning (F(1,23)=2.16, p=0.155, η^2=0.086). The post-hoc tests showed that *AA* (M=0.65 m/s,SD=0.40) led to a higher walking speed



(a) Distraction Level

(b) Warning Effectiveness

Figure 7: Questionnaires responses after each condition . (a) The perceived level of distraction with the question "I was distracted by the mathematical equations in the current condition." and (b) The perceived effectiveness of the warning with the question "I felt the warning was effective in the current condition".

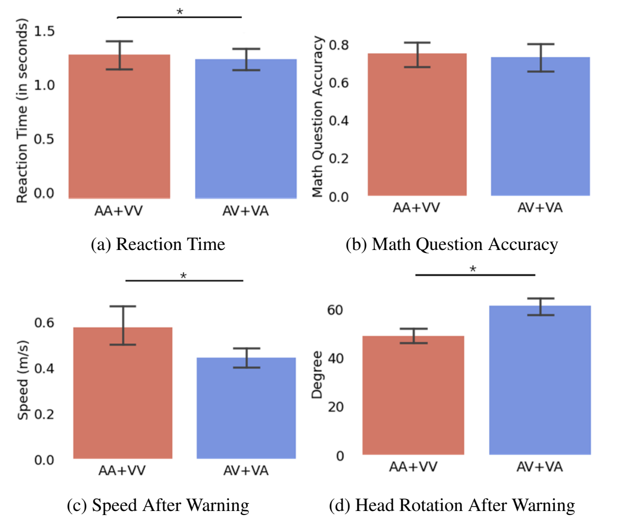
than *VA* (t(23)=22.12, p<0.001) (M=0.52 m/s,SD=0.27) and *VV* (t(23)=28.39, p<0.001) (M=0.51 m/s,SD=0.23). The condition *AV* (M=0.60 m/s,SD=0.27) had a higher speed than *VV* (t(23)=-25.38, p<0.001) (M=0.51 m/s,SD=0.23) and *VA* (t(23)=22.14, p<0.001) (M=0.52 m/s,SD=0.27). (Fig. 6a)

Head Rotation Range: No significant main effect of *distraction modality* on head rotation range before warning was found (F(1,23)=1.04, p=0.318, η^2=0.043). The RMANOVA found no significant difference on head rotation range before warning (F(1,23)=0.862, p=0.363, η^2=0.036). (Fig. 6b)

After Warning

Speed: After the warning occurred, the RMANOVA found a significant main effect of *distraction modality* on walking speed (F(1,23)=10.08, p=0.004, η^2=0.305). Participants under visual distraction (M=0.47 m/s, SD=0.19) had a significantly lower walking speed than under auditory distraction (M=0.54 m/s, SD=0.30). Additionally, RMANOVA also found a significant main effect of *warning modality* on walking speed (F(1,23)=10.59, p=0.003, η^2=0.315). Conditions with visual warning (M=0.46 m/s, SD=0.19) resulted in a significantly lower walking speed than with auditory warning (M=0.55 m/s, SD=0.30). The RMANOVA found a significant interaction between warning and distraction modality on walking speed after warning (F(1,23)=20.42, p<0.001, η^2=0.470). The post-hoc tests showed that *AA* (M=0.66 m/s,SD=0.36) led to higher walking speed than *AV* (t(23)=5.60, p<0.001) (M=0.43 m/s,SD=0.15), *VV* (t(23)=4.29, p<0.001) (M=0.49 m/s,SD=0.22) and *VA* (t(23)=4.35, p<0.001) (M=0.45 m/s,SD=0.16). *VV* (M=0.49 m/s,SD=0.22) showed a higher walking speed than *AV* (t(23)=2.20, p=0.038) (M=0.43 m/s,SD=0.15). (Fig. 6c)

Head Rotation Range: A significant main effect of the *distraction modality* on head rotation range after warning was found



(a) Reaction Time

(b) Math Question Accuracy

(c) Speed After Warning

(d) Head Rotation After Warning

Figure 8: Modality matching results for reaction time, math question accuracy, walking speed after warning and head rotation range after warning. No significant difference found for NASA TLX scores.

(F(1,23)=4.527, p=0.044, η^2=0.164). Conditions with visual distraction (M=52.77°, SD=14.33) result in a smaller head rotation range than under auditory distraction (M=57.24°, SD=11.03). No significant main effect of *warning modality* on head rotation range after warning was found (F(1,23)=0.037, p=0.849, η^2=0.002). The RMANOVA found a significant interaction between warning and distraction modality on head rotation range after warning (F(1,23)=23.34, p<0.001, η^2=0.504). The post-hoc paired t-tests showed that *AA* (M=51.28°,SD=9.12) produced a smaller head rotation range than *AV* (t(23)=-3.974, p=0.001) (M=63.20°,SD=11.02) and *VA* (t(23)=-2.13, p=0.044) (M=59.08°,SD=14.83). *VV* (M=46.45°,SD=10.79) also produced a smaller head rotation range than *AV* (t(23)=-5.77, p<0.001) (M=63.20°,SD=11.02), *AA* (t(23)=2.217, p=0.037) (M=51.28°,SD=9.12) and *VA* (t(23)=-3.87, p=0.001) (M=59.08°,SD=14.83). (Fig. 6d)

4.3 Before Warning vs. After Warning

Speed: Paired t-tests showed that *AV* after warning (M=0.43 m/s,SD=0.15) produced a slower walking speed than before warning (t(23)=4.61, p<0.001) (M=0.60 m/s,SD=0.27) and *VA* after warning (M=0.45 m/s,SD=0.16) produced a slower walking speed than before warning (t(23)=2.14, p=0.043) (M=0.52 m/s,SD=0.27). (Fig. 6)

Head Rotation Range: Paired t-tests yielded significant results indicating notable differences in the observed measures. Specifically, the *AV* condition after warning (M=63.20°,SD=11.02) exhibited a significantly larger range of head rotation compared to the condition before the warning (t(23)=-16.08, p<0.001) (M=25.28°,SD=6.05). Furthermore, the *VA* condition (M=59.08°,SD=14.83) demonstrated a significantly larger head rotation after the warning compared to the condition before the warning (t(23)=-11.94, p<0.001) (M=24.40°,SD=5.37). Similarly, the *AA* condition (M=51.28°,SD=9.12) showed a significantly larger head rotation after the warning than the condition before the warning (t(23)=-10.76, p<0.001) (M=26.67°,SD=6.74). Additionally, the *VV* condition (M=46.45°,SD=10.79) displayed a significantly larger head rotation after the warning compared to the condition before the warning (t(23)=-8.97, p<0.001) (M=25.25°,SD=4.54). (Fig. 6)

4.4 Results for Modality Matching and Modality Mismatching

No significant difference of math question accuracy (p=0.505), TLX score (p=0.710), perceived level of distraction (p=0.479) and effectiveness of warning (p=0.877) on modality matching were found.
Reaction Time A Wilcoxon signed-rank test revealed that RT on modality mismatching *AV+VA* (M=1.25 seconds, SD=0.35) was significantly shorter than modality matching *AA+VV* (M=1.34 seconds, SD=0.46) (p=0.047). (Fig. 8a)

After Warning

Note that the result of modality matching is only valid after the warning occurs, thus we only report the after warning results.
Speed: A Wilcoxon signed-rank test found the walking speed of mismatched modality *AV+VA* (M=0.44 m/s, SD=0.15) was significantly slower than matched modality *AA+VV* (M=0.57 m/s, SD=0.31) ($p<0.001$). (Fig. 8c)
Head Rotation Range: A paired t-test showed the head rotation range of matched modality *AA+VV* (M=46.87°, SD=10.18) was significantly smaller than mismatched modality *AV+VA* (M=61.14°, SD=12.52) ($t(47)=-5.601$, $p<0.001$). (Fig. 8d)

4.5 Post-study Questionnaires and Interview

Out of 24 participants, 22 felt the math question in the auditory form was more distracting (**Q1**), 18 preferred visual warning (**Q2**), and 19 preferred the warning to be in a different modality than the distraction (**Q3**). All but one participant commented that the auditory distraction is more difficult than the visual one. Seventeen participants stated that they would convert the auditory math equation into its visual form before evaluating the correctness of the equation. Six participants reported that auditory warnings were more effective in eliciting their response and creating a sense of urgency about incoming danger. One participant stated "*The auditory warning associated me with the sound of a fire alarm and gave me a greater sense of urgency.*" In addition, seventeen participants stated that visual warnings were clearer and more directional. For example, one participant stated "*A red 3D arrow told me which direction the car was coming, pointing directly in front of you.*"

Regarding the question about modality matching, nine participants commented that they felt confused and hesitated to react when seeing the warning from the same modality as the distraction. One participant commented that "*it was easier for two senses to handle two tasks than for one sense to handle two tasks*". Six participants commented that the warning at the same modality could create less surprise and did not require them to shift their attention to another place (note that three of them still preferred modality mismatching). One participant commented, "*A warning from the same sensory channel made me feel easy to react. Because instead of shifting my attention from one sensory to another, I could respond naturally.*".

5 Discussion

In summary, the result showed that when exposed to auditory distractions (*AV* and *AA*), participants walked significantly faster and reported a higher workload (compared to when they were exposed to visual distractions). Whereas conditions with visual warnings (*AV* and *VV*) were associated with participants' shorter reaction time and higher accuracy in answering secondary math tasks. Regarding the impact of modality matching, the result showed using different distraction and warning modalities (*AV* and *VA*) resulted in shorter reaction times, reduced movement speeds after receiving the warning, and a larger range of head rotation to scan the surrounding, highlighting the positive impacts of modality mismatching. Because no instances were recorded of participants engaging in acceleration upon perceiving the presence of the vehicle. This lack of response could potentially be attributed to the constraints imposed by the physical laboratory setting and the utilization of a virtual reality (VR) headset, which may have hindered the implementation of such a strategy. In this context, decelerating or slowing down was considered a favorable behavior. The post-hoc interview suggests that participants preferred warnings in a different modality from the distraction, as such a combination resulted in less confusion and facilitated more straightforward responses. The following subsections provide a detailed discussion of these phenomena, along with potential theoretical explanations for them.

5.1 Distraction Modality and Warning Modality

For the distraction modality, the results showed that conditions with auditory distraction (*AA* and *AV*) had a significantly lower math accuracy and a higher perceived workload, suggesting that the conditions with auditory distraction tasks were more difficult and demanding. Most participants (all but two) found the auditory mathematical equations to be more distracting. There could be several contributing factors to this outcome. For example, it took about 2 seconds to complete the play of an auditory equation. In contrast, most participants could comprehend the visual equations in a much shorter amount of time and started calculating. Several participants commented that they would convert the auditory arithmetic problems into a visual symbolic form before solving them, which could have added to the difficulty. It is also possible that participants, all university students, had more experience solving mathematical questions in their visual than auditory form. Regarding the warning modality, our results show that when presented with visual warnings (*AV* and *VV*), participants exhibited better behavioural responses. Specifically, participants achieved a shorter reaction time to identify the oncoming vehicle and walked at a reduced speed, indicating an increased awareness of the warning and the oncoming vehicle. However, the smaller head rotation range in these conditions seems to suggest that the visual warning might have an undesired detrimental effect of limiting the scan range of the user. We might interpret the phenomenon as the user's attention being drawn to the warning, which is located centrally in the field of view. As a result, the user may unconsciously reduce their head rotation range and scanning of the surrounding environment. Further investigation, preferably using an eye-tracker, is needed for this effect. Regarding the auditory warnings, multiple participants commented that while the auditory warnings were still useful, yet sometimes they had difficulty determining the direction of the incoming vehicle from the sound. These results align with previous research on road safety, which has shown that visual warnings are often more effective than auditory warnings for drivers [49, 68]. Additionally, our participants generally preferred visual warnings, which has also been observed in other studies on AR-based/visual safety interventions [1, 13].

5.2 Modality Matching between Distraction and Warning

When considering both the modalities of distraction and warning, our findings suggest that participants performed better when these modalities were different (*AV* and *VA*). In particular, participants had a faster reaction time to notice the oncoming vehicle, walked at a reduced speed, indicating an increased response to the warning, and exhibited a larger head rotation range, suggesting increased scan range of the surrounding environment for the oncoming vehicle. Moreover, our results indicate that the majority of participants (19 out of 24) preferred receiving distractions and warnings through different sensory channels. In follow-up interviews, participants reported that stimuli presented through different sensory channels made them more alert and aware. Some participants noted that when focused on a task presented in one sensory channel, they were more likely to ignore or be confused by warnings presented in the same channel. Note that the different perceived workloads between visual and auditory mathematical questions should not affect the result as both distraction modalities were presented in the combinations of (*AA*, *VV*) against (*AV*, *VA*). The participant's preference for receiving a warning from a different modality from the distraction might be explained by two theories:

Multiple Resource Theory (MRT) MRT [63] suggests that cognitive control is limited by several specific cognitive resources and is commonly used to explain why multitasking is hard when it demands similar cognitive resources. However, according to MRT, modality mismatching could be beneficial when it reduces demands on common cognitive resources. For example, the distraction task of listening to the auditory mathematical equation might rely on the phonological loop (the cognitive resource for processing verbal information), whereas the task of paying attention to the visual warning (red arrows pointing to the incoming vehicle) would primarily rely on the visuospatial sketchpad (the cognitive resource for visual and spatial information). Since these two tasks rely on different cognitive resources, there is less competition for the resources, and as a result, this might explain why some participants felt higher awareness of the conditions with modality mismatching. In addition, although switching between different modalities might require extra time and effort to reconfigure the attention system, it has been shown to allow the cognitive system to refresh and avoid fatigue on a single task, resulting in a higher preference for receiving the warning in a different modality from the distraction [34].

Following MRT, one may be inclined to rely on warnings communicated through all available sensory channels in the expectation that at least one of the cognitive resources will detect the warning. However, it's important to note that utilizing multi-modal information may result in increased workload [25] and could lead to additional costs associated with information integration and prioritization [60, 63].

Inattentional Blindness (IB) IB [36, 53] refers to the phenomenon where an individual fails to notice an unexpected event in plain sight when their attention is focused on something else, as demonstrated in the famous "invisible gorilla" experiment. A similar phenomenon in the auditory channel is inattentional deafness, where an individual fails to notice other sounds in their environment while focusing on one particular sound or conversation. Both phenomena demonstrate the limitations of human attention and perception and highlight the importance of being aware of one's own attentional biases and limitations.

In the context of our experiment, IB could potentially manifest itself in the form of a delayed reaction to a visual warning (red arrow) when an individual's attention is focused on a visual mathematical equation. This occurs because the individual's attention is already occupied by the mathematical task and they may not have the attention resources to process additional stimuli in their visual field. Some previous IB research also concurs with our result that using a different warning modality from the distraction is more effective. For example, Sinnett et al. [54] investigated IB for words presented visually and auditorily within and across sensory modalities. The study found that IB is less prevalent when attention is divided across modalities than within the same modality, indicating shared attentional resources across sensory modalities. In sum, this might serve as another explanation to support the findings that using a warning modality different from the distraction might lead to a better performance of the street crossing behaviour.

6 LIMITATION

While our study provided valuable insights into the effects of modality matching for street crossing behaviours, there are several limitations that future research could address.

First, our study used a sample size of $N = 24$ participants, which satisfies a minimum sample size necessary for counterbalancing the four experimental conditions. However, using a larger and more representative sample of participants could have increased the accuracy and reliability of our conclusions. There were also more males among the participants, which might affect the generalizability of the result. In addition, some papers [20, 52, 57] have explored the difference between the behavior of young and old adults. The average age of the participants we recruited was 23.67, we will recruit participants of other age groups to explore the effect of age in future work.

Second, we made several experiment design choices that might affect the results. Our experiment did not simulate the limited FoV of the AR headset. We made this design decision to avoid the smaller FoV becoming a confound to the experiment. We also tried to minimise the impact by putting the mathematical equations and warnings at the centre of the FoV. Still, further investigation might be needed when applying the finding to the smaller FoV in the current generation AR devices. Another design decision we made was the randomization of the starting positions of the virtual vehicle with the intention to reduce the learning effect. Nevertheless, this approach may influence how participants prioritize their attention allocation between distractions and the surrounding environment. Future studies might consider treating the starting position as an independent variable and systematically assess its impact. Our experiment considered only visual and auditory modalities for warning and distraction, but it's worth noting that other modalities, such as haptic feedback, could also be explored, as discussed in Section 2.2. Additionally, we did not compare the effectiveness of single modality warning versus multi-modal warnings, although recent research suggests that multi-modal warnings can be effective in certain contexts [32, 38]. However, it's important to note that our study primarily aimed to investigate the impact of modality matching, and our results have contributed to the knowledge in this area.

Last but not least, it is crucial to consider whether the findings of our study can be applied to real-world situations. While our experiment used an abstract secondary task involving mathematical calculation, which is common in psychology experiments, real-world activities such as listening to a podcast or performing web browsing on an AR display are different. Similarly, different sizes and types of visual and audio warnings might also lead to different outcomes. Additionally, the real world environment contains significantly more noise and distracting factors than the virtual environment in the experiment. Although we recognize these differences, we, and other researchers [8–10, 27], chose virtual simulations to conduct road safety experiments to avoid the potential risk of physical harm to participants in real-world. However, future research should take into account variables like individual variances and real-world environmental factors, while also exploring the applicability and generalizability of findings from this line of research.

7 CONCLUSION

This paper investigated the impact of distraction and warning modality matching on pedestrian crossing behaviour. A direct comparison of modality matching and mismatching conditions revealed that conditions with mismatched modalities (*AV* and *VA*) performed better than conditions with matched modalities (*AA* and *VV*), as evidenced by faster reaction time, reduced walking speed, and increased scan coverage. Participants in our study also expressed a preference for warnings to be presented in a modality different from the modality of the distraction. The findings suggest that future AR interface design should consider using warning modalities that are different from those causing distraction when alerting users about hazards in the surrounding environment.

ACKNOWLEDGMENTS

We would like to thank anonymous ACs and reviewers for their valuable feedback, which greatly enhanced the quality of this work. We would also like to thank all those who contributed to the project, especially William Josiah White and Branko Vanjorek.

REFERENCES

[1] T. Akhmetov and H. A. Varol. An augmented reality-based warning system for enhanced safety in industrial settings. *IEEE Transactions on Industrial Informatics*, 2022.

[2] R. Alsaleh, T. Sayed, and M. H. Zaki. Assessing the effect of pedestrians' use of cell phones on their walking behavior: A study based on automated video analysis. *Transportation research record*, 2672(35):46–57, 2018.

[3] P. Baudisch and R. Rosenholtz. Halo: A technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, p. 481–488. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/642611.642695

[4] T. M. Bhojwani, S. D. Lynch, M. A. Bühler, and A. Lamontagne. Impact of dual tasking on gaze behaviour and locomotor strategies adopted while circumventing virtual pedestrians during a collision avoidance task. *Experimental brain research*, 240(10):2633–2645, 2022.

[5] T. Campisi, I. I. Otković, S. Šurdonja, and A. Deluka-Tibljaš. Impact of social and technological distraction on pedestrian crossing behaviour: a case study in enna, sicily. *Transportation research procedia*, 60:100–107, 2022.

[6] J. Chen, E. Šabić, S. Mishler, C. Parker, and M. Yamaguchi. Effectiveness of lateral auditory collision warnings: should warnings be toward danger or toward safety? *Human factors*, 64(2):418–435, 2022.

[7] F. Chiossi, S. Villa, M. Hauser, R. Welsch, and L. Chuang. Design of on-body tactile displays to enhance situation awareness in automated vehicles. In *2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp. 1–6. IEEE, 2022.

[8] M. Colley, E. Bajrovic, and E. Rukzio. Effects of pedestrian behavior, time pressure, and repeated exposure on crossing decisions in front of automated vehicles equipped with external communication. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2022.

[9] M. Colley, S. Li, and E. Rukzio. Increasing pedestrian safety using external communication of autonomous vehicles for signalling hazards. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, pp. 1–10, 2021.

[10] M. Colley, M. Walch, J. Gugenheimer, A. Askari, and E. Rukzio. Towards inclusive external communication of autonomous vehicles for pedestrians with vision impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.

[11] C. Department of Infrastructure, Transport and R. D. A. Government. Road trauma australia 2018 statistical summary. 2018.

[12] D. Dey, K. Holländer, M. Berger, B. Eggen, M. Martens, B. Pfleging, and J. Terken. Distance-dependent ehmis for the interaction between automated vehicles and pedestrians. In *12th international conference on automotive user interfaces and interactive vehicular applications*, pp. 192–204, 2020.

[13] T. A. Dingus, S. Jahns, A. Horowitz, and R. Knipling. Human factors design issues for crash avoidance systems. *Human factors in intelligent transportation systems*, pp. 55–93, 1998.

[14] D. Dobbelstein, P. Henzler, and E. Rukzio. Unconstrained pedestrian navigation based on vibro-tactile feedback around the wristband of a smartwatch. *Proceedings of the 2016 CHI*, 2016.

[15] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 754–768. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474784

[16] M. R. Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting*, vol. 32, pp. 97–101. Sage Publications Sage CA: Los Angeles, CA, 1988.

[17] J. B. F. V. Erp, H. A. H. C. V. Veen, C. Jansen, and T. Dobbins. Waypoint navigation with a vibrotactile waist belt. *ACM Trans. Appl. Percept.*, 2(2):106–117, apr 2005. doi: 10.1145/1060581.1060585

[18] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang. Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4):1149–1160, 2009.

[19] M. Frey. Cabboots: Shoes with integrated guidance system. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, p. 245–246. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1226969.1227019

[20] M. Gérin-Lajoie, C. L. Richards, and B. J. McFadyen. The circumvention of obstacles during walking in different environmental contexts: a comparison between older and younger adults. *Gait & posture*, 24(3):364–369, 2006.

[21] U. Gruenefeld, A. E. Ali, S. Boll, and W. Heuten. 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*, number Article 40 in MobileHCI '18, pp. 1–11. Association for Computing Machinery, New York, NY, USA, Sept. 2018.

[22] U. Gruenefeld, D. Ennenga, A. E. Ali, W. Heuten, and S. Boll. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. SUI '17, p. 109–118. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132175

[23] S. Gustafson, P. Baudisch, C. Gutwin, and P. Irani. Wedge: Clutter-free visualization of off-screen locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 787–796. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357179

[24] S. G. Gustafson and P. P. Irani. Comparing visualizations for tracking off-screen moving targets. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pp. 2399–2404, 2007.

[25] E. C. Haas and J. B. F. van Erp. Multimodal warnings to enhance risk communication and safety. *Saf. Sci.*, 61:29–35, Jan. 2014.

[26] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[27] K. Holländer, A. Krüger, and A. Butz. Save the smombies: App-Assisted street crossing. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, number Article 22 in MobileHCI '20, pp. 1–11. Association for Computing Machinery, New York, NY, USA, Oct. 2020.

[28] S. Jain, C. Borgiattino, Y. Ren, M. Gruteser, Y. Chen, and C. F. Chiasserini. Lookup: Enabling pedestrian safety services via shoe sensing. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 257–271, 2015.

[29] J. Jung, H. Lee, J. Choi, A. Nanda, U. Gruenefeld, T. Stratmann, and W. Heuten. Ensuring safety in augmented reality from trade-off between immersion and situation awareness. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 70–79, Oct. 2018.

[30] A. Jylhä, Y.-T. Hsieh, V. Orso, S. Andolina, L. Gamberini, and G. Jacucci. A wearable multimodal interface for exploring urban points of interest. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, p. 175–182. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2818346.2820763

[31] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3

[32] J. D. Lee, J. D. Hoffman, and E. Hayes. Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, p. 65–72. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/985692.985701

[33] S. Licence, R. Smith, M. P. McGuigan, and C. P. Earnest. Gait pattern alterations during walking, texting and walking and texting during cognitively distractive tasks while negotiating common pedestrian obstacles. *PLoS one*, 10(7):e0133281, 2015.

[34] G. W. Lindsay. Attention in psychology, neuroscience, and machine learning. *Frontiers in Computational Neuroscience*, 14, 2020. doi: 10.3389/fncom.2020.00029

[35] H. Liu, T. Hirayama, and M. Watanabe. Importance of instruction

for pedestrian-automated driving vehicle interaction with an external human machine interface: Effects on pedestrians' situation awareness, trust, perceived risks and decision making. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pp. 748–754. IEEE, 2021.

[36] A. Mack. Inattentional blindness: Looking without seeing. *Current directions in psychological science*, 12(5):180–184, 2003.

[37] T. Maeda, H. Ando, T. Amemiya, N. Nagaya, M. Sugimoto, and M. Inami. Shaking the world: Galvanic vestibular stimulation as a novel sensation interface. In *ACM SIGGRAPH 2005 Emerging Technologies*, SIGGRAPH '05, p. 17–es. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1187297.1187315

[38] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, and J. Schöning. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3389–3401, 2020. doi: 10.1109/TVCG.2020.3023605

[39] A. Matviienko, S. Ananthanarayan, S. S. Borojeni, Y. Feld, W. Heuten, and S. Boll. Augmenting bicycles and helmets with multimodal warnings for children. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–13, 2018.

[40] A. Matviienko, S. Ananthanarayan, R. Kappes, W. Heuten, and S. Boll. Reminding child cyclists about safety gestures. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*, pp. 1–7, 2020.

[41] T. L. McIsaac, E. M. Lamberg, and L. M. Muratori. Building a framework for a dual task taxonomy. *BioMed research international*, 2015, 2015.

[42] D. Moore, R. Currano, and D. Sirkin. Sound decisions: How synthetic motor sounds improve autonomous vehicle-pedestrian interactions. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 94–103, 2020.

[43] J. Nasar, P. Hecht, and R. Wener. Mobile telephones, distracted attention, and pedestrian safety. *Accident analysis & prevention*, 40(1):69–75, 2008.

[44] J. Orschiedt, J. Schmickler, V. Nußer, T. Fischer, J. Hermsdörfer, and M. Krüger. Writing while walking: the impact of cognitive-motor multi-tasking on collision avoidance in human locomotion. *Human movement science*, 88:103064, 2023.

[45] M. Pfeiffer, T. Dünte, S. Schneegass, F. Alt, and M. Rohs. 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*, CHI '15, pp. 2505–2514. Association for Computing Machinery, New York, NY, USA, Apr. 2015.

[46] M. Prasad, P. Taele, D. Goldberg, and T. A. Hammond. HaptiMoto: turn-by-turn haptic route guidance interface for motorcyclists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 3597–3606. Association for Computing Machinery, New York, NY, USA, Apr. 2014.

[47] A. San Martín and J. Kildal. Audio-visual ar to improve awareness of hazard zones around robots. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, p. 1–6. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290607.3312996

[48] S. M. Schabrun, W. van den Hoorn, A. Moorcroft, C. Greenland, and P. W. Hodges. Texting and walking: strategies for postural control and implications for safety. *PloS one*, 9(1):e84312, 2014.

[49] F. Schwarz and W. Fastenmeier. Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness. *Accident Analysis & Prevention*, 101:55–66, 2017.

[50] D. C. Schwebel, D. Stavrinos, K. W. Byington, T. Davis, E. E. O'Neal, and D. De Jong. Distraction and pedestrian safety: how talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis & Prevention*, 45:266–271, 2012.

[51] M. Serino, K. Cordrey, L. McLaughlin, and R. L. Milanaik. Pokémon go and augmented virtual reality games: a cautionary commentary for parents and pediatricians. *Curr. Opin. Pediatr.*, 28(5):673–677, Oct. 2016.

[52] W. S. Silva, B. J. McFadyen, J. Fung, and A. Lamontagne. Reading text messages at different stages of pedestrian circumvention affects strategies for collision avoidance in young and older adults. *Gait & Posture*, 76:290–297, 2020.

[53] D. J. Simons and C. F. Chabris. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *perception*, 28(9):1059–1074, 1999.

[54] S. Sinnett, A. Costa, and S. Soto-Faraco. Manipulating inattentional blindness within and across sensory modalities. *The quarterly journal of experimental psychology*, 59(8):1425–1442, 2006.

[55] R. Soret, P. Charras, C. Hurter, and V. Peysakhovich. Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1–8, 2019.

[56] R. Soret, P. Charras, I. Khazar, C. Hurter, and V. Peysakhovich. Eye-tracking and virtual reality in 360-degrees: Exploring two ways to assess attentional orienting in rear space. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Adjunct. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3379157.3391418

[57] W. Souza Silva, B. McFadyen, J. Fung, and A. Lamontagne. Effects of age on obstacle avoidance while walking and deciphering text versus audio phone messages. *Gerontology*, 65(5):524–536, 2019.

[58] D. Stavrinos, K. W. Byington, and D. C. Schwebel. Distracted walking: cell phones increase injury risk for college pedestrians. *Journal of safety research*, 42(2):101–107, 2011.

[59] H. Tapiro, T. Oron-Gilad, and Y. Parmet. Pedestrian distraction: The effects of road environment complexity and age on pedestrian's visual attention and crossing behavior. *Journal of safety research*, 72:101–109, 2020.

[60] C. A. Tirado Cortes, H.-T. Chen, D. Sturnieks, J. Garcia Marin, S. Lord, and C.-T. Lin. Evaluating balance recovery techniques for users wearing Head-Mounted display in VR. *IEEE Trans. Vis. Comput. Graph.*, July 2019.

[61] Y. Tong, B. Jia, and S. Bao. An augmented warning system for pedestrians: User interface design and algorithm development. *Applied Sciences*, 11:7197, 08 2021. doi: 10.3390/app11167197

[62] T. Wang, G. Cardone, A. Corradi, L. Torresani, and A. T. Campbell. WalkSafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, number Article 5 in HotMobile '12, pp. 1–6. Association for Computing Machinery, New York, NY, USA, Feb. 2012.

[63] C. D. Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.

[64] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman. *Engineering psychology and human performance*. Psychology Press, 2015.

[65] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[66] J. Woodward and J. Ruiz. Analytic review of using augmented reality for situational awareness. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[67] X.-D. Yang, K. Hasan, N. Bruce, and P. Irani. Surround-see: enabling peripheral vision on smartphones during active use. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 291–300, 2013.

[68] Z. Yang, J. Shi, Y. Zhang, D. Wang, H. Li, C. Wu, Y. Zhang, and J. Wan. Head-up display graphic warning system facilitates simulated driving performance. *International Journal of Human–Computer Interaction*, 35(9):796–803, 2019.

[69] H. Zhang, C. Zhang, F. Chen, and Y. Wei. Effects of mobile phone use on pedestrian crossing behavior and safety at unsignalized intersections. *Canadian Journal of Civil Engineering*, 46(5):381–388, 2019.

[70] H. Y. Zhu, H.-T. Chen, and C.-T. Lin. The effects of virtual and physical elevation on physiological stress during virtual reality height exposure. *IEEE Trans. Vis. Comput. Graph.*, PP, Dec. 2021.