A High-Fidelity VR Simulation Study: Do External Warnings Really Improve Pedestrian Safe Crossing Behavior?

by

Fan He

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2020

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Fan He was the sole author for Chapters 1 and 2, which were written under the supervision of Dr. Catherine Burns and co-supervision of Dr. Yeti Li and were not written for publication.

Exceptions to sole authorship of material are as follows:

Research presented in Chapters 3, 4 and 5:

Dr. Catherine Burns, Dr. Yeti Li and Murat Dikmen are co-authors on any publications relating to this work.

This research was conducted at the University of Waterloo by Fan He under the supervision of Dr. Catherine Burns. Fan He and Dr. Yeti Li contributed to study design. Fan He conducted the one-on-one in person experiments and interviews. Dr. Yeti Li and Fan He were the primary coders. Fan He, Dr. Yeti Li and Murat Dikmen participated in data analysis. Fan He wrote the draft manuscripts, which all co-authors contributed intellectual input on.

As lead author of these three chapters, I was responsible for contributing to conceptualizing study design, carrying out data collection and analysis, and drafting and submitting manuscripts. My coauthors provided guidance during each step of the research and provided feedback on draft manuscripts.

Abstract

To better communicate with pedestrians, adding external displays to autonomous vehicles (AVs) has been proposed as a potential communication method to encourage safe crossing behavior by pedestrians. Whereas, most researchers have conducted intercept interviews, lab studies, or simulation studies to explore the efficacy of these displays, these approaches only studied crossing intention but did not explore crossing behavior. We developed a high-fidelity virtual reality scenario where participants could demonstrate actual crossing behavior within an adequately replicated real-world street. We simulated a local street with scalability of the real world in a VR environment, conducted an experiment in an empty space large enough for participants to move across the road in the VR environment. A mixed-method approach assessed attitudinal and behavioral interactions with potential warning patterns. The results showed that the warning patterns contributed significantly to pedestrians' perceptual vigilance, as in past studies, but safer crossing behavior was not observed. This suggests that crossing intention measures may not be an adequate substitute for behavioral measures of crossing.

Acknowledgements

I wish to thank my supervisor Professor Catherine M. Burns for bringing me into the field of cognitive human factors engineering. It was a great privilege and honor to work and study under her guidance.

I also wish to thank my co-supervisor Dr. Yeti Li for his invaluable guidance throughout this research. His vision, sincerity and empathy have deeply inspired me.

NSERC CREATE Training Program is especially thanked for supporting me during my MASc studies.

I want to express my gratefulness to my colleagues in the Advanced Interface Design Lab, for their patience, support, generous help and great sense of humor. Special thanks to Ramtin Lotfabadi, Murat Dikmen, Rodrigo Arcuri, Reicilis Casares Li and Haoyun Chen for their company and encouragement during early research preparation.

Thanks to all the student volunteers at the University of Waterloo, who spent their valuable time participating in the experiments.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Designing a communication interface to express intentions between automated vehicles (AVs) and pedestrians is a key element for safety. Recent studies showed the popularity of exploring external visual displays on AVs to facilitate pedestrian and vehicle interactions at street crossings. There has been a particular interest in adding external displays to AVs to achieve direct, simple, and unambiguous communication with pedestrians (Ayoub et al., 2019). The main interest seems to lie in the ideation of novel design concepts, including colours (Li et al., 2018; Dey et al., 2020), text messages (Clercq et al., 2019), eye contact (Chia-Ming Chang, 2017), light-emitting diode (LED) indicators (Clamann, 2016), and light projections (Overview: Mercedes-Benz F 015 luxury in motion , 2017).

However, how these displays could improve safety and how pedestrians might actually behave when faced with these new displays is still unclear. To date, only the perception of these design concepts has been studied, and behavioural compliance to the designs has not been evaluated.

In a few cases, naturalistic studies, even Wizard-of-Oz studies, have been employed in AVs research to study road users' behaviours. While these studies suffer the disadvantages of high costs, significant efforts to create controlled stimuli conditions and the risk of undermining road safety (Ayoub et al., 2019). More advanced information systems, such as VR (Virtual Reality)/AR (Augmented Reality) and advanced driving simulators, have been increasingly utilized in recent interaction studies, and offer a high-fidelity, safe, controlled, and reproducible environment to further explore road users' behavioural interactions. VR/AR affords quick, flexible, iterative and progressive ideation prototyping, experiencing and evaluation.

With the adoption of a wireless VR system, we proposed simple external visual displays as flashing warning and replicated a two-way local street with both directions in a virtual environment for this study. Participants who have experience crossing this street were recruited. This study allowed for free movement in the simulated street environment that would encourage real pedestrian behaviours. Then a mixed-method approach was conducted to analyze pedestrians' crossing intentions and behaviours warned by visual cues at uncontrolled locations (e.g., mid-block). The analysis results identified a gap between pedestrians' intentions and behaviors on jaywalking with external vehicular warning displays. Even though warnings with various flashing rates could help pedestrians distinguish hazardous situations attitudinally, the overall behavioral measures suggested that the warning displays had limited influence on crossing behaviors. Understanding this gap could serve to improve the translation of intentions into safety actions when designing warnings. For instance, pedestrian may conduct more vigilant crossings only when the flashing rates properly communicate the situational hazardousness of the traffic. The arguing of the limited transferability of attitudinal assessment results of AV external displays to a real-world interaction is one of the main contributions in this study.

The other contribution of this study is the exploration of using the VR for an enhanced close-to-reality experience to encourage pedestrian's real crossing behavior in a simulated environment. Involving complicating factors as simulated traffic noise, multiple people crossing and multiple lanes with numerous vehicles are quite challenging in most recent studies. Unlike in the real world, where people may have high levels of experience with certain crossing locations, in simulations users are often asked to cross in new situations. In other words, the simulated environment (e.g., street, building, and traffic design) tested in most research does not adequately replicate a real-world street (Colley et al., 2020). Thus, it can be difficult for participants to reflect on a real crossing experience within a distorted environment. This study presents a high-fidelity simulation in traffic volume, noise control and space scale to extend these efforts. potential limitations. First, we replicated a two-way local street with both directions in a VR environment and recruited participants who have experience crossing this street. This should increase the reliability of our research. Second, we took a mixed-method approach that included both intention and behaviour measures for pedestrian crossing at unmarked

This thesis focuses on the relationship between intention and behaviour with designed external warning patterns at crossing with unmarked crosswork by implementing a reasonably realistic simulated VR immersive environment (Chapter 3). The procedure, and the results of the experiment conducted in a high-fidelity VR environment were presented (Chapter 4). The experiment manipulated two parameters of traffic combinations, which were fleet speed and gap distance. In the experiment, perceptual rating tasks and behavioral crossing tasks were completed by 13 participants. Participants simultaneously completed an auditory task, in which an auditory task was performed.

This thesis is organized as follows: Chapter 2 provides a review of background and related literature that contributed to our study: the proposed external vehicular warning design concepts, the investigation of evaluation methods and potential concerns regarding the gap between intention and behaviour. Chapter 3 presents our design and experiment approach used Virtual Reality techniques. Chapter 4 discusses the settings, procedures, and results of experiments. In Chapter 5, I draw conclusions from the study and discuss limitations and ideas for future work.

Chapter 2
Background

In this chapter, I review relevant research to lay a foundation for assessing external warning displays at pedestrian crossings without marked crosswalks. First, I introduce the necessary and concerns with employing external warning displays on automated vehicles. Then I present recent novel design concepts of external warnings for AVs and their implementation and form factors. After this, I discuss prevalent methodologies used in understanding pedestrian behavior at crossings without painted crosswalks. This discussion provides a reference to compare experiments on their generalizability to real world behaviours. It also explores some of the concerns about each approach's limitations when understanding traffic safety and pedestrian interactions. This discussion is the basis for understanding some of the major characteristics of our study method and objects.

2.1 Autonomous Vehicle's External Warning Designs

I discuss warning designs in two sections. First, I discuss the need to explore communication mechanisms between vehicles and external road users in the context of future traffic systems with a large number of AVs. Then I review of the recent prevalent AVs external warning patterns. This review explains the current trends of vehicular warning patterns on AVs and provides inspirations for exploring AVs communication mechanisms with external road users. Finally, I briefly introduce the shortcomings of these warning pattern designs.

2.1.1 Communication at Pedestrian Crossings

Human behaviour is, directly or indirectly, an important question of some negative traffic consequences. Under different traffic conditions, the road users interact with each other and make decisions. The interactions among road users would significantly affect information exchange style and the effectiveness of the traffic system. Some may lead to risky behaviours and might eventually evolve into critical events. Almost half of the people killed around the world each year in collisions are more vulnerable road users, i.e. motorcyclists, cyclists and pedestrians (WHO, 2016). While road users such as pedestrians may be vulnerable to becoming victims of these incidents, their risky behaviours, just like drivers' risky behaviours, may also contribute to such outcomes (King et al., 2009).

At present, there are a wealth of studies on driver behaviour obtained in the laboratory and naturalistic driving studies. However, the understanding of what happens to other vulnerable road users within the traffic system in different situations is limited. For example, pedestrians crossing at uncontrolled locations (also known as "jaywalking") is a typical traffic situation where pedestrians' risk of being injured by an automobile is high. It was reported that most pedestrian fatalities (73%) did not occur at signalized intersections (NHTSA, 2019). Therefore, this situation requires accurate communication between drivers and pedestrians to make safety-critical decisions without delay. Understanding pedestrian-vehicle communication mechanisms is a prerequisite to identifying the main causes of, and the most promising approaches to mitigate any negative consequences.

On the other hand, with the acceleration of the development of new technology, the modern vehicle is no longer considered just a geographic transportation tool, but rather as a mobile device with communication and information exchange virtually everywhere. AVs are expected to make fewer errors than humans. Consequently, increased safety and facilitated mobility are top of the anticipated benefits of AVs (Fagnant & Kockelman, 2015).

However, even if AVs were to behave more safely around pedestrians than human-operated cars, would car-pedestrian interactions be, and be perceived as, less risky, given pedestrian behaviour also plays a role in the outcome of these interactions? Eighty-four percent of pedestrians advocated they sought eye contact with drivers before crossing and 2.7% raised their hand/fingers to indicate intentions to cross (Dey & Terken, 2017). While the degree of autonomy is increasing to driving automation (SAE Level 3 and above), it is highly likely that drivers/passengers in the vehicle may not pay attention to other road users or there may be no drivers or passengers in the vehicle, which makes it difficult to interact with other road users. Without a human driver's attention or presence, eye-contact or hand gestures as conventional nonverbal interpersonal communications between pedestrians and drivers may vary or vanish.

Additionally, a substantial number of pedestrians (about 20–30 percent) exhibit distracted walking by listening to audio content or using a portable device (Basch et al., 2014).. Meanwhile, based on the collected data of injury reports from hospitals across the U.S. from the National Electronic Injury Surveillance System, Nasar & Troyer (2013) found the total number of pedestrian injuries increased from 0.58 percent to 3.67 percent between 2004 and 2010 due to portable electronic device usage at street crossings.

Consequently, it is very helpful to investigate the communication mechanisms between pedestrians and future driving entities to improve street crossing safety when we have high levels of autonomous traffic and distracted pedestrian walking at the same time. Also, if any designed external communication modalities could be assessed as effective, trustworthy and robust to facilitate communication in such a highly automated system, it would shed some light on the exploration of AVs and other artificial intelligent agent's interaction approaches.

2.1.2 Novel Design Concepts of External Warnings

Interactions with road users can be divided into formal interactions (e.g., braking lights, turn signals, and warning lights) and informal interactions (e.g., body expressions) (Färber, 2016). Both informal and formal interactions have been studied by many researchers. Vehicle size and speed have also been investigated to understand accepted gaps and estimated time to cross (Beggiato et al., 2017). Since ability to cross can change with age, accepted gaps and estimated time to cross have been investigated by cohort (Beggiato et al., 2017). Pedestrians' emotion and decisions, as influenced by different vehicle behaviours have been investigated (Zimmermann & Wettach, 2017).

Many researchers have explored how to provide external displays to communicate the intentions of the vehicle and facilitate interactions with pedestrians. For example, Li et al. (2018) proposed external display warnings with different colours to indicate three levels of safety-related information (alert, dangerous, and safe situations). Dey et al. (2020) asserted that animation patterns do not have as much impact as colours in this communication, although uniform animations may be more favourable than laterally sweeping patterns. By adding eyes to the headlights, Chang (2017) tried to help pedestrians make direct, simple, and unambiguous street-crossing decisions. To overcome this lack of communication, external communication modalities such as displays (Florentine et al., 2016), LED strips (Habibovic et al., 2016), movement patterns (Zimmermann & Wettach, 2017), projections (Ackermann et al., 2019) auditory or tactile cues (Mahadevan et al., 2018) as well as combinations thereof (Mahadevan et al., 2018) and enhancement of the infrastructure (Sieß et al., 2015) have been studied.

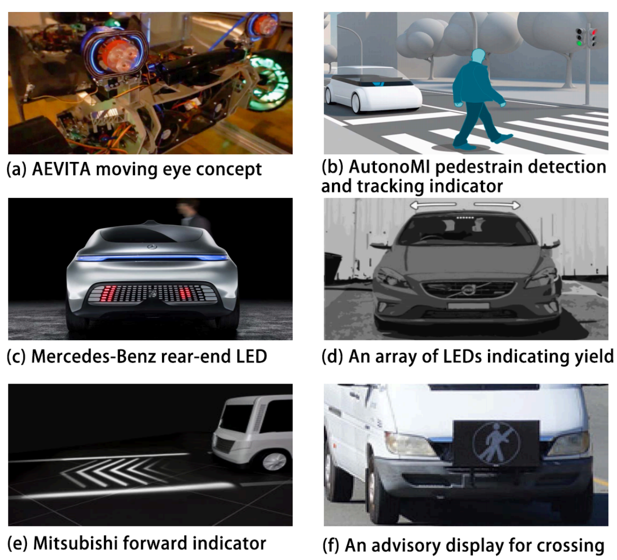

Figure 1: Difference concepts of communication for autonomous vehicles

(Rasouli et al., 2017).

The challenge of these novel external communications is how to make sure that such displays convey the vehicle's intentions directly, simply, and unambiguously. To do so, text messages, such as "Walk" and "Don't Walk" are effective (Clercq et al., 2019), however, text messages might not work for those who do not understand the language (e.g., foreigners). Another study done by Currano et al. (2018) highlighted regional differences in how pedestrians interacted with automated vehicles, which helped predict pedestrian behaviour and inform customized design strategies for future automated vehicles in different regions. Studies have shown that cultural differences can influence the interaction between automated vehicles and pedestrians (Currano et al., 2018). Hence, cultural-specific symbols and metaphors should be avoided.

Then, other possible stimuli were used, including colors (Li et al., 2018) and eye contacts (Chang, 2017). More recent study results suggested that while red and green have immediate associations with 'stop' and 'go', the meaning of these messages is not always clear and may lead to confusion (Dey et al., 2020).

Interestingly, as AVs are expected to bring enormous changes to our life, many researchers believe that the complex future interactions between road users should employ novel communication modalities to adapt to the impending changes and the potential impacts on future traffic systems. Although novel external communication modalities promote inspirations, whether they are effective, efficient and robust modalities is still unclear.

To summarize, recent research and development have produced several concepts of Human Machine Interface (eHMI) design which vary in their implementation and form factors. However, despite the absolute efficacy of these designs being questionable, their popularity makes it worthwhile to investigate finer details of implementation.

2.1.3 Limitations of popular designs of external displays

Firstly, as mentioned above, cultural differences have an undeniable impact on the interaction between automated vehicles and pedestrians (Chan A. N., 2009). Hence, cultural-specific symbols as local traffic signs or metaphors as abstract icons would be less favourable considering the regional cultural context. Needless to say, the interpretation time of symbols or anthropomorphic interfaces would be a lag for communication and decision making in such time-critical situations (Huang et al., 2014).

Single modality interfaces, such as visual interfaces, tend to be less effective. Multimodal interfaces combined auditory, tactile, and visual warnings enhanced alert system effectiveness as higher ratings of urgency, quicker response time. For example, a pedestrian alert system has utilized visual and auditory warnings to reduce collision risks (Merenda et al., 2017). To better emphasize the urgency of the warning and alert people, some researchers choose to employ multimodal interfaces in their studies, such as auditory and tactile (Politis et al., 2014) and auditory, visual, and tactile (Shim et al., 2015). Consequently, such multimodal interfaces are validated to reduce risks promptly but elicit more driver annoyance at the same time.

Another limitation of current studies on external warning design is overlooking the robustness of the designs in different situations. External displays should work effectively in poor light conditions and adverse weather conditions (e.g., snow and fog) for different types of road users (children, the elderly, and the handicapped). For instance, the complex urban city street lighting display patterns elaborated in the evening may interfere with road users' judgment within these designs and thus should not be overlooked. Moreover, colour designs may ignore the traffic safety regulation (Onatrio Highway Act , 1990) which limits colours utilized by commercial vehicles' lighting.

Critical factors such as noise, multiple pedestrians, and multiple lanes with vehicles have also not yet been properly addressed in most eHMI design prototypes evaluations. According to Rasouli et al. (2017), group size is a relevant factor for crossing decisions. However, it would be an extremely complex controlled factor for study as simulating participants brings a multitude of variation into a simulation (Colley et al., 2020).

Those limitations reveal the fact that, to some extent, most current simulation approach studies on eHMI design prototype evaluation results are not transferable to a real-world experience. Currently, there is no consensus regarding the eHMI design that should be used for crucial communication messages within uncontrolled road-crossing situations.

2.2 Assessment Methodology

Traditionally, road user behaviour research methods such as using driving simulators, naturalistic driving data analysis and self-reports have greatly contributed to the understanding of road user behaviour. However, it is often challenging to conduct studies for automated driving due to the unavailability of automated vehicles and traffic safety regulations, especially for SAE Levels 3-5. Hence, most of the current studies of automated driving were conducted using a simulator-based lab study (Hock et al., 2018) or the Wizard-of-Oz method (Currano et al., 2018) with instrumented vehicles on the road.

Another rapidly growing trend is the use of online sampling and crowdsourcing. Novel ideas of eHMI could be efficiently tested by using internet services to collect feedback from a large and diverse population of participants (Hock et al., 2018).

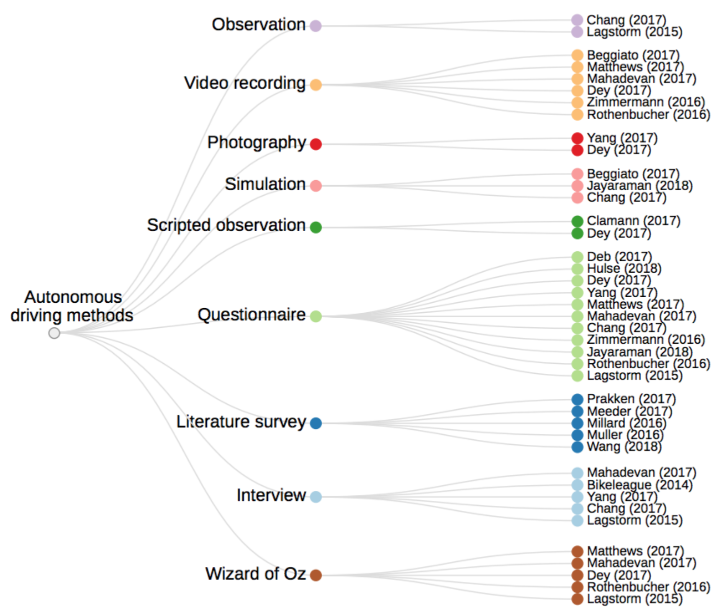

Figure 2: Data collection methods used in the pedestrian behavior studies involving autonomous vehicles. (Rasouli & Tsotsos, 2019)

Still, these methods also have several limitations. Table 1 shows a brief comparison of the methodologies.

Table 1: Assessment Methodologies of Autonomous Vehicles.

Methods	**Pros**	**Cons**
Crowdsourcing	Quickly, inexpensively, vast quantities and diverse data;	Previous experience and memory limitation bias; Mostly perceptual and limited behavioral skills;
Naturalistic Driving Study	Actual and long-lasting dynamic real-world behaviour;	Strict traffic regulation for high levels of driving automations; Limited perspective of other road users; Potential Observer bias;
Wizard-of-Oz Study	More accessible, rapid iterations;	Require significant training and preparations;
Simulation Study (Virtual-Reality / Driving Simulator)	Less risky, controllable, various perspectives of different road users, more flexible for designs and iterations, portable devices;	Potential risk of motion sickness;

We present each method’s potential limitations and merits in the context of traffic interaction between pedestrians and AVs in the following parts. These comparisons explain our choice in implementing a Virtual Reality device as a test tool in the study.

2.2.1 Crowdsourcing

The term “crowdsourcing” was defined as an online distributed problem-solving and production model in scholar research (Brabham, 2008). Crowdsourcing is firstly used to generate training data for machine learning algorithms, and commonly as a tool to evaluate and debug machine learning models (Ribeiro et al., 2016). Then, crowdsourcing has flourished as a research tool outside of the machine learning community as well (Vaughan, 2018).

Online crowdsourcing platforms, such as Amazon’s Mechanical Turk (MTurk), Prolific Academic, Qualtrics Panels, StudyResponse, Survey Monkey Audience and Witmart.as, provide researchers with a tool for quickly and inexpensively obtaining the vast quantities of diverse data needed. Through the platform, vast populations (e.g., hundreds) were recruited at a relatively low cost in a very short timeframe.

Online crowdsourcing allows us to harness the power of crowds to solve tasks that are notoriously difficult to solve with inadequate, non-random or non-independent opinions. The famous MIT moral machine (2018) is trying to gather a human perspective on moral decisions made by machine intelligence, such as self-driving cars. It also offers a solution for researchers to move beyond the biased samples and access more diverse and representative populations (Buhrmester et al., 2011; Roulin, 2015).

Although online crowdsourcing could gain a massive-scale dataset from diverse populations globally, it still has limitations. Firstly, performance on crowd-sourced assessments may differ because participants hav1)e previously been exposed to a similar experimental manipulation or measure. Furthermore, the majority of online crowdsourcing studies assessed a human's perceptual-cognitive skills rather than their executed behaviours. Therefore, the vast results of online surveys may be biased by perceptual as well as memory limitations but overlook the gap between perceptual and behavioral compliance.

I believe that after critically evaluating online crowdsourcing and integrating its use with other options it would lead to more comprehensive research results (Landers & Behrend, 2015).

2.2.2 Naturalistic Driving Study

Naturalistic Driving Studies (NDS) are a method for providing insight into the actual real-world behaviour of road users, unaffected by experimental conditions and related biases (Barnardet al., 2016). To record vehicle manoeuvres (such as speed, acceleration/deceleration, direction), internal/external conditions (such as road, traffic and weather characteristics) and other road users (conspicuity, distance, number, estimated age and gender), experimented driving entities are facilitated with unobtrusive data gathering equipment. The recorded data streams include Controller Area Network (CAN) messages, high definition video streams of the forward roadway, the vehicle's surroundings and various road users. In most cases, the gathered data through NDS lasted over several months and enough long distances, thus the datasets of experiences are dynamic (Dingus et al., 2006).

The most important concern for this methodology is that when driving assistance systems are shifting to driving automation (SAE Level 3 and above), there are more strict traffic regulation restrictions of AVs (DMV, 2020), which makes it difficult to interact with actual road users in the real world.

Moreover, NDS generally lacks the perspective of external road users' experiences. Opposite to perceptual assessment, the recorded data of pedestrians were solely behavioral. Another drawback about observations based on NDS is the strong observer bias, which can be caused by both the observers' misperception of the traffic scenes or their subjective judgments (Rasouli & Tsotsos, 2019).

To summarize, NDS could provide the most close-to-real-world long-term dynamic experience and dataset, yet it suffered the disadvantages of high costs, significant efforts in creating controlled stimuli conditions and the risk of violating road safety regulations.

2.2.3 Wizard-of-Oz Study

A good compromise between naturalistic driving and driving simulators is to use a Wizard-of-Oz study running on the road, where some functions of the vehicle are simulated (Ayoub et al., 2019). In general, the Wizard-of-Oz technique is common where the experimenters simulate the behavior of an artificial intelligent system to observe the reaction of subjects (Lagström & Victor, 2016). In the context of AVs study, implementing the Wizard-of-Oz technique, researchers wear costumes to disguise themselves as a car seat (Dey et al., 2017) or control the vehicle from a hidden place inside the vehicle that is not observable by the participants (Lagström & Victor, 2016) (see Figure 3).

(a) (b)

Figure 3: Examples of Wizard of Oz techniques to disguise the driver.

a) The driver is disguised as a car seat; and b) The driver is driving the car from a right-hand steering wheel while a dummy driver is sitting in the actual driver seat.

(Rasouli & Tsotsos, 2019)

Such a technique could establish a real-world experiment environment as long as participants are not aware of the “wizard”. Compared to a driving simulator study or NDS, the method is more accessible for gaining realistic road users’ interactions by not violating safety regulations but still has the merit of rapid iterations of design prototypes. Yet, the Wizard-of-Oz method also sparked concern for the additional cost of training a “wizard” as such a method usually requires significant preparation and training, so the simulated response is credible.

In addition, some researchers also criticize the efficacy of implementing such methodology as there is a risk that the “wizard” fails to maintain his/her behavioral consistency during the experiment, which means that participants will be aware of the trick (Bernsen et al., 1998).

Hence, although WOZ elicits much more realistic road user interaction information, the quality of the study results greatly depends on how feasible the simulations have been planned, trained, executed and iteratively evaluated (Bernsen et al., 1998).

2.2.4 Driving Simulator Study

Over the past ten years, there have been over three times the number of studies using a simulator-based lab study compared to those that employ naturalistic environments (Ayoub et al., 2019). The trend reveals the fact that compared to NDS, such simulator-based studies not only reduce the risks of the participants during driving but also reduce the cost of running the studies. Another critical reason for this trend is the difficulty of conducting large-scale naturalistic environment studies with the limited accessibility of automated vehicles on the roads, whereas implementing driving simulators could achieve an agile iteration for design ideations prototypes with a safe, reproducible, and controllable environment at hand.

However, the main challenge of driving simulator studies is that driving simulators may not be able to provide an ecologically valid driving experience, which raises the question of transferability, reliability, and the validity of the study results (Ayoub et al., 2019). Driving simulator study, as well as NDS, also only focus on experiences from the vehicles' internal users' views rather than external road users', which makes it a very weak option for conducting pedestrian communication studies.

2.2.5 Virtual Reality Study

Virtual Reality (VR) is a digital environment that allows users to experience and interact with the environment as if it were real (Jerald, 2015). Over half a century ago, VR has been used for training purposes in multiple fields (Lasse & Flemming, 2018), such as Health (Rizzo & Kim, 2003) and education (Dede, 2009). The most critical merit of VR is that it could establish an immersive environment that could be dangerous or life-threatening in the real world (Bailenson et al., 2008). Therefore, implementing this technique affords opportunities risky or even impossible in the physical world, such as autonomous driving entities with pedestrian interactions in a future traffic system, a controllable and feasible test tool.

Compared to the NDS and wizard-of-oz approaches, VR can be used to train and test interactions between humans and AVs in a realistic environment and without the physical risks of the real environments (Nascimento et al., 2019). Using virtual and controlled scenarios, it also makes it possible to evaluate human perceptions and behaviours at the same time. Moreover, VR maintains the possibility of observing conflicts, nuances or possibly even near-crashes/crashes without the potential biases of post-hoc reports.

Compared to driving simulator studies, VR studies provide a border view of the comprehensive traffic system user's experience rather than focusing on the vehicle's internal users. Using VR environments as research tools allows for greater flexibility for switching study scenarios and controlling the stimuli the user and the AV's algorithms are exposed to, thereby efficiently increasing the utilization of equipment and study space while reducing the maintenance cost (Blascovich et al., 2002).

In fact, VR has been widely accepted as a promising tool for studies addressing pedestrian behavior, pedestrian/vehicle interaction, and risk assessment in the last decade. Doric et al. (2016) conducted a two-way street crossing study in a VR environment with an acceptable level of realism. They detected that although pedestrians did not verbally estimate the "real" distance values, they assessed the risk appropriate as acting more carefully with reduced distances between passing cars. It was also pointed out that by utilizing VR headset, unseen feet and head rotation lag may induce insecurity and motion sickness. Schneider et al. (2018) and Maruhn et al. (2019) observed an overall underestimation of distance perception in virtual reality. However, Maruhn et al. (2019) demonstrated that distance compression has become less severe than in earlier study (Renner et al., 2013) as technological enhancements in VR techniques (Kelly et al., 2017). Bhagavathula et al. (2018) also validated that when the virtual environment is accurately modelled, several metrics of pedestrian response, such as intention to cross, perceived risk and safety of crossing and perceived distance of the vehicle, showed no significant difference with their responses in the real world.

Some other VR studies focus more on external communication concepts efficacy at crossing safety. Böckle et al. (2017) compared the pedestrians perceived safety and comfort by conducting a VR experiment on various communication features and claimed the attitudinal impact by adding interfaces. Deb et al. (2018) investigated pedestrians’ preferences for the visual and audible features or combinations based on their crossing behavior within a virtual environment. Researchers found that receptivity toward AVs and pedestrian vigilance in crossing increased with inclusion of external features. However, neither visual displays nor combination of visual and audible features were found significantly affecting crossing time or waiting time. Holländer et al. (2019) evaluated three external display concepts in a Virtual Reality (VR) based user study. The study included attitudinal measures and pre-crossing behavioral measures as decision time and error rate in two different traffic scenarios (i.e. One scenario is an AV approaching from behind a pedestrian and the other is where the AV approaches pedestrian from around a corner.). The results presented that external displays significantly reduced the crossing decision time of pedestrians, but overall evaluations of crossing behavior were still scarce. Interestingly, the designs of these traffic scenarios and concepts in experiments validated VR approach as a flexible and controllable test tool for pedestrian behavioral studies.

The use of VR for training AVs' AI algorithms also seems to be a promising and unexplored field. For example, Roberson et al. (2017) used a simulation and 3D engines for the training of deep learning approaches for object classification and thus indicated the viability of the simulation-only training approach for classifying real-world imagery. This kind of training can significantly reduce costs and the time to generate reliable data sets for AV image recognition and many other aspects of traffic participation that are difficult to study using traditional research.

Hence, our study intends to deal with the potential for using VR as an instrument to clarify the behavioral interaction between AVs (SAE Level 3 and above) external warnings and pedestrians crossing under different traffic conditions; to systematically understand the communication between AVs and vulnerable road users, which is difficult to study using traditional research.

2.3 Crossing Intention Vs. Crossing Behavior

It is noteworthy that the bulk of the empirical work on pedestrian-AV communication concepts has been conducted only with crossing intentions and attitudinal evaluations.

Deb et al. (2018) investigated pedestrians' preferences for the warning cues within a virtual environment. Researchers found that receptivity toward AVs and pedestrian vigilance in crossing increased with the inclusion of external cues. Bhagavathula et al. (2018) also validated the responses in the VR environment such as the intention to cross, perceived risk and safety of crossing and perceived distance of the vehicle, were not significantly different than responses in the real world. Holländer et al. (2019) evaluated attitudinal measures and pre-crossing behavioral measures within a VR environment. The study included attitudinal measures and pre-crossing behavioral measures, such as decision time and error rate. Still, these works did not explicitly examine the compliance between crossing intention and behavior.

Understanding intentions is undeniably crucial in predicting and interpreting pedestrian behavior at unmarked crosswalk crossing with AVs. However, it is equally important that people understand that behavioral intentions does not guarantee behaviors (Sheeran, 2002). Despite the considerable research on AVs' external displays improving pedestrians' perception of safety, communication comfort/preferences, and self-assessed intentions to cross, much of this work did not explicitly examine the compliance between attitudes/intention and behaviour.

The intention-behaviour gap exists as intentions get translated into action approximately only one-half of the time (Sheeran, Intention—Behavior Relations: A Conceptual and Empirical Review, 2002). Sheeran & Webb (2016) addressed important properties of intentions that might determine how well intentions predict behavior. These properties include intention temporal stability (Conner & Gaston, 2007; Cooke & Sheeran, 2013), strength, degree of formation, attitudinal versus normative control, and certainty. Other assumptions, such as that stereotypes or traits outside participants' consciousness (Bargh et al., 1996), could moderate the intention-behavior relationship as well.

In addition to above mentioned intention properties, competing goals as distractions and bad habits may also disturb goal pursuit (Sheeran & Webb, 2016). Forgetting to act (Einstein et al., 2003), missing good opportunities, and failing to engage in preparatory behaviors (Corradi et al., 2013), seem quite likely at the stage of goal pursuit initiation (Sheeran & Webb, 2016), especially when such opportunities are brief or infrequent (Dholakia & Pbagozzi, 2006). As jaywalking is a typical time-critical situation, the study of how intention realization facilitated by external warning displays would be valuable. If there exists any intention-behavior gap hindering the effective translation for pedestrian safety crossing, future study should consider applying helpful approaches for closing the intention-behavior gaps within the pedestrian-AVs communication.

In sum, although there has been a good deal of research in identifying intentional and attitudinal evaluations on AVs external communication concepts at crossing, more research is required to examine and explore their effects in the intention-behavior compliance.

2.4 Conclusion

Besides being a dynamic control task, driving is a social phenomenon and requires interactions between all road users involved to ensure the flow of traffic and to guarantee the safety of others (Rasouli & Tsotsos, 2019). Over the last few decades, existing studies of AVs' interaction mechanisms with external road users have provided considerable communication ideations to fill the void of interaction by replacing human drivers. The experiment evaluation methods review, and discussion set the theoretical and empirical basis for our approach to conduct both perceptual and behavioral assessment of warning design patterns at unmarked crosswalk crossing in a virtual environment. Moreover, our comparison between these two perspectives was trying to dig into the overlooked gap between intention and behavior in a line of studies.

Chapter 3
Assessment of Designed External Warning Patterns

This chapter proposes an approach by using Virtual Reality technique to assess external warning design facilitating AVs to communicate with pedestrians at crossing without marked crosswalk. Herein, I establish a high-fidelity simulated local street in a virtual environment and clarify a detailed plan for assessing our external warning designs from both attitudinal and behavioural aspects. The content begins with details of replicating the local street using virtual simulation. Then our designed external warning patterns would be introduced. After that, the evaluation approach of both perceptions and behaviours are presented. In the last part, I also introduce a distractor task implemented for pedestrians' auditory control of traffic sound. Together, this chapter elaborates a preliminary view of the exploration of communication mechanisms between pedestrians and AVs facilitated by designed external warnings through an immersive experience with VR.

3.1 Experimental Environment Design

This part illustrates the simulated experimental environment. The characteristics of the environmental factors (the street characters and peripheral environment) and how they are replicated in the virtual environment are introduced. Colley et al. (2019) highlighted the need to include traffic factors (e.g., speed, flow, density, road pattern and etc.) in VR studies on external communication concepts. I hope such an enhanced close-to reality environment could contribute to explorations of using a wireless VR device for examining the efficacy of external visual warning designs.

3.1.1 Real Environment Vs. Virtual Environment

Philip Street is a typical two-way city street connecting the University of Waterloo campus and a dense student residential area, as shown in Figure 4. The 620.51m-length street (Road data, 2020) consists of two bi-directional driving lanes (each lane width is between 3.0m to 3.7m) and a left turn lane in between (the width is between 2.7m to 3.3m) as a waiting area while crossing (Service, 2017). The street does not have any marked crosswalks and has seen pedestrian safety issues because of a high volume of students and its proximity to the campus (for an accident news report, see (Record, 2018)).

Figure 4: University campus and Philip Street (circled).

The virtual environment was a scene depicting the setting that served as the real environment. Figure 5 shows a comparison between Philip Street and the VR environment.

Figure 5 Real environment (Up) and virtual environment (Bottom).

Most street (e.g., number of lanes, position of sidewalk and space scales) and building details (e.g., key features of buildings' appearances, locations, distributions of trees and lawns) were built to reflect the real environment.

I made two exceptions to reduce potential confounding to the study: First, I intentionally excluded safety islands from the VR environment to make sure pedestrians always crossed at an uncontrolled location. Second, I did not implement faded street paint (e.g., lines), which was occasionally seen in the real environment. This would ensure pedestrians to cross the street more consistently.

The other details and procedures of creating the virtual reality environment is described in Chapter 4.

3.1.2 Fleet Speed

I implemented an infinite fleet of AVs traveling in both lanes and towards opposite directions. The infinite fleet was designed for simulating real-world traffic. Yet, it helped us run a controlled experiment as in each scenario, all AVs in the fleet moved at the same speed and the gap between every two AVs was also consistent.

The vehicle speed is the constant driving speed of the fleet. A recent study (Bertulis, 2014) has shown that the vehicle speed was the major factor in the driver’s yielding rates and thus it could reflect the vehicle's driving intentions for pedestrians' crossing decisions (Sun et al., 2015). As the default speed limit on city roads in the province the experiment conducted in is 50 km/h if there are no posted speed limit signs on the tested road (Rules of the Road, 2017), the fleet speed of 50km/h and 35km/h were controlled as vehicles’ distinct speeds for the study.

3.1.3 Gap distance

The fleet gap distance refers to the space between the former vehicle and the subsequent vehicle in the fleet, which could, in our study, prevent the pedestrian waiting for the only vehicle to pass. The fleet gap addressed in this study is also to investigate to what extent traffic density affects pedestrian crossing decision-making (VL Knoop, 2018). Traffic flow density, defined as vehicles per km (veh/km, sometimes additionally reported per lane), generally ranged from 20 to 165 veh/km (Greenberg, 1959).

3.2 Warning Patterns Design

As stated in Chapter 2, although popular designs of external displays promote inspirations, the challenge is how to make sure such displays convey the vehicle's intentions directly, simply, and unambiguously in a timely manner. To minimize the effect of cultural differences and to shorten the time for interpretation, symbols, metaphors and text messages as visual cues were excluded in this study.

3.2.1 Flashing Pattern

There is evidence to suggest the human visual system is hard-wired to pay special attention to flashing rather than steady lights, as an innate rather than learned response (Teller, 1979; Teller, 1986; Jouen, 1990). A more recent study also validated that pedestrians might be more attentive to communication signals that are flashing (Li, et al., 2018).

Additionally, flashing the headlights of a car is a common explicit communication tool used by drivers to negotiate with pedestrians and other road users (Dey et al., 2020). Emergency vehicles also typically use various forms of rotating or flashing lights to warn other traffic of their presence (Dey et al., 2020). Given this existing context, pedestrians would have certain expectations or intuitive associations regarding the meaning behind such warning cues.

Considering these factors, the flashing patterns were assumed as visual cues minimizing the effect of culture difference as well as timely attract pedestrians' attentions.

3.2.2 Colored Lighting

Besides flashing patterns, colored lighting is another effective approach to draw pedestrians' attention in a short time period. However, some recent research suggested disagreements of preferences upon colors for communicating a yielding message to support the interaction between AVs and pedestrians through an external human-machine interface (eHMI) (Li et al., 2018; Dey et al., 2018). Furthermore, traffic regulations do not permit certain colors to be used by commercial vehicles' lamps. Considering the color-blind population and the difficulty of identifying color in urban environments, our study would not investigate color pattern effects on pedestrians crossing.

3.2.3 Flashing Rate

There is one particular feature of warnings – their urgency – which is of practical importance in that the urgency of the warning should relate in some systematic way to the hazardousness or risk of the situations (Edworthy and Adams, 1998). Warnings can be said to be appropriately mapped when the rank ordering of the urgencies of the warnings with which situations are associated is positively correlated with the rank order of the urgencies of those situations. Figure 6 demonstrates appropriate mapping. In practical situations, the urgent situations may not have appropriately urgent warnings, and more in keeping with Figure 6 (b).

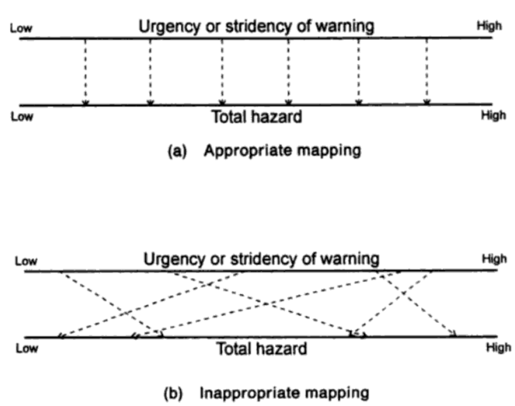

Figure 6. Urgency Mapping (a) Appropriate (b) Inappropriate.

(Edworthy & Adams, 1998)

Much effort has been put into the elucidation of the effects of flashing rates in urgency mapping. Politis et al. (2013) presented the warnings consisted of pure colours with a varying flashing rate depending on the level of urgency. In line with (Baldwin, et al., 2012; Lewis & Baldwin, 2012), pulse rate could be varied to signify escalating urgency across all modalities. So, using flashing rate as parameters allowed us to match these warnings to levels of risk actually present within the situations.

An earlier research shows that hazard perception increases with increasing flashing rates, with a particular impression of urgency achieved at 4 Hz (Van Cott, 1972). The turn signals installed in almost every vehicle blink on and off, or "flash", at a steady rate of 1–2 Hz (UN Regulation, 2014). Therefore, the momentarily flashing white 4 Hz and 2 Hz headlights, as shown in Figure 7, were employed as our study's designed visual warning rates.

Each AV in experimental scenarios was equipped with a pair of symmetrical external displays attached to the headlight position. Depending on the scenario, the external displays were set to present different warning rates or be turned off implying traffic situations (e.g. vehicular speed and traffic density). The varying frequencies is designed to detect whether they would appropriately map to situational hazards and so properly translate into safer crossing behaviours.

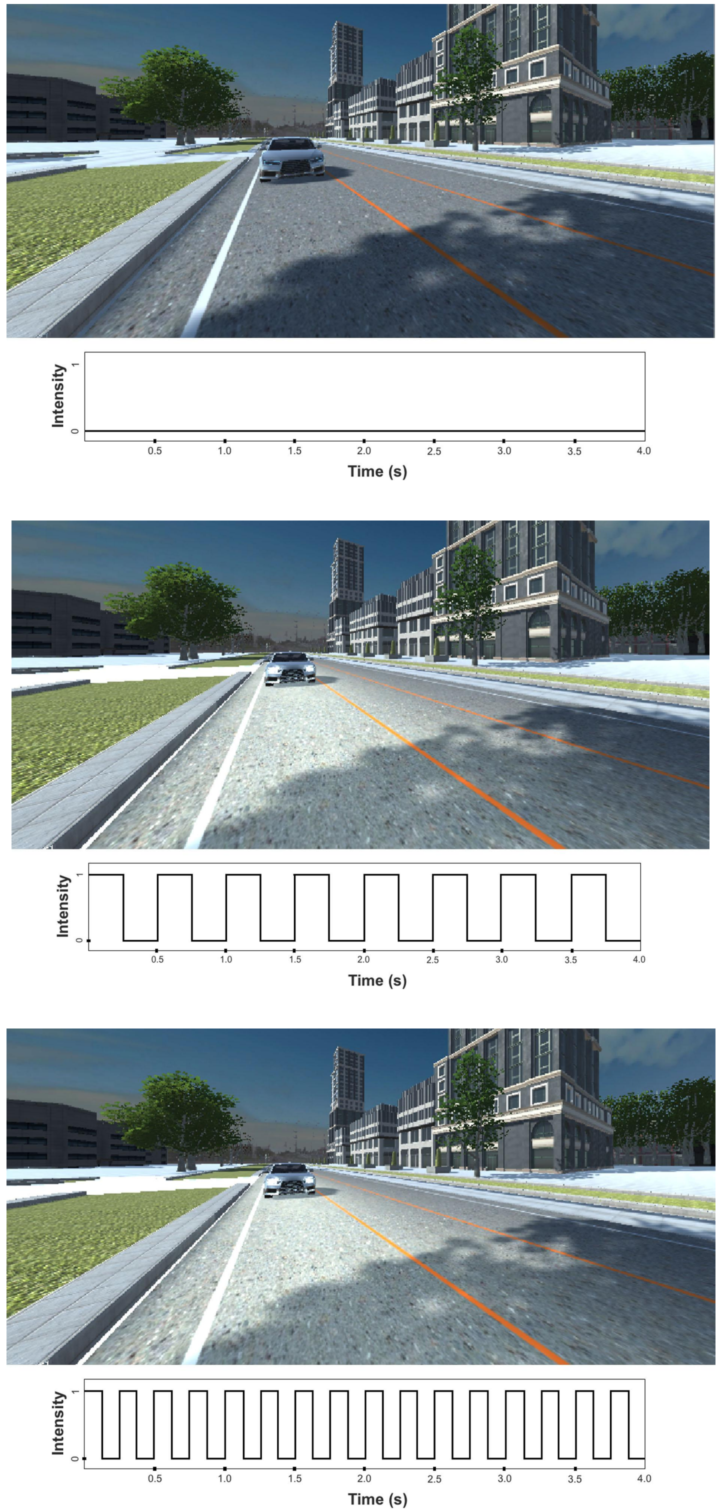

Figure 7 "No pattern" (Up), "2 Hz pattern" (Middle) and "4 Hz pattern" (Bottom).

3.3 Assessment Design

This discussion of assessment includes both attitudinal and behavioural evaluation metrics and techniques.

3.3.1 Attitudinal Assessment

Perceived hazardousness.

Perceived situational hazardousness assesses how hazardous a subject perceives a situation to be (Chan A. N., 2009). Scores range from 1 –100, where 1 would be considered the "Not Hazard" and 100 the "Most hazardous". The participants were explained the scores by the researcher before assessing. When participants drag the slider, the label would change accordingly.

Figure 8: Rating slider for the Hazardousness of situations.

Perceived urgency.

Perceived urgency of warning patterns is a scale that evaluates the perception of warning urgency, rated after each crossing task. The rating could also be considered as a preliminary assessment of attitudes on warning compliance when compared with the perceived hazardousness ratings. The full range varies from a minimum of 1 (Not urgent) to a maximum of

100 (most urgent) (Wickens et al., 1998).   When participants drag the slider, the label would change accordingly.

Figure 9: Rating slider for perceived urgency of warnings.

3.3.2 Emotion Ratings

Various traffic tasks involve complicated communications which can elicit emotions. These elicited emotional responses typically involve changes in several response systems, including perception, feelings, expressive behavior, and peripheral and central physiology (M'bailara et al., 2018). Emotional reactivity can be seen as the production of a specific affective state in response to a given stimulus and the regulation of this affective state and emotional behavior (Henry et al., 2012). The emotional reactivity can bring both positive and negative effects on traffic safety. However, the involvement of emotional factors in traffic system safety has received little attention. Some recent research has discovered that social awareness of AVs can improve robustness and reduce errors (Müller, 2016). Social awareness should help AVs to interact with other road users' emotional states and behaviours in critical situations.

Consequently, I herein employ a simple emotional assessment, "Politeness of warning," to study emotional reactivity and its potential impacts in response to the designed warning patterns.

Politeness of warning.

Politeness of warning rates perceived politeness of designed warnings, whose full range is also from 1 (i.e. the Not Polite) to 100 (i.e. the Politest). When participants drag the slider, the label would change accordingly.

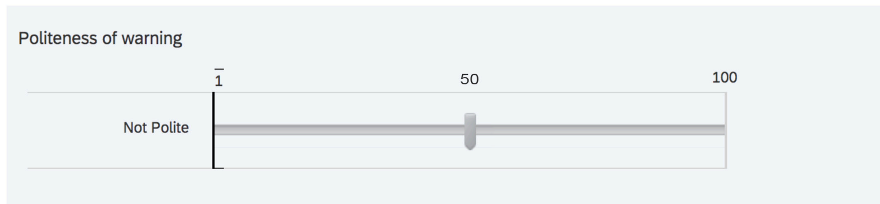

Figure 10: Rating slider for politeness of warnings.

3.3.3 Behavioral Measures

Time-to-arrival (TTA) ratio.

Time-to-arrival is defined as the duration of time it takes an approaching vehicle to reach the pedestrian who is crossing from the near lane sidewalk (Terwilliger et al., 2019; Sun et al., 2015). Accordingly, TTA was decided by both the pedestrian crossing behavior (i.e., when the pedestrian took a step to start crossing) and the traffic (i.e., speed, acceleration and the gap between vehicles).

Since I manipulated vehicle speed and gap across scenarios, I introduced TTA ratio to describe crossing behavior. TTA ratio was calculated when the participant stepped into the near or far lane, using the following formula:

$$\text{TTA Ratio} = \frac{\frac{|S_p - S_{v(n)}|}{V_v}}{\frac{G}{V_v}}$$

n – the approaching vehicle closest to the participant when they stepped into a specific lane. Let it be the nth vehicle after the scenario started.

Sp – position of the participant along the street.

Sv(n) – position of the nth vehicle along the street.

G – gap between every two vehicles in the feet, an independent variable.

Vv – vehicle speed, another independent variable.

There were near and far lane TTA ratios. The maximum of TTA ratio was 100%, indicating that the participant started crossing immediately after a former vehicle in the fleet drove passed. The minimum of TTA ratio would be 0%. Though unlikely, it showed that the participant started to cross and was immediately hit by the closest approaching vehicle.

Safety Margin Time (SMT) ratio.

SMT refers to the time it takes for the vehicles to reach the pedestrian's position after the pedestrian has passed safely to the other side of the lane. Similarly, I calculated SMT ratio to eliminate the influence of vehicle speed and gap that were varied across scenarios. SMT ratio was calculated when the participant completed crossing the near or far lane:

$$\mathbf{SMT\ Ratio} = \frac{\frac{|\mathbf{S_p} - \mathbf{S_{v(m)}}|}{\mathbf{V_v}}}{\frac{\mathbf{G}}{\mathbf{V_v}}}$$

m – the approaching vehicle closet to the participant when they finished crossing a specific lane. Let it be the mth vehicle after the scenario started. Apparently, m > n.

Sp – position of the participant along the street.

Sv(n) – position of the mth vehicle along the street.

G – gap between every two vehicles in the feet, an independent variable.

Vv – vehicle speed, another independent variable.

SMT were distinctly influenced by factors as vehicle speed, gap and pedestrian walking speed. In the most extreme case, pedestrians may make a highly risky decision that could use up this buffer zone to cross with the minimum gap: the gap that allows them to pass with a safety margin time of zero (SMT ratio was 0). This distance is the vehicle's stopping distance. Therefore, the ability to make safe estimations of vehicle stopping distances is also important for safety (Erp et al., 2015). Theoretically, the maximum of SMT ratio would be equal to TTA ratio of the same lane, which was calculated earlier in the scenario. It won't be possible in reality because it took time for the participant to walk across the lane.

Crossing time.

The complete crossing time is made up of near-lane crossing time, mid-lane-area waiting time, and far-lane crossing time. As the crossing distance was fixed in the virtual environment, the crossing time should be inversely proportional to pedestrian's walking speed. Earlier work shows that the road capacity decreases with the square of the pedestrian crossing time, hence short crossing time is desired (VL Knoop, 2018).

Lateral displacement.

Lateral displacement is defined as the lateral spatial differences between the pedestrian's starting position and end position on the other side of the lane. It can represent how pedestrians' trajectories are affected by traffic or warning patterns.

Step-back rate.

The pedestrian can step back if he or she does not feel confident about incoming vehicles motions and not safe to cross. The quantity of participants' step-back for each lane upon every scenario was compared and measured as step-back rate. It indicates how warning patterns are perceived and thus change pedestrians' behaviours in the context.

3.4 Distractor Task

An observation study of pedestrians crossing urban intersections found that 19 percent pedestrians were engaged in a distracting activity while crossing. Nine percent of them were wearing headphones and 8 percent were using a mobile device (talking, texting, or looking down at it). Thompson et al. (2013) discovered that nearly 30 percent were engaged in a secondary task while crossing. In an observation of pedestrians crossing high-risk intersections, about 11.2 percent were using a portable music player with headphones, 7.3 percent were texting, and 6.2 percent were talking on a cell phone. Basch et al. (2014) also conducted an observation research of pedestrians in a busy urban area. Over one-quarter of pedestrians were using electronic devices while crossing during the 'walk' (29 percent) and 'don't walk' (26 percent) signals. Sixteen of them were using headphones as the most common activity (16 percent), followed by looking down at a cell phone (7 percent), and talking on a cell phone (5 percent).

A substantial number of pedestrians (about 20–30 percent) exhibit distracted walking by listening to audio content or using a portable device (Basch et al., 2014).When auditory information is critical to the safety of the pedestrian, such as the sound of vehicles approaching or rolling to a stop, listening to music may be potentially hazardous. Other traffic noise, which may implicitly convey important information such as traffic flow (Calculation of road traffic noise., 1988), percentage of heavy vehicles, traffic speed, road surface, and barriers, is an important factor for auditory stimuli and evaluating pedestrians' decision-making (de Waard et al., 2011). Thus, auditory stimuli such as traffic noises and sound of oncoming vehicles should be cautiously controlled in a high-fidelity simulation study of visual warning assessment.

The secondary task technique has been used as a detection and measure of the amount of "cognitive capacity" dedicated to primary tasks that are of attention, effort, elaboration and a continuous process of perception, cognition, and action (Owen, 1991). Pedestrians' crossing tasks require three stages of processing, as illustrated in Figure 11 (Wickens et al., 1998). Secondary task performance should decrease as the capacity demands of the primary crossing task increase.

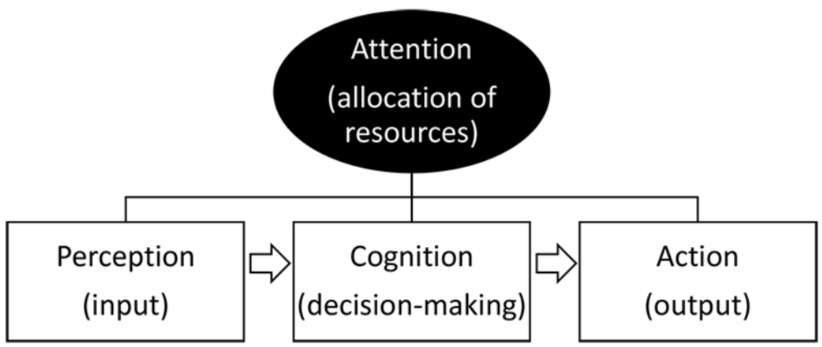

Figure 11: The simplified three-stage information processing model.

Therefore, I have designed an auditory task intending to compare the cognitive demands of pedestrian crossing under various kinematic and visual warning conditions. Moreover, the designed auditory task is expected to reduce boredom, and alleviate learning effects from repetitive crossing tasks, and meanwhile cancelled the impact of peripheral traffic sounds affecting pedestrians' decision-making.

In response to the study results of Neider et al.(2010) that suggested that listening to music may not be complex enough, the complexity of the auditory task was manipulated by the introduction of a daily news podcast short clips in a noise-cancellation earphone, instead of music, at the primary crossing task in our study. Participants were asked questions about the content of the news podcasts to see if there is any significance between different kinematic scenarios and warning patterns on auditory task performance. This provides some extra data for the preliminary evaluation of external warning design and potential detrimental effects of auditory tasks on distracted pedestrian crossing performance and safety.

The audio content trans and questionnaire is listed in Appendix A. 2.

3.5 Conclusion

In this chapter I proposed an approach for evaluating designed external warning patterns facilitating AVs and pedestrian’s communication at crossing. First, I chose a real city road in the local area where uncontrolled crossing happens frequently, and I planned to replicate it in high-fidelity virtual reality as the experiment environment. After that, I defined assessment metrics for the study in attitudinal, emotional, and behavioural perspectives based on empirical and theoretical studies. Finally, I explained the necessity of auditory control in the traffic visual warning studies and introduced the design of an auditory task in our study.

Chapter 4
Experiment: External Warning Design at Uncontrolled Crossing

This chapter follows the design approach in Chapter 3. First, we introduce the structure of design warning patterns deployed and implemented in the VR platform using Unity and a few asset packages. After that, we present our hypothesis and objects of this experiment. Then, we discuss the apparatus, procedure, and the results of the experiment.

We have examined both attitudinal and behavioral assessment of designed warning patterns aiming to facilitate communication of AVs' intentions at an uncontrolled crossing to improve communication efficiency and pedestrian safety. We also examine the attitudinal and behavioral compliance based on these results.

4.1 Designs in Virtual Reality

Unity was used to build the virtual reality environment for our study. Unity provides users with the ability to create experiences in both 2D and 3D, and the engine offers a primary scripting API in C. Editors can design, buy, or import digital assets like forests, sound effects, and create the logic guiding how all these elements interact with players. Unity is also a very common Game Engine used to develop VR experiences. It has native support for Google Daydream and can be installed with Google VR SDK for extending the Unity Event System to handle the input from its headset controller.

Daydream is a high-quality VR platform developed by Google. Lenovo's Mirage Solo headset is the first standalone headset running on Google's Daydream platform. Its highlight feature is support for Google "WorldSense," an inside-out positional tracking technology (Matney, 2018). The headsets are packaged with a wireless controller. This controller can be used for interacting with the virtual world. The controller has a touch pad, two circular buttons (one functioning as a home button and one functioning as an app-specific button), and two volume buttons. In a review of the Google Daydream View, users complemented its comfort as it offers a lighter and more portable experience than other VR headset competitors. The improved tracking system, self-contained and wireless headset, great touch controllers, comfortable design for long-time wearing experience, and affordable price made us decide to use it in the study.

Figure 12: Lenovo's Mirage Solo headset and the controller (Savage, 2018)

To achieve a high-fidelity replicated local street, we employed a few packages downloaded from Unity Assets Stores, which is a growing library of Assets provided by 3D assets designers, to replicate Phillip Street's details from Google Street View.

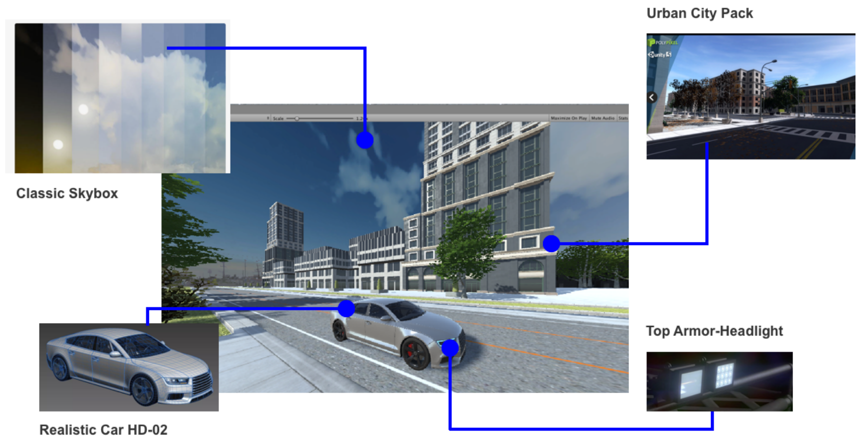

Figure 13: Packages downloaded from Unity Assets Stores.

An asset package named “Urban City Pack” was downloaded, imported, reassembled to simulated peripheral buildings, houses, tresses, lawns and sidewalks. Another critical material utilized in the simulated environment was a self-painted road configuration material mimicking Philip street’s color patterns and configurations. For creating a natural daytime light effect, we also added Skybox from Unity asset store named “Classic Skybox”. The Skybox could be flexibly extended to different weather or daytime for future study. For vehicle modeling, we utilized a common sedan model from the asset named “Realistic Car HD-02”. To test our designed external warning, we downloaded and modified a LED light asset “Top Armor-Headlight” to the vehicle models’ headlight position. Then, specifically, when a certain situation was triggered, such as when a pedestrian started crossing the street, the designed warning pattern was being activated.

Our script for controlling each scenario’s flashing patterns’ and traffic kinematics configurations was developed with Microsoft Visual C#. More details of the stimuli are discussed in the next section.

4.2 Experiment

4.2.1 Experiment Design

This study was a 2 (Speed: 35 km/h vs. 50 km/h) × 2 (Gap Distance: 60 m vs. 100 m) × 3 (Warning Pattern: No Pattern vs. 2 Hz vs. 4 Hz) repeated-measures design. The experiment lasted about 60 minutes, and the order of trials was randomized. The participants experienced the simulated city street environment and interacted with a two-way upcoming autonomous fleet to safely cross the street.

The AVs were travelling with four sets of kinematic motion combinations in speed and gap distances (i.e., 35 km/h × 60 m, 35 km/h × 100 m, 50 km/h × 60m and 50 km/h × 100 m) and presented three sets of designed warning patterns (i.e. No pattern, 2 Hz and 4 Hz flashing rates). In this study, the vehicular speed and gap would not change no matter how participants behave. Each scenario started with the street fully occupied by AVs, and one of them just passed the participant's start point next to the near lane. Across all scenarios, the AVs could travel at different speeds and maintained different gaps.

During each trial, the participants listened to a one-and-a-half-minute long podcast through a pair of noise-cancellation wireless earphones (i.e., Air Pods) for the purpose of distracting and cancelling sound of traffic (Colley et al., 2020).

Table 2 shows a total of 12 scenarios designed for the experiment. The dataset tracking participants' positional movements, traffic situations were later compiled and analyzed.

Table 2: List of Scenarios.

Scenario	Speed (km/h)	Gap (m)	Pattern Frequency
Fast-Dense-N/A	50	60	n/a
Fast-Dense-2Hz	50	60	2 Hz
Fast-Dense-4Hz	50	60	4 Hz
Slow-Dense-N/A	35	60	n/a
Slow-Dense-2Hz	35	60	2 Hz
Slow-Dense-4Hz	35	60	4 Hz
Fast-Light-N/A	50	100	n/a
Fast-Light-2Hz	50	100	2 Hz
Fast-Light-4Hz	50	100	4 Hz
Slow-Light-N/A	35	100	n/a
Slow-Light-2Hz	35	100	2 Hz
Slow-Light-4Hz	35	100	4 Hz

4.2.2 Hypotheses

I summarize our hypotheses and study objects as follows:

H1: More intense traffic (i.e., higher vehicular speed and narrower gap distance between AVs) would be perceived as more hazardous by pedestrians and result in more risky crossing decisions and behaviors.

H2: The perceived urgency of warning patterns would increase with higher flashing rates and elicit more vigilant pedestrian crossing decisions and behaviors.

H3: The perceived politeness of AV's intention communication would increase with the existence of warnings compared to no warning, but a higher flashing rate might be regarded as impolite by pedestrian.

4.2.3 Apparatus

Participants were asked to wear a wireless standalone VR headset (Lenovo Mirage Solo) and experienced the virtual environment in a 3300 sq. ft. indoor open space at University of Waterloo campus.

VR Headset.

The textured graphics with 1280 x 1440 resolution were projected onto the built-in 5.5-inch screen, providing participants with 180 degrees of immersive visual imagery of the virtual environment. The viewpoint of the scene was adjusted for each participant's eye height and they viewed the scene binocularly.

Start point.

Participants' starting position was 2 meters back from the near lane (see Figure 14). The vehicle model utilized in the virtual environment is a typical sedan.

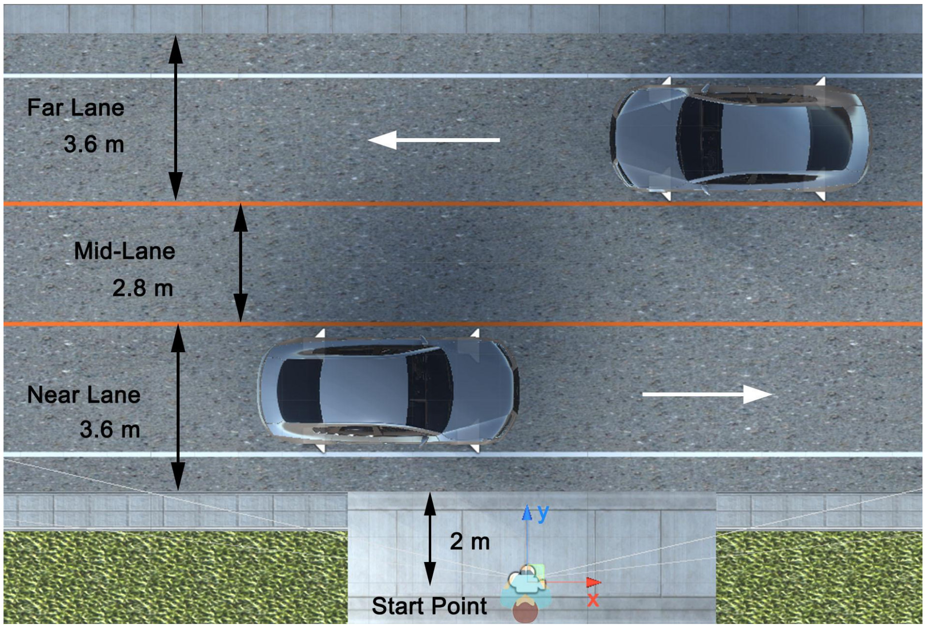

Figure 14: Top view of the experimental street and peripheral environments (White arrows refer to driving directions).

Recalibration

With eighteen participants (seven males, eleven females, with the average height of 168.91cm), the apparatus and scenarios were tested (the open space chosen for study, the questionnaire, the projected monitor and the subjects' acceptance of the 12 trials traffic situations) before formal experiments. Based on their average height and step length, we recalibrated the street width so that it would be more accurate to the real street configuration based on their feedback on spatial issues in virtual reality and real-world experience. We also adjusted the fleet gap distances from 50 to 60 metres in dense scenarios so that the subjects would not feel too risky to cross.

4.2.4 Procedure

The experiment involved 12 trials in randomized order and lasted about 60 minutes.

Before all experimental sessions, participants were asked to fill out a questionnaire on their own experiences, background knowledge and overall preferences between the designs on AVs' external displays for warning. The questionnaire asked the participants to comment on familiarity with the tested local street, as well. This was included because if recruited participants' have experience crossing this street, the results could reflect on a real crossing experience to some extent. The sample questionnaire is listed in Appendix A.1. The questionnaire took approximately 10 to 20 minutes to complete.

Before experimental trials, participants put on the VR headset device to familiarize themselves with the simulated city street environment and improve distance judgments as warm-up training until they felt comfortable and ready to start the formal experiment (Kelly et al., 2013). After the warming-up trials, the virtual fidelity was evaluated by requesting each participant to identify which street was in the environment.

During the formal experiment sessions, the participants experienced the simulated city street environment and interacted with a two-way upcoming autonomous fleet. The AV was travelling with four sets of kinematic motion combinations in speed and gap distances (i.e. 35 km/h × 60 m, 35 km/h × 100 m, 50 km/h × 60 m and 50 km/h × 100 m). They also presented three sets of designed warning patterns (i.e., No pattern, 2 Hz and 4 Hz flashing pattern).

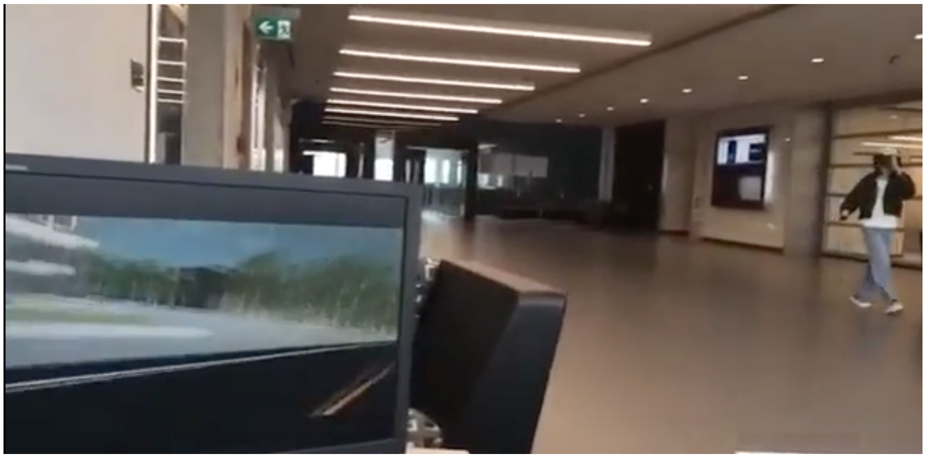

Figure 15: Participant wearing the VR headset in Study.

Then the participants were asked to walk across the street when they feel safe and/or appropriate to do so in the immersive world. Participants were told that there would be some external vehicular external warnings implying traffic situations. And they could wait in the mid-lane as a safe area before crossing the far lane. The participants' primary task was to cross the street safely.

In order to cross the street in the simulated environment, the participants need to physically walk across the room, which is approximately 12-15 meters in length. A research investigator walked alongside to ensure their safety.

During each trial, the participants listened to a one-and-a-half-minutes long recent news podcast through a pair of noise-cancellation wireless earphones.

After the completion of each trial, the participant was asked to take off the HMD to assess its situational hazards and perceived urgency of the warning patterns. Participants were firstly asked to assess through a rating slider on a tablet immediately after each trial. The rating scale of 1 to 100 was based on Hellier et al. (1993) urgency assessment method. Participants moved their finger to drag the slider and click the “Next” button to complete a rating.

After completing the attitudinal assessment task, participants were required to answer a question about an explicit detail mentioned in the first 30 seconds of the played podcast. Participants were asked to choose the correct answer from three selections as quickly as possible.

Participants were also asked to share opinions on when they decide to cross the street in the VR environment, comment on the effectiveness of each warning design, and were given the opportunity to share their final thoughts, preferences, comments, or concerns.

4.2.5 Participants

This study was approved by the University of Waterloo ethics committee. All participants were recruited from the university community, and each received $15 Canadian for approximately 60 minutes of their time. All participants indicated that they had recent experience crossing the Phillip Street. Participants also reported their experience listening to audio content (i.e., podcasts, radio, and music). Each participant read and signed a consent form before the start of the experiment.

The minimum number of 13 participants required was determined by an a priori power analysis (Gpower: Faul and Erfelder, 1992) with a power of .80 and an alpha of .05.

Fourteen participants completed this study. Their ages ranged from age 20 to 38 years old (M = 26.64, SD = 4.70). All participants have recognized the virtual street as Phillip Street. Thirteen participants reported having crossed the studied street in the real world within the last seven days, and seven reported their last seven days' crossing frequencies of this street were daily or more than once a day.

4.2.6 Results

Data analysis of the questionnaire, attitudinal ratings, and the dataset of behaviors are discussed as follows:

Questionnaire questions

Twelve of fourteen participants reported having the habit of listening to audio content (e.g., podcasts, radio, audiobooks, or music) while walking/jogging on the street, and nine reported this habit as "frequently," "often" or "always."

Table 3: Participants' habit of listening to audio contents.

How often do you listen to Podcasts/ Radios /Audio Books / Music while walking (jogging) on the street alone?	Participants' Responses
Always (f > 90%)	14.3%
Often (90% > = f > 60%)	21.4%
Frequently (60% > = f > 30%)	35.7%
Occasionally (30% > = f > 0)	21.4%
Never (f = 0)	14.3%

Half participants reported have no experience with AVs before, while only four of all participants have no Virtual Reality experience. Participants' preference for future AVs having distinctive design from conventional vehicles implies the reason of the prevalence of novel external warning displays design ideas. Meanwhile, it reflects the importance of interaction with AVs.

Table 4: Preference for future autonomous cars designs.

In your opinion, how would you compare the design of future autonomous cars and the design of conventional cars?	Participants' Preference
There should be a clear distinction between the design of future autonomous cars and the conventional cars.	42.9%
The more similar the better.	7.1%
The ability to interact with drivers/passengers in the autonomous car.	35.7%
It doesn't matter.	14.3%

Perceived hazardousness.

The urgency rating scores were recorded by the researcher through online survey tool Qualtrics via a tablet and exported to a summary table in Microsoft Excel.

The assumption of sphericity was met, $p > .05$. There was a significant speed × warning interaction effect, $F(2, 22) = 3.99$, $p <. 05$, $f = .60$, a significant speed main effect, $F(1,11) = 18.46$, $p = .001$, $f = 1.30$ (see Figure 16), and a significant gap main effect, $F(1, 11) = 28.42$, $p < .001$, $d = 1.60$. No other effects were significant, *p*s > .05.

Overall, participants indicated higher perceived situational hazardousness as the vehicle speeds increased and as the gap distances decreased. Figure 16 shows the two-way speed × warning interaction. Further analysis of the interaction effect showed that the differences in perceived situational hazardousness between 35 km/h and 50 km/h speeds was amplified when the flashing pattern was 4 Hz.

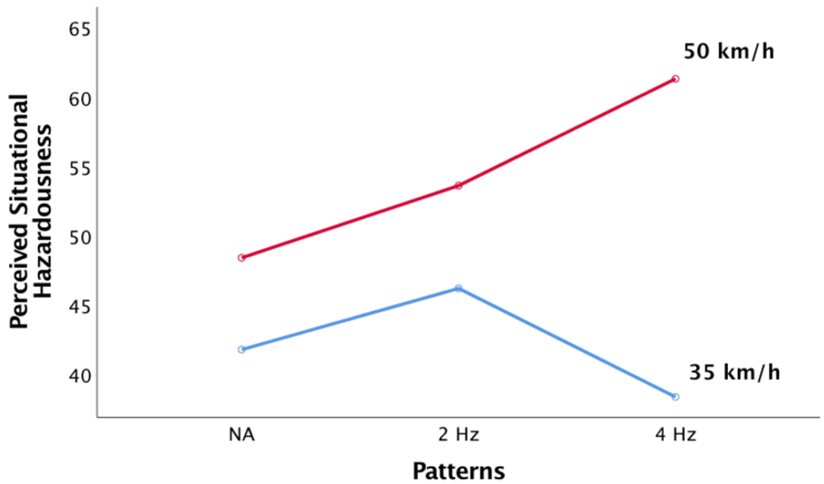

Figure 16: Line chart of Perceived Situational Hazardousness Pattern × Speed interaction (from SPSS)

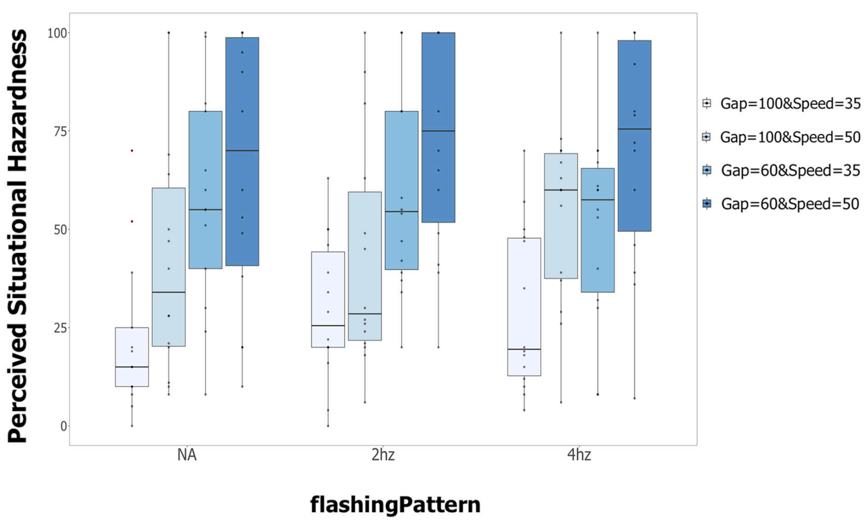

Figure 17: Boxplot of Perceived Situational Hazardousness Pattern × Speed (from R studio)

Perceived urgency.

The assumption of sphericity was met, $p > .05$. There was a significant warning main effect, $F(2, 22) = 30.12$, $p < .001$, $f = 1.66$,and a significant speed main effect, $F(1, 11) = 6.89$, $p < .05$, $f = .79$, as shown in Figure 18. No other effects were significant, *ps* > .05. Post-hoc tests using Bonferroni corrections demonstrated that the 4 Hz ($M = 49.75$, $SD = 3.67$) warning did not significantly differ from the 2 Hz ($M = 41.77$, $SD = 3.67$) warning, but both warnings were perceived as more urgent than when there was no warning pattern ($M = 6.83$, $SD = 2.78$). Furthermore, as the vehicle speeds increased, perceived urgency ratings also increased.

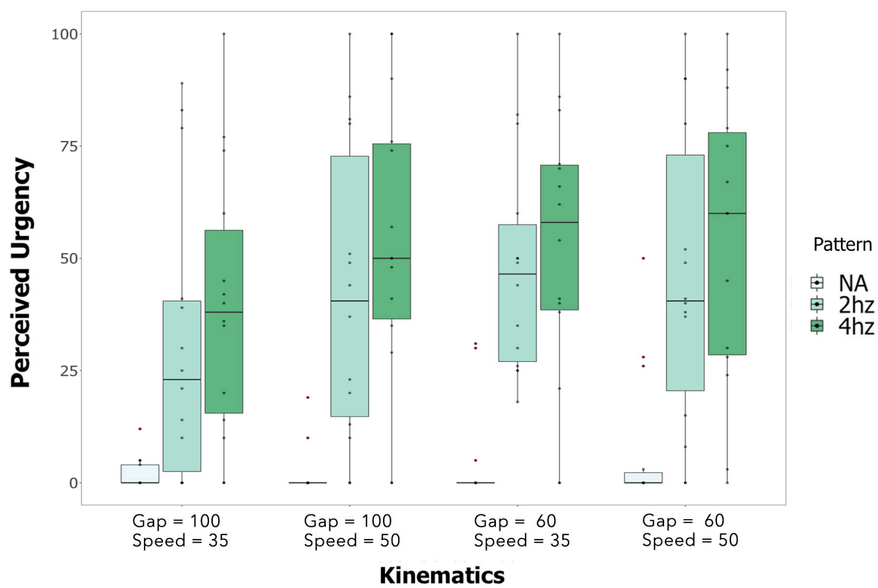

Figure 18: Boxplot of perceived urgency of warning designs (from R studio).

Politeness of warning.

The assumption of sphericity was met at $p > .05$. There was a significant warning main effect, $F(2, 22) = 12.42$, $p < .001$, $f = 1.14$ (see Figure 19). Still, the 4 Hz ($M = 44.458$, $SD = 6.125$) warning condition was not significantly different from the 2 Hz ($M = 45.60$, $SD = 5.50$) warning condition by the Post-Hoc test with Bonferroni corrections. Similarly, both warnings were perceived as more polite than when there was no warning pattern (*M* = 21.29, *SD* = 6.26).

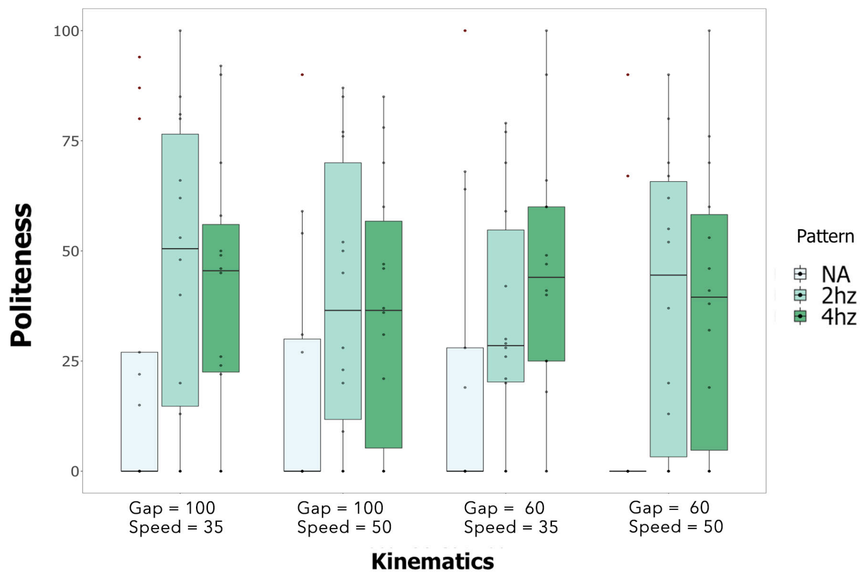

Figure 19: Boxplot of perceived politeness of warning designs (from R studio).

TTA ratio.

No significant effect has been observed for near lane TTA ratio, *p*s > .05.

For far lane TTA ratio, there was a significant three-way interaction effect, $F(2, 22) = 5.07$, $p = .02$, as shown in Figure 20. Following the significant interaction, post-hoc tests with Bonferroni corrections showed that with a gap of 100m, a significant speed × warning interaction effect has been found, $F(2, 24) = 6.84$, $p < .01$. The difference of far lane TTA ratios between 35 km/h and 50 km/h were larger in 2 Hz warning pattern condition compared to no warning and 4 Hz conditions. Further, the far lane TTA ratio was notably smaller in 50 km/h than in 35 km/h when the warning pattern was 4 Hz. No other significant effect was observed.

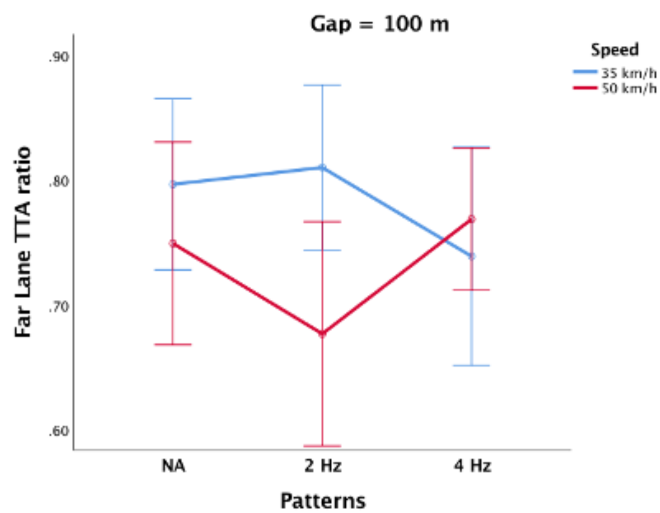

Figure 20: Line chart of Speed × Gap Distance interaction in Far Lane TTA ratio with Gap distance of 100 m (from SPSS).

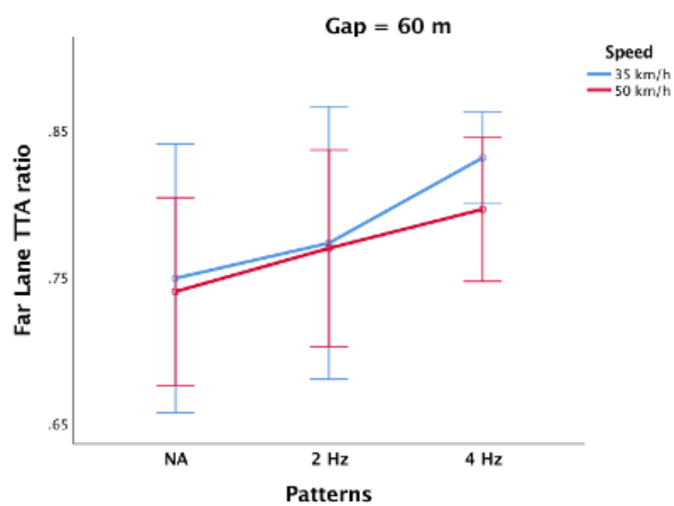

Figure 21: Line chart of Speed × Gap Distance interaction in Far Lane TTA ratio with Gap distance of 60 m (from SPSS).

For far lane TTA ratio, there was a significant three-way interaction effect, $F(2, 22) = 5.07$, $p = .02$, $f= .44$, as shown in Figure 21. Following the significant interaction, post-hoc tests with Bonferroni corrections showed that with a gap of 100m, a significant speed × warning interaction effect has been found, $F(2, 24) = 6.84$, $p < .01$, $f = .75$. The difference of far lane TTA ratios between 35 km/h and 50 km/h were larger in 2 Hz warning pattern condition compared to no warning and 4 Hz conditions. Further, the far lane TTA ratio was notably smaller in 50 km/h than in 35 km/h when the warning pattern was 4 Hz. No other significant effect was observed.

SMT ratio.

For near lane SMT ratio, we observed a significant speed main effect, *F*(1, 11) = 218.95, p < .001, *f* = 1.5. No other effects were significant, *p*s > .05. The near lane SMT ratio was larger in 35 km/h condition (*M* = .62, *SD* = .02) than in 50 km/h condition (*M* = .54, *SD* = .02).

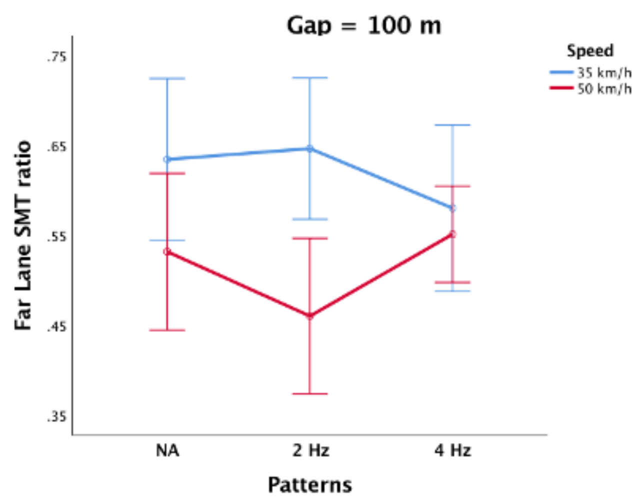

Figure 22: Line chart of Patterns × Speed interaction × Gap in Far Lane SMT ratio with Gap = 100m (from SPSS)

For far lane SMT ratio, a significant three-way interaction effect, *F*(2, 22) = 4.92, *p* = .02, *f* = .67, has been observed as shown in Figure 22. Following the significant interaction, post-hoc tests with Bonferroni corrections showed that with a gap distance of 100 m, a significant speed × warning interaction effect has been found, *F*(2, 24) = 6.59, *p* = .01, *f* = .74.

Similarly, the difference in far lane SMT ratio between 35 km/h and 50 km/h speeds were larger in 2 Hz warning pattern condition compared to No pattern and 4 Hz conditions with a gap distance of 100m. Other simple effects were not significant.

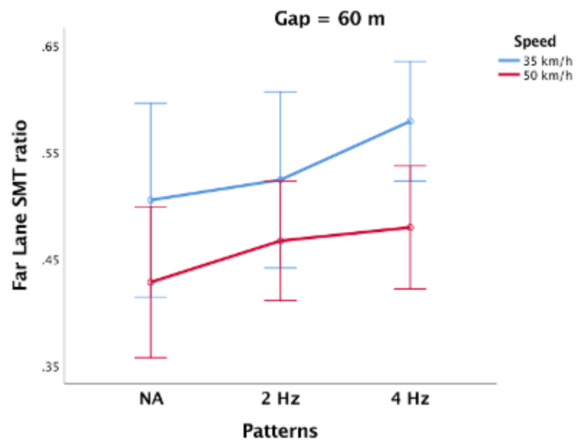

Figure 23: Line chart of Patterns × Speed × Gap in Far Lane SMT ratio with Gap = 60m (from SPSS)

There were also a significant gap main effect, $F(1, 11) = 6.83$, $p = .02$, $f = .79$, and a significant speed main effect, $F(1,11) = 24.37$, $p < .001$, $f = 1.49$, for far lane SMT ratio, indicating that SMT ratio was larger in 35 km/h condition ($M = .58$, $SD = .03$) than in 50 km/h ($M = .49$, $SD = .02$) condition on far lane crossing, and was larger with gap of 100 m ($M = .57$, $SD = .03$) than with 60 m ($M = .50$, $SD = .03$).

Segmented crossing time.

The non-parametric statistical method, Kendall's W test and Friedman test, were adopted for statistical analysis of crossing time within different trials. There were merely significant differences in near lane cross time and mid lane waiting time as shown in Table 5.

Table 5: Friedman Test and Kendall's W Test Results of segmented crossing time.

	Difference in Mean (0.1s)	Friedman Test			Kendall's Coefficient of Concordance
		Chi-Square	*df*	Asymp.Sig.	Kendall's *W*
Near Lane cross time	12.333	19.595	11	0.051	0.148
Mid Lane wait time	38.750	19.959	11	0.046	0.151
Far Lane cross time	16.417	29.914	11	0.002	0.227

For far lane crossing time, according to post-hoc pairwise comparison, the significant differences were detected between 50 km/h × 60 m × 4 Hz vs. 35k m/h × 100 m × No pattern (p=0.03), 50 km/h × 60 m × 4 Hz vs. 35 km/h × 100 m × 4 Hz (p<0.01) and 50 km/h × 60 m × 4 Hz vs. 35 km/h ×100 m × NA (p<0.01). No significant main effect of flashing rates of waning patterns has been validated in any of those pairs.

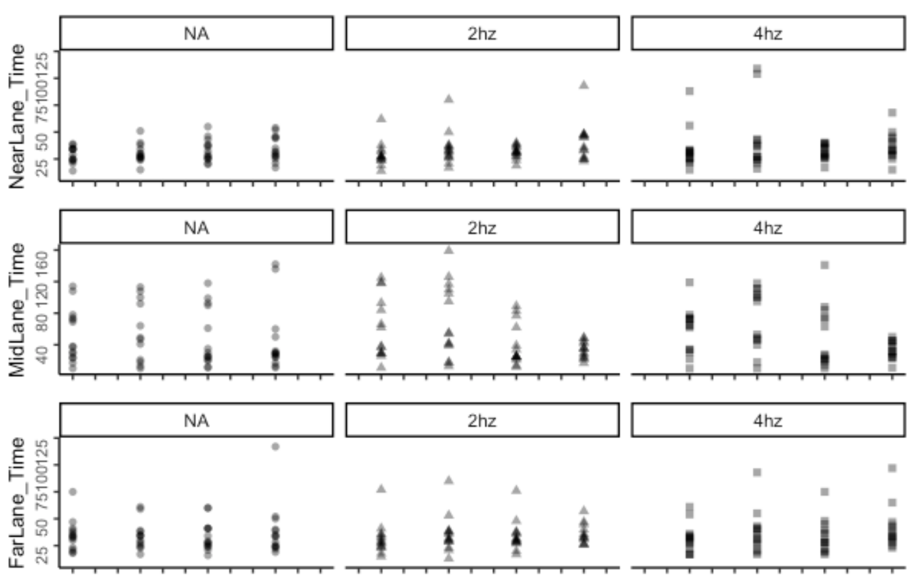

Figure 24: Plots of each lane crossing time.

Lateral displacement

Still, an ANOVA with repeated measures suggested none of warning pattern flashing frequencies, vehicle speed or gap distances effect were statistically significant on lateral displacements at .05 significance level, as illustrated in

Table 6,

Table ***7***,

Table ***8*** and

Table ***9***.

Table 6: Repeated ANOVA measures of near lane waiting lateral displacements.

Source of Variations	df	F	*p* Value
Lateral Displmt. of near lane waiting			
Patterns	2	.896	.422
Gap	1	1.871	.199
Speed	1	.279	.608
Patterns × Gap	2	.566	.576
Patterns × Speed	2	.270	.766
Gap × Speed	2	.001	.975
Patterns × Gap × Speed	2	1.058	.364
Error	22		

Table 7: Repeated ANOVA measures of near lane crossing lateral displacements.

Source of Variations	df	F	*p* Value
Lateral Displmt. of near lane crossing			
Patterns	2	.124	.884
Gap	1	.687	.425
Speed	1	1.140	.308
Patterns × Gap	2	.453	.642
Patterns × Speed	2	.280	.759
Gap × Speed	2	.388	.546
Patterns × Gap × Speed	2	.645	.534
Error	22		

Table 8 : Repeated ANOVA measures of far lane waiting lateral displacements.

Source of Variations	df	F	*p* Value
Lateral Displmt. of far lane waiting			
Patterns	2	1.136	.339
Gap	1	3.946	.072
Speed	1	2.582	.136
Patterns × Gap	2	1.564	.232
Patterns × Speed	2	1.119	.334
Gap × Speed	2	.212	.654
Patterns × Gap × Speed	2	.331	.722
Error	22		

Table 9: Repeated ANOVA measures of near lane crossing lateral displacements.

Source of Variations	df	F	*p* Value
Lateral Displmt. of far lane crossing			
Patterns	2	.233	.794
Gap	1	.757	.403
Speed	1	1.083	.320
Patterns × Gap	2	1.071	.360
Patterns × Speed	2	2.648	.093
Gap × Speed	2	.517	.483
Patterns × Gap × Speed	1.303	.384	.616
Error	14.338		

Figure 25 shows all subjects' trajectories of each trial. Consistent with the quantitative analysis, there is an absence of noticeable trends overall. In the figure, the light-yellow area refers to the waiting mid lane, and the gradient of color indicates the density of time elapsed. The lane between -5 in y axis and the light-yellow belt area represents the near lane, where fleet was forward to the right side and the lane between the waiting area and 5 in y axis represents the far lane, where the fleet drove to the other direction.

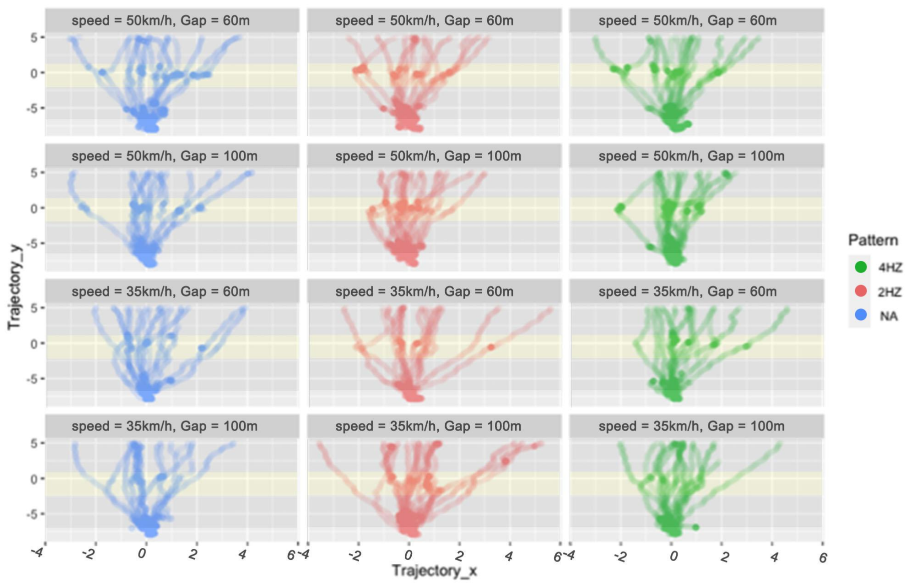

Figure 25: Trajectories in all trials.

Step-back Rate.

Table **10** presents the step back times upon each scenario's context. The total crossing tries includes all participants trying times in each situation.

Table 10: Step-back Rates.

Speed (km/h)	Gap (m)	Pattern Frequency	Step back times	Total crossing tries
50	60	n/a	0	14
50	60	2 Hz	0	14
50	60	4 Hz	3	17
35	60	n/a	0	13
35	60	2 Hz	0	14
35	60	4 Hz	2	16
50	100	n/a	0	14
50	100	2 Hz	0	14
50	100	4 Hz	0	14
35	100	n/a	0	13
35	100	2 Hz	0	14
35	100	4 Hz	0	14

Most participants were not affected by traffic kinematics to step back as long as they decided to cross. The only five step-back cases happened in the fleet gap distance of 60 meters, and a 4 Hz flashing pattern displayed. It has validated the potential efficacy of making pedestrians more vigilant with a high frequency of flashing patterns under denser traffics while the effectiveness was limited.

Auditory Task.

The chosen podcast episodes used as auditory tasks were from Podcast "The Daily" (https://www.nytimes.com/column/the-daily) by the American newspaper The New York Times. The selected clips were about recent (from December 2019 to January 2020) topical news and cultural stories.

On average, the auditory task questions had an overall accuracy of just over 40 percent (43.4%), as shown in Table 11.

Table 11: Auditory Task Accuracies.

Scenario Auditory Task	Accuracy
Scenario 1 Task: The Escape of Carlos Ghosn	42.86%
Scenario 2 Task: Why Australia Is Burning	64.29%
Scenario 3 Task: Why So Many Hospitals Are Suing Their Patients	14.29%
Scenario 4 Task: America's Education Problem	57.14%
Scenario 5 Task: Why Iran Is in Mourning	35.71%
Scenario 6 Task: What Boeing Knew	35.71%
Scenario 7 Task: Firebombs in Hong Kong	64.29%
Scenario 8 Task: An Interview with Andrew Yang	57.14%
Scenario 9 Task: When #MeToo Went on Trial	14.29%
Scenario 10 Task: Walmart Enters the Gun Control Debate	0.00%
Scenario 11 Task: Why China Went to War with the N.B.A.	42.86%
Scenario 12 Task: What Should Happen to the Navy SEAL Chief?	57.14%

I have performed Pearson Chi-square test of independence to examine the relationship between auditory task accuracies and independent variables. The relation between fleet speeds and accuracies was significant, $\chi2$ (4) =35.32, $p < .001$. There was also a meaningful relationship between flashing rates and accuracies, $\chi2$ (4) =52.72, $p < .001$. Thus, it has demonstrated that the auditory task performance was affected by the mental capacity demands change on crossing at various fleet speeds and warnings.

4.2.7 Discussion

Overall, the results suggested that the warning patterns had limited influence on participants' crossing behavior under various controlled kinematic motion combinations.

From an attitudinal perspective, participants generally reported higher hazardousness with higher fleet speed. In addition, the fleet motion kinematics (i.e., fleet speed and gap distance) were the dominant factors in pedestrians' crossing decisions and behavior. In general, TTA and SMT ratios were smaller with more intense traffic (i.e., faster speed and shorter gap distance). H1 was supported.

However, it is worth noting that with a fleet speed of 35 km/h, the perceived situational hazardousness of the 4 Hz pattern was lower than 2 Hz, which implied that 4 Hz pattern did not properly communicate situational hazardousness in scenarios with a fleet speed of 35 km/h.

Although the perceived urgency was significantly different between no warning condition and warning conditions, contrary to our hypothesis H2, the perceived urgency of flashing warning patterns did not increase with higher flashing rates. Further, the overall behavioral measures suggested that the warning patterns had limited influence on participants' crossing behavior under various speed-gap distance combinations, which led us to reject H2.

Both warning conditions were perceived as more polite than no warning condition but there was no difference between 2 Hz and 4 Hz warning conditions. H3 was partially supported.

The only noticeable behavioral compliance to warning patterns was observed in far lane SMT and TTA ratios. This effect could be because before crossing the far lane, participants had to stand in the mid lane where there was traffic flowing behind them. This situation might have resulted in them being more attentive to warning patterns compared to standing at the beginning position, which highlights the importance of contextual factors influencing the intention-behavior gap.

These inconsistent findings could be a result of several disturbing factors on intention realization suggested by Sheeran and Webb (2016). Participants may be distracted by the auditory task or miss the opportunity to behave in such a brief window. Pedestrians' individual characteristics (e.g. aggressiveness) or past crossing experience would guide behaviors through "unconscious mechanisms of the mind" rather than by intentions (Sheeran, 2002). A lack of knowledge (Sheeran, 2002) about traffic norms may elicit behaviors inconsistent with intentions as well. Explaining why such noncompliance exists will require further investigation of the participant's background knowledge, experience, circumstances and reasoning, whether self-conscious or not.

The three-way interaction effects of far lane TTA and SMT ratios revealed that flashing patterns could facilitate accurately estimating vehicle-stopping distances and crossing behaviorally in specific kinematic combinations to some extent. Specifically, in high-density traffic with a 60 m gap, increasing the warning frequency from 2 Hz to 4 Hz created a more considerable gap acceptance and safety margin, suggesting that pedestrians could make less risky crossings. Surprisingly, in low-density traffic with a 100 m gap and 35 km/h fleet speed, increasing the warning frequency from 2 Hz to 4 Hz led to a considerably smaller gap acceptance and safety margin, which was an unexpected result.

Interestingly, in low-density traffic, the warning patterns with higher flashing rates demonstrated poor performance in communicating situational hazardousness and facilitating safer crossing behaviors. One reason might be that pedestrians made crossing decisions based on inherent traffic characteristics (e.g. vehicular speed and gap) beyond imposed warning cues. When the 4 Hz warning improperly represented the moderate traffic, pedestrians did not comply with such warnings. Another possibility is that when the traffic was moderate, pedestrians might have thought that don’t need to pay attention to the warnings. In that case, excessive warnings might have been perceived redundant and disturbing, leading to distrust, complacency and misjudged crossing behaviors. Therefore, I believe that future studies should examine the appropriate matching of these communication cues to the various traffic characteristics.

To sum up, even though warning patterns with various flashing rates could help pedestrians distinguish hazardous situations attitudinally, the vehicles’ kinematic behaviors remained deciding factors in pedestrian behaviors on the crossing. Hence, the discrepancy in these results challenges the efficacy of attitudinal approaches to understanding actual crossing behavior when testing external visual displays.

4.3 Summary

In this chapter, the designed external warning patterns were evaluated in both attitudinal and behavioral perspectives of eliciting participants' vigilance and the practical impact on behaviors. Compliance between perception and practices were examined at the same time.

First, we developed a replicated virtual local street scene with Unity and deployed it to an Android standalone wireless headset. During the experimental sessions, participants were wearing the headset to experience a high-fidelity real-world street in Virtual Reality and complete the crossing task with designed external warning patterns and various traffic kinematic combinations. Subjective rating analysis represented participants' attitudes towards designed warning patterns with specific traffic context, and the study of the dataset collected from the headset tracking system reflected behavioral changes. Results show that the proposed different flashing warning patterns have some effects on perceived urgency, but kinematic behavior of the fleet are still deciding factors on perceived situational hazardousness. By implementing higher frequency flashing warnings in more risky contexts, the researcher can make pedestrians more vigilant at the crossing but may evoke negative emotions. From the results of the behaviors dataset analysis, flashing warning patterns proved limited effects on pedestrians' actions and thereby seemed not expertly facilitated pedestrian decision making. The lack of compliance between attitudes and behaviors lay doubt on the efficacy of attitudinal assessment on external visual warnings studies. Thus, future communication mechanisms between road users at uncontrolled crossing and the intention-behavior discrepancies within such context require further investigations.

Chapter 5
Conclusions

5.1 Summary

This thesis is an exploration of future communication mechanisms between AVs and pedestrians. We have utilized VR technique to evaluate external visual warning efficacy in a future autonomous traffic context. In Chapter 1, I present the necessity and importance of achieving direct, unambiguous, and straightforward communications at risky traffic conditions. Then I explain the interest in studying external warning display's efficacy to improve the communication efficiency with the accelerating development of AVs as well as the potential concerns of employing novel designs of displays for human safety. A review of recent prevalent studies on new external display design concepts with external users re-elected two problems. First, the creative design ideas suffer the limitations as culture differences, interpretation latency, lack of robustness and etc. Therefore, there has not been a consensus of design concepts yet. The second problem was the popular experiment methodologies as crowdsourcing, observation, simulation, or naturalistic driving demonstrated unavoidable shortcomings to acquire robust, effective, and comprehensive results in the study of future scenarios. Therefore, some of the study results are questioned as not transferable to real-world experience.

To solve these problems, we utilized the VR technique to establish a high-fidelity immersive experience of a real-world local street with an autonomous fleet. Within this replicated virtual world, we were capable of implementing designed external warning patterns on AVs and closely observing participants' interactions with future traffic systems without any risks and restrictions. The designed warning patterns recruited simple flashing, which is validated adequately, drawing human attention as an innate from the previous study. Besides external warning displays as visual stimuli, we also designed the kinematic motion of driving entities on the street, as they were proved to be the critical factors for pedestrians crossing decision making. Moreover, as auditory stimuli' crucial impact on road users' communication mechanisms, we conducted an auditory control task for acoustic control and crossing tasks attention assessment. Not only the subjective attitudinal rating, which quantified perceived traffic hazardousness, perceived urgency, and evoked emotions of designed warnings, were collected, but their behavioral compliance was compared and analyzed as well.

The experiment procedure was following the proposed approach in Chapter 3. We employed several Unity Assets Store packages for reassembling the replicated Phillip Street and developed scripts for controlling kinematics motions of the fleet. The participants wore the wireless headset to complete the crossing tasks on a familiar local street at an indoor open free space. Participants were requested to rate their perceived urgency, situational hazardousness accordingly by using the same rating scale, to verify the design. Standard metrics for evaluating pedestrians' intention to cross (e.g., TTA) and post-crossing analysis (e.g., Safety Margin Time) were quantified and analyzed, to present the behavioral performance at each trial.

The study results revealed that kinematic motions of traffic itself still an unignorable dominant decision-making factor at the crossing. Flashing external warning has not achieved the vigilant attitude and risk-averse behaviors at the same time. Consequently, external flashing warnings failed to close the pedestrian intention behavior gap in such traffic contexts.

5.2 Contributions

From the results of the three experiments, the design of flashing warning patterns' efficacy seemed consistent with the results of previous studies of using high frequency flashing stimuli. Thus, the thesis confirmed the potential of using flashing patterns as visual interfaces in traffic-related contexts.

During the experiment, we noticed the problem of auditory dimensions of traffic not evaluated sufficiently, thus proposed a reasonable and feasible auditory-control task to mitigate the overlooked aspect.

The results of the experiment also showed that attitudinal and behavioral effects of stimuli designs may differ. This result supported our concerns in prior studies that only collect participants' intentions, preferences, and opinions. It is important to collect actual behavior, as this may differ from reported intention.

The experiment also explored the feasibility of using wireless VR to provide a controllable and flexible approach study interaction that is closer in reality to the real world. By using wireless VR as the experiment tool bed, we provided a possible improved practice for the evaluation of warnings on.

5.3 Limitations and Future Works

Firstly, in terms of the technical limitations involved in the implementation of VR study, there are limitations to be considered for future studies.

The warning pattern's real-time performance was subject to rendering effects (. e.g., partially compressed and therefore of a lower resolution than the real-world models) for maintaining high-speed movement continuity. Although the effect was barely perceptible during the trials, it may have made slight alterations to the flash patterns that were viewed by participants.

We tried to explore the feasibility of employing podcasts as an auditory control at pedestrian crossing context. The auditory control task did not strictly examine our participants' familiarity with the audio content, so its results may reflect some bias.

Additionally, we have found physical crossing distances changed slightly due to frame latency. As the frame speed is 60 fps within the study, it bore the frame latency of 6.7 milliseconds. This effect has been reported during the recalibration test. As in one trial, the traffic kinematic motion was too intense to run across, the issue perceived running with the body, but frame static, which resulted in the final actual crossing distance was much longer than street width. Thus, potential effects of crossing time measurement accuracy may have been present. And it should be noted that this conflict under fast speed movement between the visual and vestibular systems may introduce motion sickness. Hence, we have adjusted experiments traffic performance cautiously to avoid the possibility of inducing the perception of frame latency.

We have conducted warm-up training trials before formal experiments to mitigate a subtle egocentric distance tolerance effect in the study. Pedestrians were required to freely cross the empty street to get familiar with the virtual environment. Further research should also take a subtle egocentric distance tolerance effect into account. The study suggested that egocentric distances in VR are commonly underperceived by up to 50% of the predetermined intervals. However, a brief period of interaction in which participants walk through the virtual environment while receiving visual feedback can dramatically improve distance judgments (Kelly et al., 2013).

To guarantee high realism, future experimenters should be able to move around an ample space on the same physical scale as the VR world (e.g., A typical two-way city street with a width of 10-12 meters). Currently, conducting such high-fidelity VR research is constrained primarily by the size of the room available to the researchers. However, the 'room-scale VR' is generally set up for open space around 6m × 2m (Terwilliger et al., 2019), which is too restricted for conducting behavioral observation experiments safely through the Virtual Reality approach.

With the high realism details and specificity in the designed experiments, there may be a trade off with generalizability of the results. The Philips street may be a special case due to its unique location and user populations as students. The revealed facts by a university student population may differ from regular pedestrians.

Last but not least, in the context of intelligent driving, intention estimation algorithms have been developed to predict the future actions of pedestrians (Rasouli et al., 2017) and drivers (Molchanov et al., 2015). Technologies, such as wireless communication mechanisms, and various visual intent displays have been introduced continuously. Yet, the majority of these approaches, nonetheless, disregard the theoretical findings of traffic interaction and treat the problem as dealing with a rigid dynamic object rather than a social being (Zhang et al., 2020). Hence, future studies may need to focus more on AV's human-like characteristics (e.g., social awareness) as inspiration for exploring communication mechanisms.

References

Beggiato, M., Witzlack, C., & Krems , J. F. (2017). Gap Acceptance and Time-To-Arrival Estimates as Basis for Informal Communication between Pedestrians and Vehicles. *AutomotiveUI '17: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (pp. 50-57).

King, M. J., Soole, D., & Ghafourian, A. (2009). Illegal Pedestrian Crossing at Signalised Intersections: Incidence and Relative Risk. *Accident Analysis & Prevention*, 41(3):485-90.

Ackermann, C., Beggiato, M., Schubert, S., & Krems, J. F. (2019). An Experimental Study to Investigate Design and Assessment Criteria: What Is Important for Communication Between Pedestrians and Automated Vehicles? *Applied Ergonomics*, 272-282.

Ayoub, J., Zhou, F., & Yang, X. (2019). *From Manual Driving to Automated Driving: A Review of 10 Years of AutoUI. AutomotiveUI.* AutomotiveUI .

Böckle , M.-P., Brenden, A., Klingegård, M., & Habibovic, A. (2017). SAV2P: Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians' Experience. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*, (pp. 136-140).

Bailenson, Yee, & Blascovich. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences*, 102-141.

Baldwin, C. L., Eisert, J. L., Andre, G., Lewis, B., Pratt, S. M., & Gonzalez, C. (2012). Multimodal urgency coding: auditory, visual, and tactile parameters and their impact on perceived urgency. *Work, 41*(Suppl 1), 3586-3591.

Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*.

Barnard, Y., Utesch, F., & Nes, N. v. (2016). The study design of UDRIVE: the naturalistic driving study across Europe for cars, trucks and scooters. *European Transport Research Review .*

Basch , Ethan, C., & Rajan . (2014). (2014). Technology-related distracted walking behaviours in Manhattan's most dangerous intersections., 20(5), 343-346. *Injury Prevention*.

Bernsen, N., Dybkjær, L., & Dybkjær, H. (1998). Wizard of Oz Simulation. In *Designing Interactive Speech Systems.*

Bertulis, T. &. (2014). *Driver Approach Speed and Its Impact on Driver Yielding to Pedestrian Behavior at Unsignalized Crosswalks. .* Transportation Research Record: Journal of the Transportation Research Board.

Bhagavathula, R., Williams, B., Owens, J., & Gibbons, R. (2018). The Reality of Virtual Reality: A Comparison of Pedestrian Behavior in Real and Virtual Environments. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2018.

Blascovich, J., Loomis, J., Beall, A. C., & Swinth, K. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*.

Buhrmester, M., Kwang, T., & Gosling, S. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science, 6(1), 3-5. *Perspectives on Psychological Science*.

Calculation of road traffic noise. (1988). Retrieved from Department of Transport Welsh Office.

Chan , M. (2016). *Distracted Road Use: A Literature Review.* Retrieved from www.edmonton.ca: www.edmonton.ca/transportation/RoadsTraffic/DistractedRoadUseReport.pdf

Chan, A. N. (2009). *Perceptions of implied hazard for visual and auditory alerting signals.* Safety Science.

Chia-Ming Chang, K. T. (2017). *Eyes on a Car: An Interface Design for Communication between an Autonomous Car and a Pedestrian.* In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI ’17. https://doi.org/10.1145/3122986.3122989.

Clamann. (2016). *Method:Field Experiment.*

Clercq, K. d., Dietrich, A., & Velasco, J. P. (2019, March 26). External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. (61), 1353-1370.

Colley, M., Walch, M., & Rukzio, E. (2019). *For a Better (Simulated) World: Considerations for VR in External Communication Research.* In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular.

Colley, M., Walch, M., & Rukzio, E. (2020). *Unveiling the Lack of Scalability in Research on External Communication of Autonomous Vehicles.* Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts.

Conner, M., & Gaston, G. (2007). Temporal stability of behavioural intention as a moderator of intention–health behaviourrelationships. *Psychology & Health*.

Cooke, R., & Sheeran, P. (2013). Properties of intention: Component structure and consequences for behavior, information processing, and resistance. *Journal of Applied Social Psychology*.

Corradi, N., Priftis, K., Jacucci, G., & Gamberini, L. (2013). Oops, I forgot the light on! The cognitive mechanisms supporting the execution of energy saving behaviors. *Journal of Economic Psychology*.

Currano, R., Park, S., & Domingo, L. (2018). ¡Vamos!: Observations of Pedestrian Interactions with Driverless Cars in Mexico. *In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18),210–220.*

de Waard, D., Edlinger, K., & Brookhuis, K. (2011). Effects of listening to music, and of using a handheld and handsfree telephone on cycling behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*.

Deb, S., Strawderman, L., & Carruth, D. (2018). Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transportation Research Part F*.

Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *" Science*, pp. 66–69.

Dey, D., & Terken, J. (2017). *Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication* (pp. 109–113). AutomotiveUI '17: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Dey, D., Habibovic, A., Pfleging, B., Martens, M., & Terken, J. (2020). Color and Animation Preferences for a Light Band eHMI in Interactions Between Automated Vehicles and Pedestrians. *CHI '20: CHI Conference on Human Factors in Computing Systems.* Honolulu, HI, USA.

Dey, Martens, Eggen, & Terken. (2017). The impact of vehicle appearance and vehicle behavior on pedestrian interaction with autonomous vehicles. *International Conference on Automotive User Interfaces and Interactive Vehicular Applications,.*

Dholakia, U., & Pbagozzi, R. (2006). As Time Goes By: How Goal and Implementation Intentions Influence Enactment of Short-Fuse Behaviors1 . *Applied Social Psychology* .

Dingus, Klauer, Neale, & Petersen. (2006). The 100-car naturalistic driving study, phase II: results of the 100-car field experiment. NHTSA. National Highway Traffic Safety Administration (NHTSA).

Dirk Rothenbücher Stanford Center for Design Research, 4. P., Li, J., Sirkin, D., Mok, B., & Ju, W. (17 November 2016). Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles . *IEEE.* New York, NY, usa.

DMV, C. (2020). Retrieved from https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-testing-permit-holders/

Doric, I., Frison Anna, A.-K., Wintersberger, P., Riener , A., & Wittmann, S. (2016). *A Novel Approach for Researching Crossing Behavior and Risk Acceptance: The Pedestrian Simulator.* AutomotiveUI '16 Adjunct: Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Edworthy, J., & Adams, A. (1998). *Warning Design: A Research Prospective.* London: Taylor & Francis.

Einstein, G., McDaniel, M., Williford, C., Pagan, J., & Dismukes, R. (2003). Forgetting of intentions in demanding situations is rapid. . *Journal of experimental psychology.*

Erp, J., Toet, A., & Janssen, J. (2015). Uni-, bi- and tri-modal warning signals: Effects of temporal parameters and sensory modality on perceived urgency. *Safety Science, 72*.

Färber, B. (2016). Communication and Communication Problems Between Autonomous Vehicles and Human Drivers . *Autonomous Driving* , 125-144.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice, 77*, 167-181.

Florentine, E., Ang, M. A., Pend, D. S., & Andersen , H. (2016). Pedestrian Notification Methods in Autonomous Vehicles for Multi-Class Mobility-on-Demand Service. *HAI '16: Proceedings of the Fourth International Conference on Human Agent Interaction*, (pp. 387-392).

Greenberg, H. (1959). An analysis of traffic flow. *Operations research*.

Habibovic, A., Andersson, J., Lagström, T., & Nilsson, M. (2016). Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context? *Advances in Human Aspects of Transportation* (pp. 485-497). Florida, USA: AHFE 2016 International Conference on Human Factors in Transportation.

Hassan., J. M. (n.d.). *How Do Vision and Hearing Impact Pedestrian Time-to-Arrival Judgments?*

Hellier, E., Edworthy, J., & Dennis , I. (1993). Improving Auditory Warning Design: Quantifying and Predicting the Effects of Different Warning Parameters on Perceived Urgency. *Human Factors*.

Henry, C., Mary, P., & Ellen, L. (2012). Emotional Dysfunction as a Marker of Bipolar Disorders. *Front Biosci (Elite Ed)*.

Hock, P., Kraus, J., & Babel, F. (2018). How to Design Valid Simulator Studies for Investigating User Experience in Automated Driving: Review and Hands-On Considerations. *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.*

Holländer, K., Colley, A., Mai, C., Häkkilä, J., & Alt, F. (2019). Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services.*

Huang, S.-C., Bias, R., & Schnyer, D. (2014). How are icons processed by the brain? Neuroimaging measures of four types of visual stimuli used in information systems. *Association for Information Science and Technology*.

Jerald, J. (2015). *The VR Book: Human-Centered Design for Virtual Reality.* Association for Computing Machinery and Morgan & Claypool.

Jouen, F. (1990). *arly visual-vestibular interactions and postural development.* Sensory-motor Organization and Development in Infancy and Early Childhood.

Kelly, J. W., Donaldson, L. S., Sjolund, L. A., & Freiberg, J. B. (2013). More than just perception–action recalibration: Walking through a virtual environment causes rescaling of perceived space. *Atten Percept Psychophys .*

Kelly, J., Lucia, C., & Siegel , Z. (2017). Perceived Space in the HTC Vive. *ACM Transactions on Applied Perception*.

Lagström, T., & Victor, M. (2016). AVIP - Autonomous vehicles' interaction with pedestrians - An investigation of pedestrian-driver communication and development of a vehicle external interface.

Landers, & Behrend. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, 142-164.

Lasse, J., & Flemming, K. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*.

Lewis, B., & Baldwin, C. (2012, 9). Equating Perceived Urgency Across Auditory, Visual, and Tactile Signals. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 1307-1311.

Li, Y., Dikmen, M., Hussein, T., Wang, Y., & Burns, C. (2018). *To Cross or Not to Cross: Urgency-Based External Warning Displays on Autonomous Vehicles to Improve Pedestrian Crossing Safety.* 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Lo, C. (2012). *Driverless train technology and the London Underground: the great debate* . Retrieved from railway-technology: https://www.railway-technology.com/features/featuredriverless-train-technology/

Müller, L. R. (2016). *The social behavior of autonomous vehicles.* . 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16.

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.*

Maruhn, P., Schneider, S., & Bengler, K. (2019). Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. *PLOS ONE*.

Matney, L. (2018). *Google's first WorldSense VR headset, the Lenovo Mirage Solo, ships in Q2 for 'under $400'*. Retrieved from Cruchbase: https://techcrunch.com/2018/01/09/googles-first-worldsense-vr-headset-the-lenovo-mirage-solo-ships-in-q2-for-under-400/

M'bailara, K., Atzeni , T., Contrand, B., & Derguy, C. (2018). Emotional Reactivity: Beware Its Involvement in Traffic Accidents. *Psychiatry Res.*

Merenda , C., Kim, H., & Gabbard , J. (2017). Did You See Me?: Assessing Perceptual vs. Real Driving Gains Across Multi-Modal Pedestrian Alert Systems. *AutomotiveUI '17:*

Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, (pp. 40-49).

MIT moral machine. (2018). Retrieved from https://www.media.mit.edu/projects/moral-machine/overview/

Molchanov, P., Gupta , S., Kim, K., & Pulli, K. (2015). Multi-sensor system for driver's hand-gesture recognition. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG).*

Nasar, J., & Troyer, D. (2013). Pedestrian injuries due to mobile phone use in public places. *Accident Analysis and Prevention*, 91-95.

Nascimento, A., Queir, A., & Cugnasca, P. (2019). The Role of Virtual Reality in Autonomous Vehicles' Safety. *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR).*

Neider, M., McCarley, J., & Crowell, J. (2010). Pedestrians, vehicles, and cell phones. *Accident Analysis and Prevention, 42*(2), 589-594.

NHTSA. (2019, March). *Traffic Safety Facts.* Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812681

Onatrio Highway Act . (1990). Retrieved from Highway Traffic Act, R.S.O. 1990, c. H.8: https://www.ontario.ca/laws/statute/90h08

Overview: Mercedes-Benz F 015 luxury in motion . (2017, 06 30). Retrieved from Global Media Daimler: https://media.daimler.com/marsMediaSite/en/instance/ko/Overview-Mercedes-Benz-F-015-Luxury-in-Motion.xhtml?oid=9904624

Owen, R. S. (1991). Clarifying the Simple Assumption of the Secondary Task Technique . *NA - Advances in Consumer Research*.

Politis , I., Brewster, S., & Pollick , F. (2014). Speech Tactons Improve Speech Warnings for Drivers. *AutomotiveUI '14: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.*

Politis, I., Brewster, S., & Pollick, F. (2013). Evaluating Multimodal Driver Displays of Varying Urgency. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI ,13).* Eindhoven, The Netherlands.

Rasouli , A., & Tsotsos, J. (2019). *Autonomous Vehicles that Interact with Pedestrians: A Survey of Theory and Practice.* IEEE Transactions on Intelligent Transportation Systems.

Rasouli , A., Kotseruba, I., & Tsotsos, J. (2017). Agreeing to cross: How drivers and pedestrians communicate. *2017 IEEE Intelligent Vehicles Symposium.*

Rasouli, A., Kotseruba, I., & Tsotsos, J. (2017). *Are They Going to Cross? A Benchmark Dataset and Baseline for Pedestrian Crosswalk Behavior.* The IEEE International Conference on Computer Vision (ICCV).

Record, W. R. (2018, December 11). *Phillip Street closed after pedestrian struck*. Retrieved from TheRecord.com: https://www.therecord.com/news-story/9078851-phillip-street-closed-after-pedestrian-struck/.

Renner, R., Velichkovsky, B., & Helmert, J. (2013). The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*.

Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Rizzo, A., & Kim, G. (2003). A SWOT Analysis of the Field of Virtual Reality Rehabilitation and Therapy. *Teleoperators and Virtual Environments*.

Road data. (2020). Retrieved from City of Waterloo Open Data: http://data.waterloo.ca/datasets/roads/data?

Roberson, M. J., Barto, C., & Mehta, R. (2017). Driving in the Matrix: Can Virtual Worlds Replace Human-Generated Annotations for Real World Tasks? *Proceedings of International Conference on Robotics and Automation (ICRA).*

Robertson, A. (2019). *Google is discontinuing the Daydream View VR headset, and the Pixel 4 won't support Daydream*. Retrieved from The Verge: https://www.theverge.com/2019/10/15/20915609/google-pixel-4-no-daydream-support-view-vr-headset-discontinued

Roulin, N. (2015). Don't Throw the Baby Out With the Bathwater: Comparing Data Quality of Crowdsourcing, Online Panels, and Student Samples . *Cambridge Core*.

Rules of the Road. (2017). Retrieved from novascotia.ca: http://www.gov.ns.ca/snsmr/rmv/handbook/DH-Chapter2.pdf

Savage, A. (2018). *Lenovo Mirage Solo VR Headset Review!* Retrieved from Youtube: https://www.youtube.com/watch?v=JPpoQoSo4pM

Schneider , S., Maruhn, P., & Bengler, K. (2018). Locomotion, Non-Isometric Mapping and Distance Perception in Virtual Reality. (pp. 22-26). New York, New York: Proceedings of the 2018 10th International Conference on Computer and Automation Engineering—ICCAE 2018.

Service, T. (2017). *Lanewidths Guideline.* Retrieved from City of Toronto.

Sheeran, P. (2002). Intention—Behavior Relations: A Conceptual and Empirical Review. *European Review of Social Psychology*.

Sheeran, P., & Webb, T. (2016). The Intention–Behavior Gap. *Social and Personality Psychology Compass*.

Shim, L., Liu, P., & Politis, I. (2015). Evaluating multimodal driver displays of varying urgency for drivers on the autistic spectrum.

Sieß, A., Hübel, K., & Hepperle, D. (2015). Hybrid City Lighting - Improving Pedestrians' Safety through Proactive Street Lighting. *2015 International Conference on Cyberworlds (CW).* IEEE.

Sun, R., Zhuang, X., Wu, C., Zhao, G., & Zhang, K. (2015). The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*.

Teller, D. (1979). *The forced choice preferential looking procedure: a psychophysical technique for use with human infants.* Developmental Medicine and Child Neurology 33.

Teller, D. M. (1986). *Visual development.* Vision Research 26, 1483-1506.

Terwilliger, J., Glazer, M., & Schmidt, H. (2019). Dynamics of Pedestrian Crossing Decisions based on Vehicle Trajectories in Large-Scale Simulated and Real-World Data. *DRIVING ASSESSMENT CONFERENCE.*

Thompson, Rivara, L., Ayyagari, F., & Ebel. (2013). Impact of social and technological distraction on pedestrian crossing behaviour: An observational study. *Injury Prevention*, 232-237.

TRL. (2016). *Heathrow shuttles "take off" from Terminal 5* . Retrieved from Heathrow Media Centre: https://mediacentre.heathrow.com/pressrelease/details/81/Corporate-operational-24/5605

Van Cott, H. K. (1972). *Human Engineering Guide to Equipment Design.* Washington DC.: Government Printing Office.

Vaughan, J. (2018). Making better use of the crowd: How crowdsourcing can advance machine learning research. *Machine Learning Research*.

VL Knoop, C. D. (2018). *The Effect of Crosswalks on Traffic Flow.* EJTIR.

WHO. (2016). *Road Traffic Injuries.* Retrieved from World Health Organization: www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Wickens, C., Lee, J., Liu, Y., & Gordon-Becker, S. (1998). *Introduction to Human Factors Engineering.*

Zhang, H., Liu, Y., Wang,, C., & Fu, R. (2020). Research on a Pedestrian Crossing Intention Recognition Model Based on Natural Observation Data. *Conference on Intelligent Transportation Systems (ITSC).*

Zimmermann, R., & Wettach, R. (2017). First Step into Visceral Interaction with Autonomous Vehicles. *AutomotiveUI '17: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (pp. 58-64).

Appendix A Experimental Documents

A.1 Questionnaire of Experiment

Participant Number:____________

Q1. As a pedestrian, how often did you cross Phillip Street in the last 7 days?

◯ A few times a day

◯ Daily

◯ 4 - 5 times

◯ 2 - 3 times

◯ Only one time

◯ Never

Q2. How long have you had your current driving license?

◯ I do not currently have a driver's license.

◯ Less than 5 years

◯ 5 to 10 years

◯ More than 10 years

Q3. Have you ever had any experience with autonomous vehicles?

◯ Yes, as a driver.

◯ Yes, as a passenger.

◯ Yes, as a pedestrian.

◯ Never

◯ Others, please specify ____________________

Q4. Have you ever had any experience with Virtual Reality?

◯ Yes

◯ No

◯ Other, please specify ____________________

Q5. In your opinion, how would you compare the design of future autonomous cars and the design of conventional cars? (Choose one)

◯ The more similar the better

◯ There should be a clear distinction between the design of future autonomous cars and the design of conventional cars.

◯ The ability to interact with drivers/passengers in the autonomous car

◯ It doesn't matter.

Q6. If an autonomous car approaches you when you cross a street, what kind of information that is presented by the autonomous car could increase your trust in the autonomous car? (Choose all that apply)

☐ Showing explicit information about whether the car is an autonomous car or a conventional car

☐ Showing explicit information about the autonomous car’s next action

☐ Showing explicit information about the guidance to pedestrians (such as “go ahead” or “please stop”)

☐ Showing no information just like a normal family car

☐ Other (Please specify): ____________________

Q7. How often have you listened to podcasts/radio/Audio books/music while walking (jogging) on the street alone?

○ Never (0%)

○ Occasionally (0% < f =< 30% ）

○ Frequently (30% < f =<60%)

○ Often (60% < f = < 90%)

○ Always (90% < f)

A.2 Auditory Task Contents Transcriptions and Questions

All audio contents used for auditory tasks in our study were recruited from Podcast: *The Daily*. It is a daily news podcast and radio show by the American newspaper The New York Times, its episodes are based on the Times' reporting of the day with interviews of journalists from the New York Times. During each crossing trial, the participants listened to clips from The daily. Each clip lasted for no longer than 40 seconds. The correct answers for the following questions are marked as bold and underlined in the text.

Scenario 1 Task:

The Escape of Carlos Ghosn

Tuesday, January 14th, 2020

Today: The trial was poised to be the most closely-watched in Japanese history — a case involving claims of corporate greed, wounded national pride, and a rigged legal system, until the defendant pulled off an unimaginable escape. Ben Dooley with the latest in the saga of Carlos Ghosn. Ben, set the scene for us. It's just before New Year's in Tokyo. What are you doing? So I'm fast asleep. It's about 7:30 in the morning. I'm expecting to have a very easy holiday. The government offices are all closed for a week. Nothing's happening in Japan. And all of a sudden my phone starts ringing…

Question: Where was the journalist Ben before New Year's Eve?

A. New York **<u>B. Tokyo</u>** C. Lebanon

Scenario 2 Task:

Why Australia Is Burning

Monday, January 13th, 2020

Today: **<u>The prime minister of Australia</u>** is calling for a high-level investigation into how the country's government has responded to the devastating wildfires there. My colleague Livia Albeck-Ripka on the events that led up to this moment. It's Monday, January 13. Livia, you're on the ground in Australia, covering these wildfires that have been raging throughout the continent. What have you been seeing these past few weeks?

Livia Albeck-Ripka: So the scenes that are unfolding here in Australia are totally unprecedented. We do have bushfires here in Australia, but what we're seeing now has not happened before...

Question: Who was calling for a high-level investigation into how the country's government has responded to the devastating wildfires mentioned in the News?

A. Residence in Sydney
B. **<u>Prime minister of Australia</u>**
C. Australian House of Representatives

Scenario 3 Task:

Why So Many Hospitals Are Suing Their Patients

Monday, December 2nd, 2019

Sarah Kliff: So Amanda is 41 years old. She lives in this tiny little town in rural Virginia called Norton.

Amanda Sturgill: It's very small. I don't even know why it's called a city, because it's not very big.

Sarah Kliff: She has **four** children. She's a single mom.

Michael Barbaro: Sarah Kliff writes about health care for The Times.

Sarah Kliff: She works full time at an audio equipment company, where she processes orders.

Amanda Sturgill: So it's just like a little freak thing that happened.

Sarah Kliff: And a few years ago...

Amanda Sturgill: I was working...

Sarah Kliff: Her daughter Michaela was giving one of the other kids a bath.

Michael Barbaro: Mm-hm.

Amanda Sturgill: And she bent down to pick her up, to get her out the bathtub.

Sarah Kliff: She bent down to pick up her sibling and just got this terrible pain in her back.

Amanda Sturgill: Kind of go down her legs. And she was just in this horrible, excruciating pain. Took her to the emergency room —

Sarah Kliff: Amanda, obviously, is worried. She doesn't understand what's going on, so she takes her daughter to the emergency room at the local hospital system, a company named Ballad. They give her a pain shot, go back home, but this pain doesn't go away.

Question: How many children does Amanda have?

A. Two B. Three **C. Four**

Scenario 4 Task:

America's Education Problem

Thursday, December 5th, 2019

Michael Barbaro: Dana, when did the United States start to feel a sense of anxiety around the education levels of our children in relation to the rest of the world?

Dana Goldstein: Well, I think back to 1957. And that was the year that the **Russians beat us to space** with their Sputnik satellite. And this triggered a sort of national conversation and anxiety in the United States among our political leaders and the public. We were the country that beat back tyranny and saved the world in World War I and World War II. What did it mean that this other nation could outperform us in this new frontier of space?

Question: why did the United States start to feel a sense of anxiety around the education levels of our children in relation to the rest of the world in 1957?

A. **Russians beat them to space**

B. Political leaders’ anxiety

C. The rise of Japan

Scenario 5 Task:

Why Iran Is in Mourning

Tuesday, January 7th, 2020

Monday morning was the start of **the official state funeral** for General Qassim Suleimani. By 8:00 a.m., there were millions of people out in downtown Tehran. He was being celebrated as a national hero, but also as a religious martyr and a saint. There were families. There were men, women, children. They had the symbolic Shia ritual symbols out — feathers, swords, drums, music, eulogies, songs. And the crowd also had a very anti-American and defiant mood. People were sad, but they were also very angry, and we heard a lot of “revenge, revenge,” and “no more negotiations with the U.S., it’s time for battle,” chanted by the crowd.

Question: Why were there millions of people out in downtown Tehran mentioned in the News?

A. **<u>A Funeral</u>** B. Mass Protest C. Religious ritual

Scenario 6 Task:

What Boeing Knew

Thursday, October 31st, 2019

So Natalie, for the past six months, we have been talking to you about what went wrong with Boeing's 737 Max, this jet that crashed **<u>twice</u>**, basically back to back, over the past year and both times killed everybody on board. Remind us what you have uncovered about those crashes up until now. What went wrong? So the first 737 Max crashes in October. It's not really clear what's going on. It seems like it's a new automated system on the plane.

Question: How many times has Boeing's 737 Max' crash(s) been mentioned in the talk ?

A. None B. One **<u>C. Two</u>**

Scenario 7 Task:

Firebombs in Hong Kong.

Wednesday, October 2nd, 2019

Today: As China celebrated 70 years of Communist Party rule...scenes of pageantry, pride and unity in Beijing were met with firebombs, rubber bullets and mass protests in Hong Kong.My colleague Natalie Kitroeff speaks with China correspondent Javier Hernández about a day of historic contrasts. It's Wednesday, October 2.

Natalie Kitroeff: Javier, take us back to a few weeks ago when you get this knock at your door.

Question: What was happening in Beijing as there were mass protests in Hong Kong?

A. **National day celebration** B. Mass protest C. Nothing

Scenario 8 Task:

An Interview With Andrew Yang

Thursday, September 12th, 2019

So I met Andrew Yang a few years ago when he was running this organization called **Venture for America**, which he had had a couple of careers. He was a corporate lawyer, did some startups. He actually sold one of his startups and made a decent chunk of money from that. And then he was trying to do this thing where he would essentially turn recent college graduates into entrepreneurs — set them up with some money and supporting them as they went off and started companies.

Question: What was the name of Andrew Yang' organization a few years ago when Kevin met him?

A. Crooked Media

B. Dean and Deluca

C. <u>Venture for America</u>

Scenario 9 Task:

When #MeToo Went on Trial

Friday, October 4th, 2019

It's been months and months since the **<u>Weinstein</u>** story broke, and #MeToo too has become this global phenomenon. But there's also a lot of controversy and a kind of sense of mounting unfairness. A real backlash is accumulating. And we're seeing that tension play out in the news every single day. Ronan Farrow reports a very powerful story about Les Moonves. But Moonves refuses to step down.

Question: Since whose story break, #MeToo has become this global phenomenon?

A. Bill Cosby

B. <u>Weinstein</u>

C. Kevin Spacey

Scenario 10 Task:

Walmart Enters the Gun Control Debate

Thursday, September 5th, 2019

So I was sitting at my desk. It's midday Tuesday. I was actually about to get up to go grab lunch. And I looked at my screen, and an email popped up. And it said, from Doug McMillon. Now, Doug McMillon is the **C.E.O. of Walmart,** the largest retailer in the country. And the subject line was "Walmart's next steps." And he started the email — Dear Andrew, I wanted to follow back up with you on the subject of firearms and ammunition. And my eyes widened.

Question: Who is Doug McMillon?

- **A. C.E.O. of Walmart**
- B. Financial columnist
- C. President of NRA

Scenario 11 Task:

Why China Went to War With the N.B.A.

Friday, October 11th, 2019

Well, Daryl Morey is the general manager of the **Houston Rockets**, so he runs the team. And he's probably one of the more unlikely general managers you're going to find in the N.B.A. Most of those guys are former players, or they have been in the past. He was a data nerd. He worked for a data firm, and the whole league was like, who is this guy? But he was kind of of an era.

Question: Which team did Daryl Morey work as the general manager?

- **A. Houston Rockets**
- B. Los Angeles Clippers

C. North Carolina Tar Heels

Scenario 12 Task:

What Should Happen to the Navy SEAL Chief?

Monday, November 25th, 2019

Today, the trial gets underway for the Navy SEAL accused of murdering an Islamic State prisoner. Special Operations Chief Edward Gallagher is charged with **premeditated murder**. A former Navy SEAL pleaded not guilty today to war crimes, including This trial has a plot that could be turned into a movie. A stunning twist at a military tribunal in San Diego just sent shockwaves through the court.

Question: What was Special Operations Chief Edward Gallagher initially charged with?

D. **Premeditated murder**

E. Posing in a photo with a dead ISIS fighter

F. Abusing prisoners