TU Delft

Delft University of Technology

Bridging Worlds: Augmented Reality for Pedestrian-Automated Vehicle Interactions

Tabone, W.

DOI
10.4233/uuid:2879077b-f96f-4380-800a-d796611ba26a

Publication date
2024

Document Version
Final published version

Citation (APA)
Tabone, W. (2024). *Bridging Worlds: Augmented Reality for Pedestrian-Automated Vehicle Interactions*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:2879077b-f96f-4380-800a-d796611ba26a

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

BRIDGING WORLDS

Augmented Reality for Pedestrian-Automated Vehicle Interactions



WILBERT TABONE

BRIDGING WORLDS: AUGMENTED REALITY FOR PEDESTRIAN-AUTOMATED VEHICLE INTERACTIONS

BRIDGING WORLDS: AUGMENTED REALITY FOR PEDESTRIAN-AUTOMATED VEHICLE INTERACTIONS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Monday, 4 March 2024 at 15:00 o'clock

by

Wilbert TABONE

Master of Science in Artificial Intelligence,
University of Malta, Malta,
born in Pietà, Malta.

This dissertation has been approved by the promotors.

promotors: Prof.dr.ir. J.C.F. de Winter, Prof.dr.ir. R. Happee

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof.dr.ir. Joost C.F. de Winter	Delft University of Technology, promotor
Prof.dr.ir. Riender Happee	Delft University of Technology, promotor
Independent members:	
Prof.dr. J.F. Krems	Chemnitz University of Technology, Germany
Prof.dr.ir. A. Bozzon	Delft University of Technology
Prof.dr. D. de Waard	University of Groningen
Prof.dr. M.A. Neerincx	Delft University of Technology
Prof.dr.ir. J. Hellendoorn	Delft University of Technology, reserve member
Other members:	
Prof.dr. N. Merat	University of Leeds, The United Kingdom

TU Delft Delft University of Technology

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410

SHAPE-IT

Keywords: automated vehicles, augmented reality, pedestrians, interactions, virtual reality, simulators, eye-tracking, large language models, interface design

Printed by: Gildeprint Drukkerijen

Cover: Human-AI collaboration between DALL.E 3, Adobe Firefly, and the author. Further post-production by Caiseal Beardow

ISBN 978-94-6496-045-7

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

All we have to decide is what to do with the time that is given us.

Gandalf

CONTENTS

Summary xiii

Samenvatting xvii

Taqsira xxi

1 Introduction 1
- 1.1 The Future Urban Environment, and Connected Cities . . . 2
- 1.2 Automated Vehicles . . . 2
- 1.3 Encounters with Automated Vehicles . . . 4
  - 1.3.1 Implicit and Explicit Cues . . . 4
  - 1.3.2 Vehicle-to-Pedestrian (V2P) Communication . . . 4
  - 1.3.3 External Human-Machine Interfaces (eHMIs) . . . 4
  - 1.3.4 Limitations of eHMIs . . . 5
- 1.4 Computer-Mediated Reality . . . 6
  - 1.4.1 The xReality Framework . . . 7
  - 1.4.2 Experiencing Augmented Reality . . . 9
  - 1.4.3 XR in Pedestrian-Vehicle Interaction Studies . . . 10
- 1.5 Research Gaps and Motivation . . . 11
- 1.6 Objectives of this Thesis . . . 12
- 1.7 Thesis Outline . . . 12
- References . . . 17

I Part I: Design 27

2 Vulnerable road users and the coming wave of automated vehicles: Expert perspectives 29
- 2.1 Glossary . . . 30
- 2.2 Introduction . . . 32
- 2.3 Methods . . . 34
- 2.4 Researchers' views . . . 35
  - 2.4.1 Jonas Bärgman . . . 35
  - 2.4.2 Martin Baumann . . . 36
  - 2.4.3 Shuchisnigdha Deb . . . 37
  - 2.4.4 Azra Habibovic . . . 38
  - 2.4.5 Marjan Hagenzieker . . . 40
  - 2.4.6 P. A. Hancock . . . 41
  - 2.4.7 Riender Happee . . . 42
  - 2.4.8 Josef Krems and Claudia Ackermann . . . 43
  - 2.4.9 John D. Lee . . . 44

2.4.10 Marieke Martens . . . 46
2.4.11 Natasha Merat . . . 47
2.4.12 Don Norman and Colleen Emmenegger . . . 48
2.4.13 Thomas B. Sheridan . . . 49
2.4.14 Neville A. Stanton . . . 50
2.5 Discussion . . . 51
2.5.1 Comparison of the Researchers' Views on the Future of AVs . . . 51
2.5.2 Comparison of views on eHMIs and AR for eHMIs . . . 53
2.5.3 Comparison of views on VR and AR simulation . . . 54
2.6 Conclusions . . . 55
2.6.1 Limitations . . . 55
2.6.2 Recommendations for Future Research . . . 56
References . . . 57
Appendix . . . 67

3 Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes 69
3.1 Introduction . . . 70
3.2 Methods . . . 71
3.2.1 Phase 1: Collection of ideas . . . 71
3.2.2 Collection of relevant literature . . . 72
3.2.3 Phase 3: Brainstorming design sessions . . . 72
3.2.4 Phase 4: Implementation . . . 73
3.2.5 Creating the final application and testing on-site . . . 75
3.3 Results and Discussion . . . 75
3.3.1 Concept 1: Augmented zebra crossing . . . 75
3.3.2 Concept 2: Planes on vehicle . . . 76
3.3.3 Concept 3: Conspicuous looming planes . . . 76
3.3.4 Concept 4: Field of safe travel . . . 76
3.3.5 Concept 5: Fixed pedestrian traffic lights . . . 77
3.3.6 Concept 6: Virtual fence . . . 78
3.3.7 Concept 7: Phantom car . . . 79
3.3.8 Concept 8: Nudge HUD . . . 79
3.3.9 Concept 9: Pedestrian traffic lights HUD . . . 80
3.4 General Discussion . . . 81
3.4.1 Limitations and future work . . . 81
3.5 Conclusion . . . 83
References . . . 84

II Part II: Evaluation 91

4 Augmented reality interfaces for pedestrian-vehicle interactions: An online study 93
4.1 Introduction . . . 94
4.2 Method . . . 98
4.2.1 Videos . . . 98

4.2.2 Questionnaire Procedure . . . 100
4.2.3 Questionnaire Design . . . 101
4.2.4 Final questions . . . 103
4.2.5 Analysis . . . 103
4.3 Results . . . 104
4.3.1 Ratings of Videos Depicting AR Interfaces . . . 105
4.3.2 Descriptor Scale (Q21), Acceptance scale (Q22), and Composite Score . . . 106
4.3.3 Assessment of Moderator Variables . . . 106
4.3.4 Textual Responses (Q23) . . . 111
4.3.5 Preferred AR Interfaces and Use of Augmented Reality in Traffic . . . 116
4.4 Dicussion . . . 119
4.4.1 Interface Preference by Respondents . . . 119
4.4.2 Free-Text Comments . . . 120
4.4.3 Limitations and Future Work . . . 121
4.5 Conclusion . . . 122
References . . . 124
Appendix . . . 134

5 Using ChatGPT for Human–Computer Interaction Research: A Primer 141
5.1 Introduction . . . 142
5.2 Study 1: Questionnaire Textbox Data . . . 144
5.2.1 Methods . . . 145
5.2.2 Results . . . 149
5.3 Study 2: Interview Data . . . 153
5.3.1 Methods . . . 153
5.3.2 Results . . . 154
5.4 Study 3: Think Aloud Data . . . 159
5.4.1 Methods . . . 160
5.4.2 Results . . . 162
5.5 Discussion . . . 164
5.6 Conclusion . . . 168
References . . . 169

6 Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment 177
6.1 Introduction . . . 178
6.2 Methods . . . 181
6.2.1 Participants and Recruitment . . . 181
6.2.2 Apparatus . . . 181
6.2.3 Augmented-Reality Interfaces . . . 182
6.2.4 Scenario Design . . . 183
6.2.5 Experimental Design . . . 183
6.2.6 Experimental Procedure . . . 187
6.2.7 Data Logging . . . 189
6.2.8 Processing of Post-Block Interviews . . . 189

6.2.9 Dependent Measures . . . 190
6.2.10 Statistical Analysis . . . 191
6.3 Results . . . 191
6.3.1 Objective 1: Replication of Intuitiveness Ratings: CAVE study vs. Online Study . . . 192
6.3.2 Objective 2: AR interface/Attention-Attractor Position Interaction . 192
6.3.3 Post-Block Interviews . . . 199
6.4 Discussion . . . 203
6.4.1 Objective 1: Replication of Intuitiveness Ratings: CAVE study vs. Online Study . . . 203
6.4.2 Objective 2: AR interface/Attention-Position Interaction . . . 205
6.4.3 Limitations and Future Work . . . 206
6.5 Conclusion . . . 207
References . . . 208
Appendix . . . 214

7 Augmented reality for supporting the interaction between pedestrians and automated vehicles: An experimental outdoor study 217
7.1 Introduction . . . 218
7.2 Methods . . . 221
7.2.1 Participants . . . 221
7.2.2 Materials and Settings . . . 221
7.2.3 Participant Information and Task . . . 222
7.2.4 Automated Vehicle Behaviour . . . 223
7.2.5 Augmented Reality Designs . . . 224
7.2.6 Experiment Design . . . 227
7.2.7 Questionnaires and Rating Scales . . . 227
7.2.8 Data Recording and Analysis . . . 228
7.3 Results . . . 229
7.4 Discussion . . . 234
7.5 Conclusion . . . 236
References . . . 237

III Part III: Diminished Reality 245

8 Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios 247
8.1 Introduction . . . 248
8.2 Method . . . 251
8.2.1 Participants . . . 251
8.2.2 Materials . . . 251
8.2.3 Participant task . . . 251
8.2.4 Virtual environment . . . 253
8.2.5 Augmented Reality Designs . . . 256
8.2.6 Experimental Design . . . 258

8.2.7 Questionnaires . . . 258
8.2.8 Data analysis . . . 260
8.3 Results . . . 261
8.3.1 Head Movement Analysis . . . 263
8.4 Discussion . . . 265
8.5 Conclusion . . . 268
References . . . 269
Supplementary Material . . . 277

IV Part IV: Conclusion **279**

9 Conclusions and Recommendations **281**
9.1 Summary of Key Findings . . . 282
9.2 Reflections . . . 287
9.2.1 What is the Potential of Augmented Reality in Facilitating Pedestrian-Vehicle Interactions? . . . 287
9.2.2 How Should AR be Designed, Integrated, and Anchored onto the Real Environment ? . . . 288
9.2.3 Reflections on Evaluation . . . 291
9.2.4 Large Language Models as a Tool for Human Factors Research . . . 291
9.2.5 Significance of this Thesis . . . 292
9.3 Recommendations and Future Work . . . 294
9.4 Future Vision and General Conclusion . . . 296

Acknowledgments **299**

Curriculum Vitæ **305**

Outreach Activities **309**

Demo Videos **311**

List of Open Data Sets **313**

List of Publications **315**

Summary

The age of spatial computing is upon us, garnering newfound interest in technologies such as Augmented Reality (AR), which allows us to see virtual objects superimposed on, or integrated into, the surrounding real-world environment. AR has already been employed in various facets of industry and society, such as aerospace, manufacturing, military, education, and museums. While the automotive industry has utilised the technology for in-vehicle technologies, there is little research into applications external to the vehicle. With the proliferation of automated vehicles (AVs) in urban environments, and the need to communicate vehicle intentions to other road users, AR stands out as a possible solution.

In this four-part thesis, the use of AR technology for promoting transparent and trustworthy pedestrian-AV interactions is explored. The main objective was to assess the suitability of AR for such interactions by designing AR interfaces which promote intuitive communication with AVs. Additionally, this thesis developed virtual reality (VR) and AR simulation methods and investigated the interactions between pedestrians and AVs using the designed interfaces. Furthermore, this thesis aimed to identify which design elements are preferred by experiment participants, and the interfaces' ideal mapping locations in the real world, being, on the vehicle (vehicle-locked/mapped), on the road (road-mapped/world-locked), or as heads-up displays (head-locked).

Part I: Design. The first part of the thesis presents a problem statement and design process, commencing with an interview study with experts in the field. The experts' perspectives on the future of AVs, interactions with vulnerable road users (VRUs), smart infrastructure and the potential of AR in future urban traffic are presented. Experts largely agreed that full AV autonomy is not imminent, with smart infrastructure playing a bigger role. They also acknowledged the potential of AR for pedestrian-AV interactions, as an additional secondary cue to the already-present implicit communication, and as a suitable research tool for simulating and testing interfaces. Several promising AR-interface ideas were extracted from the interviews and served as the basis for the design process which produced nine novel AR interfaces. During brainstorming design sessions, three main mapping categories emerged, namely interfaces mapped to the road surface, others mapped to the AV, and interfaces mapped to the pedestrians' field-of-view (FOV), and therefore move with their head. All interfaces were designed with two states, yielding (communicating a safe situation) and non-yielding (communicating an unsafe situation). Each interface was heuristically evaluated, implemented in AR using a novel Unity library, and subsequently demonstrated in a real environment. A sample of the designed interfaces is shown in Figure 1.

Part II: Evaluation. In the second part of the thesis, the focus shifts to a series of user evaluations of the AR interface, which utilise progressively more sophisticated methods to assess the interfaces' effectiveness. Firstly, the results of a large online questionnaire conducted to study the initial feedback on the AR interfaces are presented. Interfaces were measured based on intuitiveness, convincingness, and other aesthetic and usability



Figure 1: A subset of the designed (yielding) interfaces. *Left*: road-mapped, *Augmented zebra crossing*. *Centre*: vehicle-mapped *Phantom vehicle*. *Right*: head-locked *Nudge HUD*.

scales. Initial findings indicated a higher preference for head-locked interfaces and designs incorporating traditional traffic elements, such as traffic light symbols.

An online questionnaire with interfaces displayed as videos does not convey the real-world risks associated with crossing the street. Hence, in the latter part of the evaluation section, the ecological validity was increased and the interfaces were evaluated using a number of simulation methods.

An intermezzo chapter is presented beforehand, which examines the validity of the Large Language Model ChatGPT in analysing qualitative data, as subsequent evaluation chapters use the technology in analysing interview data.

Next, the process of evaluating the interfaces in a highly immersive CAVE pedestrian simulator experiment (HIKER, Leeds University, UK) is presented, where participants cross an actual environment while presented with the AR interfaces as a virtual AV yields (or not) to them. The experiment investigates similar measures to the online questionnaire study, but also focuses, through data collected using head and eye tracking, on how head-contingent attention-attracting circles, utilised to control the participant's initial gaze and head direction, affect the different interface anchoring positions (where the interfaces are mapped in the enviornment). The findings from the CAVE experiment confirmed the initial preferences found in the online questionnaire but also found that AR interfaces that were road-mapped or head-locked, thereby allowing the pedestrian to rely on peripheral vision, elicited quicker crossing initiation times. Furthermore, head-locked and road-mapped interfaces were found to be less affected by pedestrian distraction (initially gazing to the left, centre, or right) than interfaces that were mapped to the vehicle. Interfaces that exploited the participants' peripheral vision also achieved a higher intuitive rating than in the online questionnaire study. In general, the findings of the simulator experiment revealed how head-locked interfaces achieved similar crossing initiation times across different gaze directions, in contrast to road-mapped and vehicle-mapped interfaces. The experiment also highlighted the risk that pedestrians may over-rely on the interfaces when crossing.

The last chapter in the second part of the thesis analyses a number of the interfaces (in different mapping locations) displayed through an AR head-mounted display in a real-world outdoor environment. Once again, the evaluation focuses on the AR interfaces' intuitiveness, trustworthiness, and overall acceptance among participants. In addition, the chapter offers insights into whether previous findings replicate in a real-world environment, and how issues such as the pedestrians' divided attention, proximity-compatibility of the interfaces, and egocentricity vs. allocentricity of the message, affect pedestrians' willingness to cross.

Collectively, the user evaluation found that the performance and salience of world-locked (i.e., road-mapped) interfaces were affected by increased visual clutter in the real



Figure 2: *Left*: A participant looks towards the initial attention-attractor circle during a trial in the pedestrian CAVE simulator (Chapter 6). *Centre*: An augmented vehicle equipped with a yielding vehicle-mapped interface drives in front of a participant during a trial with the AR HMD (Chapter 7). *Right*: The *TransparentVehicle* interface uses the diminished reality concept to make the occluding object translucent (Chapter 8).

environment, which also affected the participants' divided attention and hence their task response time. The chapter outcomes demonstrated the challenges faced when porting AR interfaces from a controlled simulator environment to the attentional demanding real-world environment.

Part III: Diminished Reality. Lastly, the third part of this thesis shifts focus to the concept of Diminished Reality (DR). As suggested by the results of the real-world experiment, outdoor environments may contain visual clutter, making it visually demanding for a pedestrian to detect and interpret AR information. DR, a concept where information is *removed* rather than added, is proposed and evaluated.

This section explores different DR solutions for pedestrian safety in scenarios where visibility is compromised, such as when a parked vehicle obstructs the view. The study assesses six distinct interfaces with varying degrees of local presence (the degree to which a user perceives AR content as being actually there), which aim at preventing collisions and enhancing pedestrian safety. Two interfaces displayed the occluded scenario as a live-feed video and were head-locked, another two displayed the same scenario but were world-locked onto the infrastructure, and a further two interfaces were mapped to the vehicle itself, and either displayed the feed on the rear windshield, or made the vehicle transparent to remove the occlusion itself. High local embeddedness of interfaces, where interfaces were conformed better to the surrounding environment (eg. the transparent vehicle mapping onto the actual vehicle), elicited higher local presence, and were therefore found to be the most favourable to participants. Figure 2 demonstrates the experiments presented in Chapters 6–8.

Part IV: Conclusion. The thesis concludes with synthesising the findings from all parts, discussing their contributions to the overarching aims and research questions. In conclusion, the research of this thesis primarily finds that AR is a solution for promoting safe interactions between pedestrians and AVs, compared to having no form of guidance as evidenced by higher trust and clarity ratings, and quicker crossing reaction times when interfaces are present in comparison to no interface scenarios, across all evaluatory experiments. Furthermore, thanks to the number of diverse evaluation methods used, the findings in this thesis offer deeper insights into how the technology can be properly applied to solve issues surrounding interactions with AVs. Anchoring locations were also investigated in a number of experiments, and findings reveal a number of pros and cons for each mapping category. For instance, HUDs ensure quicker response times but are prone to vergence-accommodation

conflicts (where the brain struggles to reconcile the difference between the focusing of the eyes and the convergence of the eyes' axes when looking at a three-dimensional object or image, therefore causing eye strain), and possible over-reliance. On the other hand, vehicle-mapped interfaces are preferred in a DR solution, due to high local embeddedness. Yet, they are more challenging for AR, especially in cases where pedestrians are distracted. The latter issue also applies to road-mapped interfaces, which still require similar visual attention shifts. Yet, their similarity to current traffic infrastructure in terms of placement and design makes them more preferable than their vehicle-mapped counterparts. In general, these findings offer recommendations on how to design effective interfaces that respect a number of human factor considerations that are imperative for promoting safety and trust, as well as clarity and understandability.

The concluding chapter also offers reflections, guidelines, and recommendations for future research, building upon the comprehensive work presented in the rest of the thesis. The outcomes of this thesis could assist with guiding designers, and human factors engineers in developing interaction techniques with AVs that bridge the real and virtual worlds.

SAMENVATTING

Het tijdperk van ‘spatial computing’ is aangebroken en heeft nieuwe interesse gewekt in technologieën zoals Augmented Reality (AR), waarmee we virtuele objecten kunnen zien in de omgeving. AR wordt al gebruikt in verschillende delen van de industrie en de samenleving, zoals de ruimtevaart, de productiesector, het leger, het onderwijs en musea. Hoewel de technologie door de auto-industrie wordt gebruikt voor technologieën binnenin voertuigen, is er weinig onderzoek naar externe toepassingen. Met de toename van geautomatiseerde voertuigen (oftewel automated vehicles, AV’s) in stedelijke omgevingen, en de noodzaak om de intenties van voertuigen te communiceren met andere weggebruikers, springt AR eruit als een mogelijke oplossing.

In dit vierdelige proefschrift wordt het gebruik van AR-technologie onderzocht voor het bevorderen van transparante en betrouwbare interacties tussen voetgangers en AV’s. De belangrijkste doelstellingen waren het beoordelen van de geschiktheid van AR voor dergelijke interacties door het ontwerpen van AR-interfaces die intuïtieve communicatie met AV’s bevorderen. Daarnaast ontwikkelde dit proefschrift virtual reality (VR) en AR simulatiemethoden en onderzocht de interacties tussen voetgangers en AV’s met de ontworpen interfaces. Verder onderzocht dit proefschrift welke ontwerpelementen de voorkeur hadden en hun ideale mapping (vehicle-locked, world-locked, of head-locked) in de echte wereld.

Deel I: Ontwerp. Het eerste deel van het proefschrift presenteert een probleemstelling en ontwerpproces, te beginnen met een interviewstudie met experts op dit gebied. De perspectieven van de experts op de toekomst van AV’s, interacties met kwetsbare weggebruikers (oftewel vulnerable road users, VRU’s), slimme infrastructuur en het potentieel van AR in toekomstig stedelijk verkeer worden gepresenteerd. De experts waren het er grotendeels over eens dat volledige autonomie voor AV’s niet aanstaande is en dat slimme infrastructuur een grotere rol zal spelen. Ze erkenden ook het potentieel van AR voor voetganger-AV interacties, als een extra secundaire cue naast de al aanwezige impliciete communicatie, en als een geschikt onderzoeksinstrument voor het simuleren en testen van interfaces. Een aantal veelbelovende AR-interface ideeën werden uit de interviews gehaald en dienden als basis voor het ontwerpproces dat negen nieuwe AR-interfaces opleverde. Tijdens de brainstormsessies kwamen drie categorieën naar voren, namelijk interfaces die op het wegdek worden afgebeeld, andere die op de AV worden afgebeeld en als laatste interfaces die op het gezichtsveld (FOV) van de voetganger worden afgebeeld en dus met het hoofd meebewegen. Alle interfaces werden ontworpen met twee staten, ‘yielding’ (geeft aan dat de situatie veilig is) en ‘non-yielding’ (geeft aan dat de situatie niet veilig is). Elke interface werd heuristisch geëvalueerd, geïmplementeerd in AR met behulp van een nieuwe Unity-bibliotheek en vervolgens gedemonstreerd in een echte omgeving. Een voorbeeld van de ontworpen interfaces wordt getoond in figuur 3.

Deel II: Evaluatie. De focus van het tweede deel van het proefschrift ligt op een reeks gebruikersevaluaties van de AR-interface, die steeds geavanceerdere methoden gebruiken om de effectiviteit van de interface te beoordelen. Ten eerste worden de resultaten van



Figuur 3: Een selectie van de ontworpen ('yielding') interfaces. *Links*: road-mapped, *Augmented zebra crossing*. *Midden*: vehicle-mapped *Phantom vehicle*. *Rechts*: head-locked *Nudge HUD*.

een grote online vragenlijst gepresenteerd om de initiële feedback op de AR interfaces te bestuderen. Interfaces werden gemeten op basis van intuïtiviteit, overtuigingskracht en andere esthetische en bruikbaarheidsschalen. De eerste bevindingen wezen op een hogere voorkeur voor head-locked interfaces en ontwerpen die traditionele verkeerselementen bevatten, zoals verkeerslichtsymbolen.

Een online vragenlijst met video's van verkeerssituaties geeft niet de werkelijke risico's van de verkeerssituaties weer. Daarom werd in het laatste deel van het evaluatiegedeelte de ecologische validiteit verhoogd en werden de interfaces geëvalueerd met behulp van een aantal simulatiemethoden.

Voorafgaand wordt een tussenhoofdstuk gepresenteerd, waarin de validiteit van het Large Language Model ChatGPT bij het analyseren van kwalitatieve data wordt onderzocht, omdat in latere evaluatie hoofdstukken gebruik wordt gemaakt van deze technologie bij het analyseren van interview data.

Vervolgens wordt het evaluatieproces van de interfaces in een immersief CAVE voetgangerssimulator experiment (HIKER, Leeds University, UK) gepresenteerd, waarbij deelnemers een virtuele weg moeten oversteken met de AR interfaces terwijl een virtuele AV hen voorlaat (of niet). Het experiment onderzoekt vergelijkbare maten als het online vragenlijstonderzoek, maar richt zich ook, door middel van gegevens die zijn verzameld met behulp van hoofd- en oogtracking, op de manier waarop aandacht aantrekkende cirkels, die worden gebruikt om de initiële blik en hoofdrichting van de deelnemer te controleren, de verschillende interface verankeringsposities beïnvloeden. De bevindingen van het CAVE experiment bevestigden de initiële voorkeuren die gevonden waren in de online vragenlijst, maar vonden ook dat road-mapped of head-locked AR interfaces tot snellere oversteek initiatie leidde omdat de voetganger kon vertrouwen op hun perifere visie. Bovendien bleken head-locked en road-mapped interfaces minder beïnvloed te worden door afleiding van de voetganger (gesimuleerd door eerst naar links, het midden of rechts te kijken) dan interfaces die vehicle-mapped waren. Interfaces die gebruik maakten van het perifere zicht van de deelnemer kregen ook een betere beoordeling voor intuïtieviteit dan in de online vragenlijststudie. In het algemeen lieten de bevindingen van het simulatorexperiment zien hoe head-locked interfaces vergelijkbare aanvangstijden voor oversteken bereikten voor verschillende kijkrichtingen, in tegenstelling tot road-mapped en vehicle-locked interfaces. Het experiment benadrukte ook het risico dat voetgangers te veel vertrouwen op de interfaces bij het oversteken.

Deel III: Diminished Reality. Tot slot wordt in het derde deel van dit proefschrift de focus verlegd naar het concept van Diminished Reality (DR). Zoals gesuggereerd door de resultaten van het praktijkexperiment, kunnen buitenomgevingen visuele rommel bevatten



Figuur 4: *Links*: Een deelnemer kijkt naar de initiële aandacht aantrekkende cirkel tijdens een experiment in de CAVE-simulator voor voetgangers (hoofdstuk 6). *Midden*: Een augmented vehicle uitgerust met een yielding vehicle-mapped interface rijdt voor een deelnemer tijdens een proef met de AR HMD (hoofdstuk 7). *Rechts*: De *TransparentVehicle* interface gebruikt het diminished reality concept om het occluderende object doorschijnend te maken (hoofdstuk 8).

en het visueel veeleisend maken voor de voetganger om AR-informatie te detecteren en interpreteren. DR, een concept waarbij informatie wordt *verwijderd* in plaats van toegevoegd, wordt voorgesteld en geëvalueerd.

Dit hoofdstuk onderzoekt verschillende DR oplossingen voor voetgangersveiligheid in scenario's waar het zicht beperkt is, zoals wanneer een geparkeerd voertuig het zicht belemmert. De studie beoordeelt zes verschillende interfaces met verschillende niveaus van lokale aanwezigheid, die tot doel hebben botsingen te voorkomen en de veiligheid van voetgangers te verbeteren. Twee interfaces toonden het occlusie scenario als een live-feed video en waren head-locked, twee andere toonden hetzelfde scenario maar waren world-locked op de infrastructuur, en nog eens twee interfaces waren gekoppeld aan het voertuig zelf en toonden de feed op de achterruit, of maakten het voertuig transparant om de occlusie zelf te verwijderen. Interfaces met een hoge lokale verankering, waarbij de interfaces meer in overeenstemming waren met de omgeving (bv. het transparante voertuig dat op het eigenlijke voertuig werd afgebeeld), leidden tot een hogere lokale aanwezigheid en waren daardoor het meest aantrekkelijk voor de deelnemers. Figuur 4 toont de experimenten die in de hoofdstukken 6-8 zijn gepresenteerd.

Deel IV: Conclusie. Het proefschrift sluit af met een synthese van de bevindingen uit alle delen, waarbij hun bijdragen aan de overkoepelende doelen en onderzoeksvragen worden besproken. Concluderend kan worden gesteld dat het onderzoek van dit proefschrift voornamelijk aantoont dat AR een oplossing is voor het bevorderen van veilige interacties tussen voetgangers en AV's, in vergelijking met het hebben van geen vorm van begeleiding, zoals blijkt uit de hogere scores voor vertrouwen en duidelijkheid, en snellere oversteekreactietijden wanneer interfaces aanwezig zijn in vergelijking met scenario's zonder interface, in alle evaluerende experimenten. Bovendien bieden de bevindingen in dit proefschrift, dankzij het aantal verschillende gebruikte evaluatiemethoden, diepere inzichten in hoe de technologie op de juiste manier kan worden toegepast om problemen rond interacties met AV's op te lossen. Verankeringslocaties werden ook onderzocht in een aantal experimenten en de bevindingen onthullen een aantal voor- en nadelen voor elke kaartcategorie. HUD's zorgen bijvoorbeeld voor snellere reactietijden, maar zijn gevoelig voor vergence-accomodation conflicten (wat vermoeide ogen veroorzaakt) en mogelijk te veel vertrouwen. Aan de andere kant hebben vehicle-mapped interfaces de voorkeur in een DR-oplossing, vanwege de hoge

lokale verankering, maar zijn ze lastiger voor AR, vooral in gevallen waarin voetgangers worden afgeleid. Dit laatste geldt ook voor road-mapped interfaces, die nog steeds vergelijkbare visuele aandachtsverschuivingen vereisen, maar door hun gelijkenis met de huidige verkeersinfrastructuur in termen van plaatsing en ontwerp hebben ze de voorkeur boven hun vehicle-mapped tegenhangers. In het algemeen bieden deze bevindingen aanbevelingen voor het ontwerpen van effectieve interfaces die rekening houden met een aantal menselijke factoren die noodzakelijk zijn voor het bevorderen van veiligheid en vertrouwen, maar ook duidelijkheid en begrijpelijkheid.

Het afsluitende hoofdstuk biedt ook reflecties, richtlijnen en aanbevelingen voor toekomstig onderzoek, voortbouwend op het uitgebreide werk dat in de rest van het proefschrift is gepresenteerd. De uitkomsten van dit proefschrift kunnen helpen bij het begeleiden van ontwerpers en human factors engineers bij de ontwikkeling van interactietechnieken met AV's die een brug slaan tussen de echte en virtuele wereld.

TAQSIRA

Ninsabu fiż-żmien tal-ispatial computing, żmien li rawwem interess ġdid f'teknoloġiji bħar-Realtà Awmentata (Augmented Reality, AR), li permezz tagħha wieħed jista' jara oġġetti virtwali mqegħdin jew integrati fid-dinja reali ta' madwaru. L-AR diġà ntużat f'diversi oqsma industrijali u soċjetarji, bħall-ispazju tal-ajru, il-manifattura, il-militar, l-edukazzjoni u l-mużewijiet. Għalkemm l-industrija tal-karozzi użat din it-teknoloġija għal tagħmir ġewwa l-vetturi, hemm ftit riċerka dwar l-applikazzjoni ta' din it-teknoloġija fuq barra. Bit-tifrix ta' vetturi awtomatizzati (automated vehicles, AVs) f'ambjenti urbani, u minħabba l-ħtieġa li min ikun fit-triq ikun jaf l-intenzjoni tal-vetturi, l-AR tispikka bħala soluzzjoni potenzjali.

Din it-teżi maqsuma f'erba' partijiet tesplora l-użu tal-AR biex individwi fit-triq ikunu jistgħu jinteraġixxu mal-AVs bi trasparenza u fiduċja. L-għan ewlieni huwa li jiġi evalwat kemm l-AR hija xierqa għal dawn l-interazzjonijiet, billi jitfasslu interfaċċi bl-AR li jħeġġu komunikazzjoni intuwittiva mal-AVs. Barra minn hekk, din it-teżi żviluppat metodi kif ir-Realtà Virtwali (Virtual Reality, VR) u l-AR jistgħu jiġu simulati, u investigat kif individwi fit-triq u l-AVs jinteraġixxu flimkien permezz ta' dawn l-interfaċċi. It-teżi kellha wkoll il-mira li tidentifika liema elementi fid-disinn jippreferu l-parteċipanti fl-esperimenti u l-aktar post xieraq fejn għandhom jidhru l-interfaċċi fid-dinja reali (mapping): fuq il-vettura nnifisha (vehicle-locked/mapped), xi mkien fit-triq (road-mapped/world-locked), jew bħala notifika fil-forma ta' tabella li timxi mad-direzzjoni li l-persuna jkun qed iħares (head-locked).

Parti I: Disinn. L-ewwel parti tat-teżi tippreżenta l-kwistjoni li se tiġi indirizzata u l-proċess tad-disinn, u tibda bi studju fil-forma ta' intervista ma' esperti fil-qasam. Wara dan jiġu ppreżentati l-perspettivi tal-esperti dwar il-futur tal-AVs, l-interazzjonijiet ma' utenti vulnerabbli tat-triq (vulnerable road users, VRUs), infrastruttura intelliġenti u l-potenzjal tal-AR għat-traffiku urban fil-futur. L-esperti fil-biċċa l-kbira qablu li l-AVs fadlilhom biex isiru awtomatizzati għalkollox u li l-infrastruttura intelliġenti għandha rwol akbar x'taqdi. Għarfu wkoll il-potenzjal tal-AR għall-interazzjonijiet bejn l-individwi fit-triq u l-AVs, kemm bħala ħjiel ieħor sekondarju għall-komunikazzjoni impliċita li diġà ssir, kif ukoll bħala għodda xierqa ta' riċerka biex jiġu simulati u ttestjati l-interfaċċi. Mill-intervisti nsiltu bosta ideat promettenti għall-interfaċċi bl-AR u kien fuqhom li ssejjes il-proċess tad-disinn, li tulu nħolqu disa' interfaċċi ġodda tal-AR. F'sessjonijiet ta' qsim ta' ideat feġġew tliet kategoriji ewlenin għall-mapping: interfaċċi integrati mal-art tat-triq, interfaċċi integrati mal-vettura nnifisha, u interfaċċi integrati fil-medda ta' viżjoni (field-of-view, FOV) ta' dak li jkun, jiġifieri jiċċaqalqu skont fejn wieħed iħares. L-interfaċċi tfasslu f'żewġ stati: wieħed li juri li huwa sigur li wieħed jista' jaqsam (yielding), u ieħor li juri li mhux sigur (non-yielding). Kull interfaċċa ġiet evalwata ewristikament, implimentata fl-AR b'librerija innovattiva fuq Unity, u mbagħad tqiegħdet f'ambjent reali. Figura 5 tagħti ħarsa lejn l-interfaċċi li tfasslu.

Parti II: Evalwazzjoni. It-tieni parti tat-teżi titfa' dawl fuq għadd ta' evalwazzjonijiet mill-utenti rigward l-interfaċċi tal-AR, li jużaw metodi dejjem aktar sofistikati biex jevalwaw l-effettività tal-interfaċċi. L-ewwel jiġu ppreżentati r-riżultati ta' kwistjonarju kbir online,



Figure 5: Kampjun tal-interfaċċi li tfasslu (yielding). *Xellug: Augented zebra crossing* b'road-mapping. *Nofs: Phantom vehicle* b'vehicle-mapping. *Lemin: Nudge HUD* b'head-locking.

li sar biex jiġi studjat il-feedback inizjali dwar l-interfaċċi tal-AR. L-interfaċċi tkejlu fuq skali ta' intuwizzjoni, kemm huma konvinċenti, elementi estetiċi u kemm jistgħu jintużaw faċilment. Is-sejbiet inizjali jixhdu preferenza aqwa għall-interfaċċi head-locked u disinji li jinkorporaw elementi tradizzjoni bħal simboli ta' dwal tat-traffiku.

Kwistjonarju online, fejn l-interfaċċi jintwerew bħala videos, ma jurix ir-riskji fid-dinja reali marbutin mal-qsim tat-triq. Għaldaqstant, fl-aħħar parti tal-evalwazzjoni żdiedet il-validità ekoloġika u l-interfaċċi ġew evalwati b'għadd ta' metodi ta' simulazzjoni.

Qabel dan jiġi ppreżentat kapitlu tranżizzjonali, li jeżamina l-validità ta' ChatGPT, Mudell Lingwistiku Kbir (Large Language Model), biex janaliżża data kwalitattiva, peress li kapitli oħra ta' evalwazzjoni jużaw it-teknoloġija biex janalizzaw data tal-intervisti.

Wara jiġi ppreżentat il-proċess fejn ġew evalwati l-interfaċċi f'esperiment ta' simulazzjoni pedonali b'immersjoni għolja bl-użu ta' CAVE (HIKER, Università ta' Leeds, Renju Unit). Hawnhekk, il-parteċipenti jaqsmu "triq" f'ambjent reali bl-interfaċċi tal-AR skont jekk AV virtwali tiqfilhomx jew le. L-esperiment ikejjel elementi simili ta' dawk fil-kwistjonarju online, iżda permezz ta' data li ssegwi l-moviment tar-ras u l-għajnejn, jeżamina wkoll l-effetti ta' ċrieki — intiżi li jiġbdu l-attenzjoni u jikkontrollaw fejn il-parteċipanti jħarsu u jċaqalqu rashom — fuq il-pożizzjonijiet differenti tal-anchoring fl-interfaċċi (jiġifieri fejn l-interfaċċi jkun mmappjati fl-ambjent).

Is-sejbiet mill-esperiment bil-CAVE kkonfermaw il-preferenzi inizjali fil-kwistjonarju online, iżda indikaw ukoll li b'interfaċċi tal-AR b'road-mapping jew head-locking, li bihom l-individwi jistgħu jiddependu fuq il-viżjoni periferika, il-parteċipenti damu inqas biex jaqsmu. Irriżulta wkoll li b'interfaċċi bil-headlocking jew road-mapping, il-parteċipanti ttellfu anqas (eż. billi fil-bidu jħarsu lejn ix-xellug in-nofs jew il-lemin) meta mqabblin ma' interfaċċi integrati mal-vettura nnifisha. L-interfaċċi li għamlu użu mill-viżjoni periferika tal-parteċipanti wkoll ngħataw punt ogħla ta' intuwizzjoni milli fil-kwistjonarju online. B'mod ġenerali, is-sejbiet ta' dan l-esperiment urew kif f'interfaċċi bil-head-locking il-parteċipanti bdew jaqsmu t-triq f'ħinijiet simili irrispettivament 'il fejn ħarsu, għall-kuntrarju ta' interfaċċi bir-road-mapping jew vehicle-mapping. L-esperiment enfasizza wkoll ir-riskju li min jaqsam it-triq jiddipendi żżejjed fuq dawn l-interfaċċi. L-aħħar kapitlu fit-tieni parti tat-teżi analizza interfaċċi differenti permezz ta' apparat viżiv li l-parteċipanti libsu ma' rashom, f'ambjent reali ta' barra. L-evalwazzjoni reġgħet kejlet l-intuwizzjoni u l-affidabbiltà tal-interfaċċi tal-AR kif ukoll kemm il-parteċipanti lesti jaċċettawhom. Barra minn hekk, il-kapitlu jagħti ħjiel jekk is-sejbiet preċedenti humiex l-istess f'ambjent esperimentattiv li mhuwiex virtwali. Juri wkoll kif l-għażla tal-individwi jaqsmux jew le hija affettwata minn kemm qegħdin jagħtu kas x'qed jiġri madwarhom, mill-prossimità tal-messaġġ lejn l-element li l-persuna tinteraġixxi magħha (proximity-compatibility), u jekk il-messaġġ hux



Figure 6: *Xellug:* Parteċipant iħares lejn iċ-ċirku intiz li jiġbed l- attenzjoni waqt prova fis-simulatur pedonali CAVE (Kapitlu 6). *Nofs:* Vettura awmentata bl-interfaċċa yielding u vehicle-mapped tgħaddi minn quddiem parteċipant waqt prova bl-AR HMD (Kapitlu 7). *Lemin:* L-interfaċċa *TransparentVehicle* tuża l-kunċett ta' Realtà Diminwita biex l-oġġett li jxekkel il-viżjoni jsir traslucenti (Kapitlu 8).

indirizzat direttament lill-persuna (egocentric) jew ġenerali (allocentric).

Kollettivament, l-evalwazzjoni sabet li l-prestazzjoni u l-prominenza ta' interfaċċi bil-world-locking (jiġifieri road-mapping) ġew affettwati meta kien hemm aktar oġġetti viżivi mifruxin fl-ambjent reali; dan xekkel l-attenzjoni tal-parteċipanti u għaldaqstant damu aktar biex jagħmlu l-għażla tagħhom. L-eżiti ta' dan il-kapitlu wrew l-isfidi biex, minn ambjent ta' simulazzjoni kontrollat, l-interfaċċi tal-AR jitqiegħdu f'ambjent fid-dinja reali li jitlob attenzjoni liema bħalha.

Parti III: Realtà Diminwita. It-tielet parti ta' din it-teżi tiffoka fuq il-kunċett ta' Realtà Diminwita (Diminished Reality, DR). Kif jissuġġerixxu r-riżultati tal-esperiment f'dinja reali, jista' jkun hemm oġġetti viżivi mifruxin f'ambjenti ta' barra, u b'hekk tkun aktar diffiċli għal dak li jkun biex jagħraf liema informazzjoni ġejja mill-AR u jinterpretaha. Hawhnhekk jiġi propost u evalwat il-kunċett ta' DR, fejn l-informazzjoni *tonqos* minflok tiżdied.

Din it-taqsima tesplora soluzzjonijiet differenti tad-DR għas-sigurtà tal-individwi fit-triq f'sitwazzjonijiet fejn ma jistgħux jaraw daqshekk sew, pereżempju minħabba karozza pparkjata. L-istudju jevalwa sitt interfaċċi distinti b'livelli differenti ta' preżenza lokali, bl-għan li ma jkunx hemm ħbit u biex l-individwi fit-triq ikunu aktar siguri. Żewġ interfaċċi ppreżentaw din is-sitwazzjoni b'viżjoni mnaqqsa bħala video dirett bil-head-locking, tnejn oħra wrew l-istess sitwazzjoni imma bil-world-locking fuq l-infrastruttura, u l-aħħar żewġ interfaċċi tqiegħdu fuq il-vettura nnifisha, fejn il-video deher mal-ħġieġa ta' warra jew il-vettura saret trasparenti biex wieħed ikun jista' jara aħjar. Meta l-interfaċċi kienu integrati sew b'konformità mal-ambjent fil-madwar (local embeddedness, eż. billi l-karozza saret semitrasparenti permezz tad-DR), il-parteċipanti ħassewhom aktar "preżenti" fix-xena. Għaldaqstant instab li dan il-livell għoli ta' integrazzjoni kien l-aktar approċċ vantaġġuż għall-parteċipanti. Figura 6 turi l-esperimenti ppreżentati fil-Kapitli 6–8.

Parti IV: Konklużjoni. It-teżi tagħlaq b'taqsira tas-sejbiet fil-partijiet kollha u diskussjoni tal-kontribut tagħhom għall-għanijiet globali u l-mistoqsijiet ta' riċerka. Bħala konklużjoni, ir-riċerka ta' din it-teżi tixhed fuq kollox li l-AR hija mezz li jħeġġeġ lill-individwi fit-triq jinteraġixxu b'mod sigur mal-AVs, meta mqabbel ma' sitwazzjoni fejn l-individwi ma jingħataw l-ebda ħjiel jew gwida, kif jirriżulta minn punti ogħla ta' fiduċja u ċarezza, u l-fatt li fl-esperimenti evalwatorji kollha l-individwi jdumu inqas biex jaqsmu bl-interfaċċi milli mingħajrhom. Barra minn hekk, grazzi għad-diversi metodi ta' evalwazzjoni li ntużaw, is-sejbiet ta' din it-teżi jagħtu ħarsa aktar profodna lejn kif it-teknoloġija tista' tiġi app-

likata b'mod xieraq biex jissolvew kwistjonijiet ta' interazzjoni mal-AVs. Ġew investigati wkoll il-pożizzjonijiet tal-anchoring f'ċerti esperimenti, u s-sejbiet jixhdu għadd ta' vantaġġi u żvantaġġi għal kull kategorija ta' mapping. Pereżempju, in-nies idumu inqas biex jieħdu deċiżjoni bl-HUDs, iżda jista' jkun li l-moħħ jitfixkel bejn id-distanza mill-għajnejn u l-oġġett simulat fil-VR, u d-distanza mill-għajnejn u l-apparat milbus (vergence-accommodation conflicts), b'hekk in-nies iweġġgħu għajnejhom. Jista' jkun ukoll li n-nies jiddipendu fuq dawn l-HUDs iżżejjed. Min-naħa l-oħra, l-individwi jippreferu interfaċċi b'vehicle-mapping f'kuntest ta' DR minħabba l-integrazzjoni tal-AR fl-ambjent fiżiku (local embeddedness). Dawn madankollu jippreżentaw sfida akbar għall-AR, speċjalment f'każijiet fejn dak li jkun jittellef b'li hemm madwaru. Din tal-aħħar tapplika wkoll għall-interfaċċi b'road-mapping, fejn l-individwi wkoll iridu jagħtu aktar kas x'jiġri madwarhom. Madankollu, peress li dawn l-interfaċċi jixbhu l-infrastruttura attwali tat-traffiku, kemm bħala mapping kif ukoll bħala disinn, hemm preferenza għalihom milli interfaċċi b'vehicle-mapping. B'mod ġenerali, dawn is-sejbiet joffru rakkomandazzjonijiet għal disinn effettiv ta' interfaċċi li jirrispettaw għadd ta' elementi ta' fatturi umani, li huma kruċjali biex irawmu s-sigurtà u l-fiduċja, kif ukoll iċ-ċarezza u fehim min-naħa tal-individwu.

Il-kapitlu ta' konklużjoni jagħti wkoll riflessjonijiet, gwidi u rakkomandazzjonijiet għal riċerka futura, filwaqt li jibni fuq il-ħidma komprensiva li saret fil-kumplament tat-teżi. L-eżiti ta' din it-teżi li jistgħu jkunu ta' għajnuna u gwida lid-disinjaturi, u l-inġiniera li jaħdmu fil-fatturi umani biex jiżviluppa mezzi ta' interazzjoni mal-AVs li jagħlqu d-distakk bejn id-dinja reali u dik virtwali.

1

Introduction

1

RECENT drastic developments in spatial computing technology, such as augmented reality, make it possible to investigate how it can be used to drastically change the way road users interact with automated vehicles. By designing interfaces which promote transparent and trustworthy interactions, we may alleviate the dangers faced in the urban environment as automation increasingly features in our everyday lives.

1.1 THE FUTURE URBAN ENVIRONMENT, AND CONNECTED CITIES

OUR current traffic system is rife with pedestrian accidents, which amounted to over 300,000 or 23% of all road deaths in 2016 alone (WHO, 2018), and where pedestrian error was the leading causality in 80% of all pedestrian-motor vehicle crashes (Hudson et al., 2019). Pedestrians are also nine times more at risk of death than the occupants of a car per kilometre travelled (Global Mobility Report, 2017). One of the collective efforts that attempt to curb these numbers and improve safety on the roads is the concept of smart cities (Hasan and Hasan, 2022). These connected cities leverage technology and data analytics in order to enhance the various aspects of urban living by gathering different systems and services, such as transport, energy, and waste management, and interconnecting them through the Internet of Things (IoT), allowing for optimised, real-time responses to different urban challenges (Zanella et al., 2014). Hence, advanced sensors and communications technology are incorporated into the infrastructure and vehicles in order to collect and analyse data. The gathered information can then be utilised to optimise traffic flows, reduce congestion, and improve safety through enabled new services such as automated vehicles (AVs) and vehicle-pedestrian (V2P) communication technology.

1.2 AUTOMATED VEHICLES

THE increased prevalence of smart sensor technology has enabled the development of automated vehicles which are equipped with various sensors for pedestrian safety (Rasouli and Tsotos, 2019). For a vehicle to be classified as automated, it must possess the ability to accurately determine its location, perceive the surrounding environment, and make informed decisions to navigate the optimal path utilising a process which is commonly referred to as the control loop (Troel-Madec et al., 2019). There are a number of levels (Figure 1.1) of automated capabilities which classify AVs into 6 categories (SAE International, 2021).

SAE levels 0 to 2 pertain to driver support systems, which encompass various functionalities such as automatic emergency braking (level 0), adaptive cruise control (level 1), and adaptive cruise control combined with automatic lane centering (level 2). In these cases, the driver is responsible for maintaining control by steering during adaptive cruise control (ACC), and is expected to supervise and intervene to ensure safety.

On the other hand, vehicles classified as SAE level 3 or higher are characterised by further automated driving capabilities, where the car operates itself and the driver's involvement is limited to specific instances as requested by the system (level 3) or not required at all (levels 4 and 5). The distinction between level 3 to 5 lies in the specific conditions under which the vehicle is expected to function autonomously, such as automated driving in a traffic jam

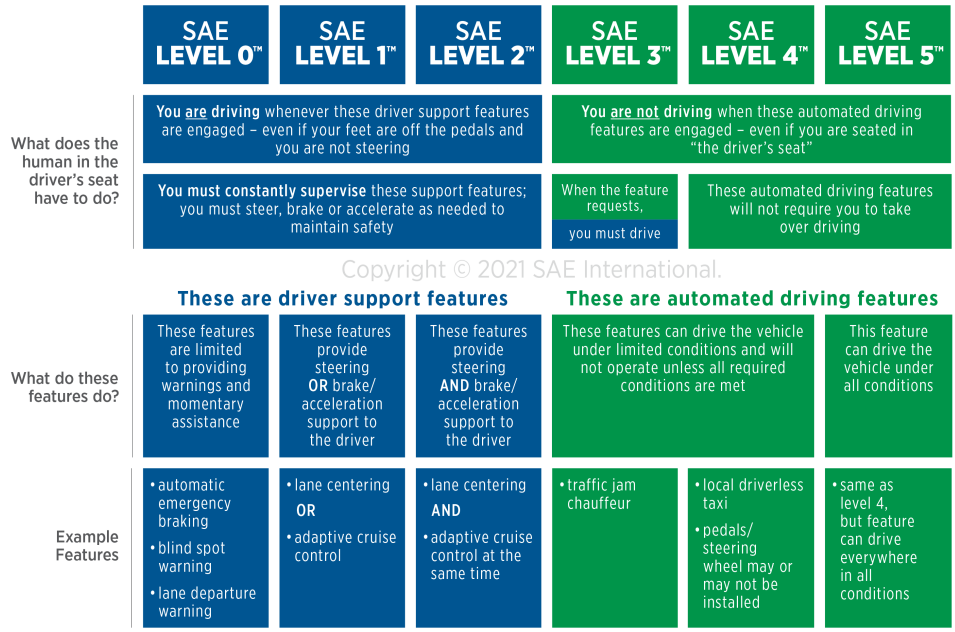


Figure 1.1: The six levels of driving automation as described by SAE International. (SAE, 2021).

(level 3), highly automated driverless taxis (level 4), or full automation in all conditions and locations (level 5). As these vehicles are introduced to the urban environment, the issue of safety arises. In an effort to ensure the safety of the people inside and outside these vehicles, a number of approaches have been developed. For drivers, a considerable amount of research led to the adoption of Automated Advanced Driver Assist Systems (ADAS), and the introduction of Automated Driver Systems (ADS), both of which offer pedestrian detection and crash avoidance on roads (Yu and Marinov, 2020). In the case of pedestrian safety, two main approaches exist: built-in sensor systems and vehicle-pedestrian communication. In the first category, the vehicle makes use of in-built exteroceptive sensors that track information related to the vehicle's surroundings (Schoettle, 2017). Examples include the use of cameras, LiDAR, and Radar, which separately collect different forms of data that allow automated vehicles to safely navigate environments without the need to communicate to any other external actor. On the other hand, in the case of vehicle-pedestrian communication, the vehicle collaborates with various interaction partners, such as intelligent road infrastructure, which exchange information with the vehicle. This collaboration enables data sharing among vehicles within a larger Internet of Things (IoT) transport network (Alam et al., 2017; Troel-Madec et al., 2019). The network facilitates cooperation between these different entities, fostering a collective behaviour that prioritises safety and efficiency for the human counterpart of this interaction loop. The particular target group, in this case, would be a subset of other road users referred to as Vulnerable Road Users (VRUs), which includes non-motorised road users such as pedestrians and cyclists, as well as motorcyclists and persons with disabilities or reduced mobility and orientation (Directive, 2010/40). To successfully

complete the aforementioned interaction loop, the AV must effectively communicate its decisions, derived from the IoT, and its intentions to the VRUs.

1.3 ENCOUNTERS WITH AUTOMATED VEHICLES

1.3.1 IMPLICIT AND EXPLICIT CUES

The flow of traffic and the safety of others could be guaranteed through the proper flow of traffic and the interactions between all road users, since intrinsically driving is a social phenomenon (Rasouli et al., 2017). VRUs could experience confusing scenarios when they interact with AVs due to possible discrepancies between their expectations from AVs and the actual behaviour of the AVs (Nuñez Velasco et al., 2017). Therefore, VRUs, such as pedestrians, tend to rely on implicit and explicit cues from the vehicle as a means of inferring the latter's intent, and therefore whether that vehicle will yield to them or not. Implicit cues include the kinematic motion of the vehicle, such as the velocity, acceleration, deceleration, or the movement and direction of the tires, while examples of explicit cues include eye contact and or hand gestures from the driver, and the use of built-in signals on the vehicle such as flashing of headlamps or indicator lights (Fuest et al., 2017; Schneemann and Gohl, 2016).

Previous studies have investigated the effects of such cues on pedestrian behaviour and found that pedestrians tend to base their decisions mostly on implicit cues (Clamann et al., 2017; Jayaraman et al., 2018; Lee et al., 2021). However, despite the significant importance of implicit communication, explicit cues are still relied upon in ambiguous situations (Kaleefathullah et al., 2022; Lee et al., 2023) so that VRUs build trust and acceptance of the AVs. Hence, while the replacement of drivers with automated and autonomous systems may potentially decrease the risk of human errors and improve overall safety (Chou et al., 2016), the loss of driver cues may decrease pedestrian confidence and trust in interactions with AVs (Lundgren et al., 2017; Matthews et al., 2017).

1.3.2 VEHICLE-TO-PEDESTRIAN (V2P) COMMUNICATION

In an effort to alleviate issues related to loss of driver cues, a subset of Vehicle-to-Everything (V2X), called Vehicle-to-Pedestrian (V2P), has emerged in the automated vehicle industry to facilitate warning VRUs such as pedestrians about approaching vehicles (Tahmasbi-Sarvestani et al., 2017), and vice-versa by warning drivers about pedestrians. In the V2P, the associated actors, such as the vehicles, pedestrians, cyclists etc. communicate by safety messages which are exchanged periodically and contain real-time status and position-related data. An example of V2P systems architecture is presented in Figure 1.2.

Where the interaction between vehicles and pedestrians is concerned, V2P allows for alternatives to the lost explicit cue communication which pedestrians rely upon. One such group of alternatives which researchers have proposed are various communication mechanisms that involve vehicle-mounted interfaces, known as external human-machine interfaces (eHMIs).

1.3.3 EXTERNAL HUMAN-MACHINE INTERFACES (EHMIS)

A number of such eHMIs have been proposed or developed by both academia and industry to serve as a substitute for explicit cues (Bazilinskyy et al., 2019; Dey et al., 2020; Rouchitsas

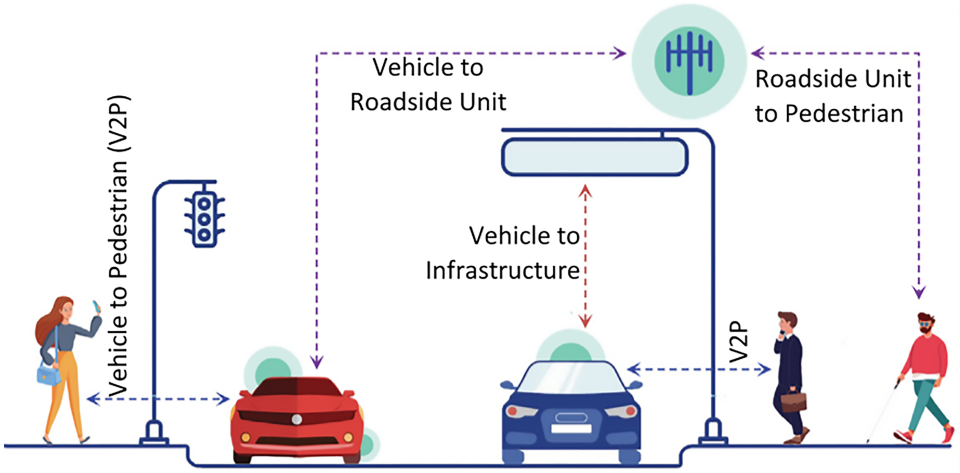


Figure 1.2: Vehicle-to-Pedestrian (V2P) communication architecture (image taken from Hasan and Hasan, 2022).

and Alm, 2019). Examples of eHMI designs include the utilisation of LED strip lights, LED screens, projections, and the use of street infrastructure amongst others. There are also designs which involve the attachment of additional hardware to anthropomorphise the vehicle by representing 'eyes' which follow the pedestrian, and actuated robotic arms to replace the waving arm of a human driver. Figure 1.3 demonstrates a number of examples of such eHMIs.

A number of communication modalities have been explored for eHMIs, mainly visual, auditory, and even haptic feedback (Bazilinskyy et al., 2019, Hasan and Hasan, 2022; Rouchitsas and Alm, 2019). Due to the modal variety, studies have been conducted to contrast their effectiveness, such as whether textual information is more effective than iconographic information, whether anthropomorphic interfaces are better at conveying their message effectively to VRUs than their counterparts (Charisi et al., 2017; Verma et al., 2019), the effectiveness of auditory feedback (Hudson et al., 2019; Ahn et al., 2021), and the effectiveness of haptic feedback from smartphones (Mahadevan et al., 2019).

1.3.4 LIMITATIONS OF EHMIS

Despite the diverse examples of eHMI designs, there are a number of challenges which have been identified. The first challenge pertains to the problem of ambiguity when an AV attempts to communicate with multiple VRUs, such as pedestrians, or vice versa. In such scenarios, pedestrians may be uncertain whether the AV yielded to them, or to another person, or if the AV stopped for an entirely different reason (Löcken et al., 2019; Rasouli and Tsotos, 2019; De Winter and Dodou, 2022). Visual eHMIs may also pose a challenge to pedestrians if they encounter multiple different designs utilised by different car brands to distinguish themselves. In this case, the pedestrians may not be able to interpret the eHMIs in a timely manner (Clamann et al., 2017). Moreover, further risks arise if text is used for eHMIs, as it creates the expectation that the interacting actor (pedestrian) would be literate and cognizant of the language in use (Ackermann et al., 2019; Colley et al., 2019).

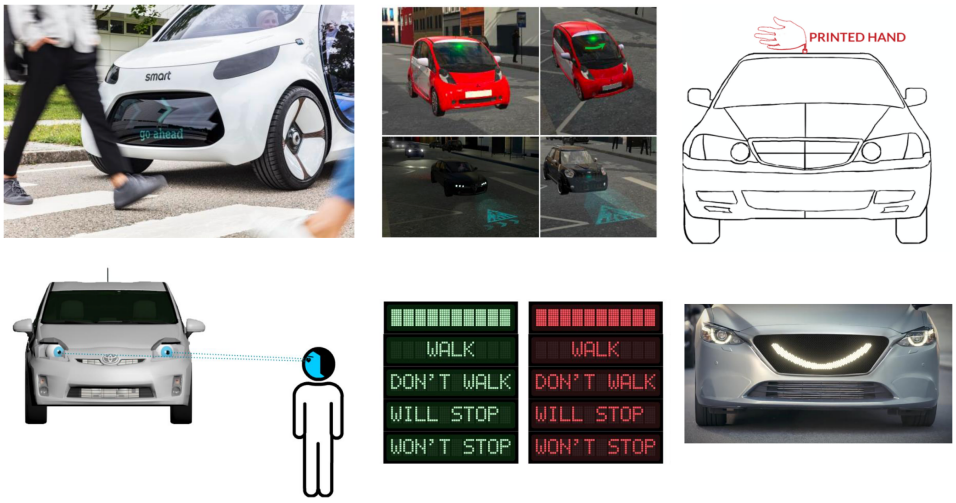


Figure 1.3: Examples of various eHMI designs. From top-left, clockwise: moving text concept (Daimler, 2017), projections on the windshield and the road surface (Dietrich et al., 2018), actuated robotic hand attached to the vehicle (Mahadevan et al., 2018), eyes-on-the-car concept (Chang et al., 2017), text on LED panels attached to the vehicle (Bazilinskyy et al., 2019), and a LED panel with a smiling car concept (Semcon, 2016).

In order to tackle some of these challenges, researchers have suggested a number of possible solutions, such as standardising eHMIs (Rasouli and Tsotos, 2019), changes in traffic policy (Litman, 2021), and other possible solutions where the pedestrian wears sensors which can be used to transmit various warning signals regarding the intentions of the vehicle to the pedestrian (Gordon et al., 2016). In this area, one particular modality which has not been extensively explored, is displaying the vehicle intents through Augmented Reality (AR) technology. The recent popular consumer device developments in the area of AR headsets such as the introduction of the Apple Vision Pro (Apple, 2023), Meta Quest series (Meta, 2023), Microsoft Hololens (Microsoft, 2023) and the Magic Leap (Magic Leap, 2023) amongst others, come with the promise of drastically changing the way we interact in the digital realm. These advances make the development of assistive technologies for pedestrians, built upon the foundational technology used by these devices, an interesting and timely use case to be explored. The overarching focus of this thesis is built upon the aforementioned notion.

1.4 COMPUTER-MEDIATED REALITY

COMPUTER-MEDIATED REALITY (CMR) refers to a spectrum of experiences where a person's perception of reality is altered, augmented, or replaced by computer-generated stimuli displayed on wearable computers or handheld devices (Mann, 1994). The two most common and familiar terms which describe technologies that generate or modify reality are Augmented Reality (AR), and Virtual Reality (VR). AR is defined as supplementing reality rather than replacing it with a virtual world, as it overlays or composites virtual elements on the real world while still allowing the user to still see the real world beyond (Azuma,

2001). On the other hand, VR can be defined as a computer-generated environment that gives the user the illusion of being in a particular location different from where they actually are (Wickens et al., 2021). Other terms in the area of CMR include Diminished Reality (DR), Mixed Reality (MR), and Extended Reality (which is often abbreviated as XR).

These terms have been often utilised inconsistently by industry and academia to refer to different concepts with the same name. For example, the influential Reality-Virtuality Continuum by Milgram and Kishino (1994) places AR and VR on opposite sides of the spectrum that spans from the *real environment* on one side, to the *virtual environment* on the other. The continuum also defines MR as an umbrella term for AR and VR, and in some literature, AR is seen as a setting on a 'mixer' or 'fader' or 'slider' that lies between reality and virtuality (Mann et al, 2018). Other literature states that MR merges both VR and AR (Wedel et al., 2020), whilst other studies contest the Milgram framework to suggest that MR lies between AR and Augmented Virtuality (AV), or MR is viewed as an extension of AR, whereby the latter is available through smartphone apps, whilst the former requires a wearable device (Hoyer et al., 2020). The meaning and abbreviation of XR also remain ambiguous throughout the literature.

Such ambiguity in the definition of these terms and concepts is detrimental to the overall user experience, as argued by Rauschnabel et. al. (2022). What is meant here is that the potential and value of the technology may not be fully realised by the end user (consumer), which restricts the value realisation of the producers. These variables related to customer perceptions need to match with the customer expectations, or else there will be issues related to satisfaction. The latter in turn, is linked to various managerial variables including equity (Poushneh and Vasquez-Parraga, 2017), which is why the user experience is important for managers (Rauschnabel et al., 2022). In an attempt to address the inadequately defined boundaries of AR, VR, and MR, Rauschnabel et. al. (2022) present the xReality conceptual framework which was developed by identifying the existing and often diverging perspectives in both academic and industry literature, and then synthesising the terms in a framework which was externally informed and validated by focus groups and in-depth interviews with industry experts. The authors however acknowledge that as the technology advances, so would the terminology and uses evolve around the technological advances and evolving user experience. At the time of writing this thesis, the framework presented here is found to be sufficient in organising common use cases in the domain of CMR, and classified and organised current use cases and devices. Hence, in this thesis, the xReality framework posited by Rauschnabel et. al. (2022) will be used as the basis of how AR, VR, and their sub-spaces are defined and categorised.

1.4.1 THE XREALITY FRAMEWORK

The framework presented in Figure 1.4 presents XR as an umbrella term with two separate sub-streams (AR and VR) that each have their own continua.

AUGMENTED REALITY

In the case of AR, the continuum ranges from a low functional level (*assisted reality*) to an interactive and realistic experience (*mixed reality*). An example of assisted reality would be displaying a simple text overlay over the real world layer to assist the user in obtaining a better understanding of the physical environment, whilst an example of mixed reality would

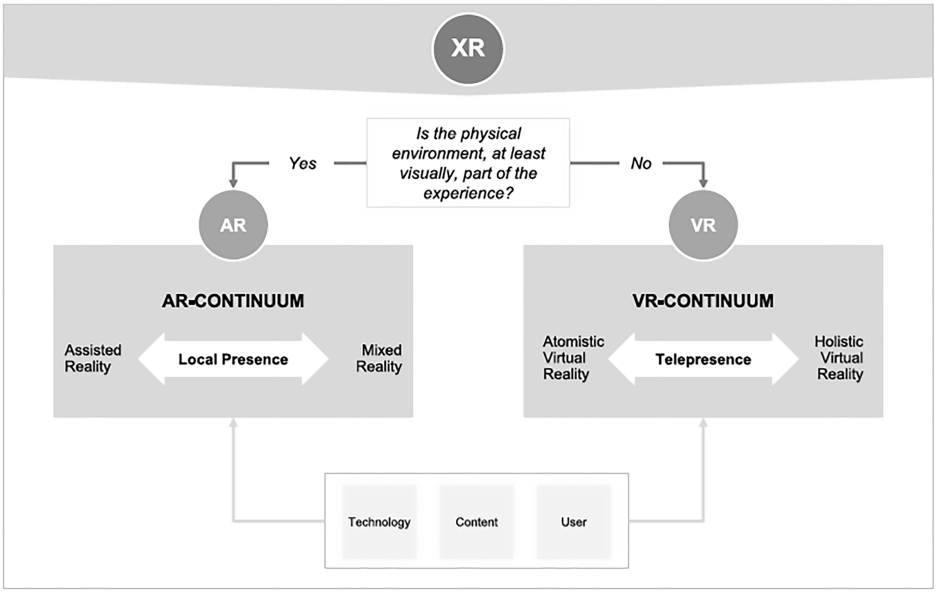


Figure 1.4: The xReality (XR) framework as presented in Rauschnabel et. al. (2022).

refer to an experience which integrates digital objects seamlessly and realistically into the real layer. The differences between assisted and mixed reality are centred upon realism, and therefore how open users are to accept information that goes against their perception of the real world. In the XR framework, the term *local presence* is used to define the degree to which a user deems an AR object as being actually present in their own real environment. Hence, in assisted reality, an object would look very artificial to the user, whilst in mixed reality the virtual object would look like it is actually part of the environment. Virtual content in the real-world could either be defined as being *head-stable* (head-locked) or *world-stable* (world-locked) (Jaekl et al., 2002). Head-locked content moves according to the orientation of the user's head while world-locked content is anchored in a fixed position within the 3D environment. In the latter context of world-stability, the content may also be *persistent*, which means that it would be spatially attached to some specific physical object or a geographical location through GPS coordinates. Finally, the two endpoints of the AR-continuum could also be used to achieve Diminished Reality (DR), where physical objects in the real layer are removed from the user's perception through the virtual layer. Assisted reality would use an unrealistic overlay (such as a black rectangle overlay), while mixed reality would seamlessly remove the object making it difficult for the user to detect it.

VIRTUAL REALITY

On the other side of the XR framework is the VR continuum with *atomistic* and *holistic* VR on the opposite side of the spectrum, and a degree of *telepresence* as the distinguishing factor in the continuum. Telepresence, in contrast to local presence in the AR spectrum, refers in the context of this framework to the degree to which the user feels present in the virtual

environment compared to the physical environment (other existent research uses 'presence', 'virtual presence' or 'mediated presence' to refer to the same concept, as noted in Bourdon, 2020; Mantovani and Riva, 1999; Sheridan, 1992). Based on the concept of telepresence, VR applications can therefore be positioned between being atomistic, where the quality of the VR experience is secondary to some other task, and holistic, where the quality of the VR experience is nearly indistinguishable from the real-world experience (Rauschnabel et. al., 2022).

In the framework described, the continua denoting the different realities discussed are grouped under the umbrella term XR, which through this framework stands for xReality, where *X* serves as a placeholder for *Augmented*, *Assisted*, *Mixed*, *Virtual*, *Atomistic Virtual*, *Holistic Virtual*, or *Diminished* Reality.

1.4.2 EXPERIENCING AUGMENTED REALITY

There are a number of ways that AR could be experienced (refer to Figure 1.5 for examples), namely handheld, head-mounted displays, and spatial AR. In handheld experiences, the augmented layer is superimposed on the real layer on a mobile device, which captures the real layer through its cameras. Head-mounted displays come in two different variations: optical see-through, where the augmented layer is superimposed on the actual real world as viewed through the lens of the headset, and video see-through, where cameras capture the real world and combine it with the augmented layer (Aukstakalnis, 2016). Lastly, spatial AR augments real world objects through the use of digital projectors which display graphical information onto physical objects (Bimber et al., 2005). Moreover, it is imperative to note that the concept of AR can also refer to non-visual modalities such as auditory AR, where for example a user is wearing screenless spectacles with an AI assistant that reacts to auditory commands or triggers. In the context of this thesis, the focus will solely lie on visual AR.

With its different modes, AR has been employed in a number of diverse domains, including industry, the military, museums, tourism, construction and architecture, engineering, medicine, and education (Aukstakalnis, 2016). Most of these applications were custom-built for that particular problem, with the software specifically built to accommodate the needs of the domain. Moving forward, researchers in the domain of spatial computing have been exploring an ultimate aim to allow for pervasiveness and context awareness in AR.

PERVASIVE AND CONTEXT-AWARE AUGMENTED REALITY

Pervasive AR refers to AR experiences which can be accessed continuously and everywhere, and are adapted to the user's current environment. It can be defined as a continuous and pervasive user interface which augments the physical world with digital information registered in 3D while being aware of and responsive to the user's context (Grubert et al., 2017). Example applications include the display of augmented navigation information on a streetscape, where the system would have to make sense of the environment and adapt the information display according to the surroundings, so that important features, such as parked cars are not occluded. A further use case would be the display and interaction with virtual objects during a face-to-face meeting (Grubert et al., 2017).

Pervasive AR can be further enhanced by making it context-aware. Context-aware AR not only takes the location into consideration, but also the user's current situation, needs, preferences, and environmental context. The xReality framework makes provisions for



Figure 1.5: Examples of different ways in which AR can be experienced. From top-left clockwise: handheld AR experience on tablets and smartphones (Apple Developer, 2023), spatial AR projection on a cultural heritage artefact (Ridel et al., 2014), the Meta Quest Pro video see-through headset (Meta, 2023), the Apple Vision Pro video see-through headset (Apple, 2023), the Magic Leap 2 optical see-through headset (Magic Leap, 2023), and the Microsoft Hololens 2 optical see-through head-set (Microsoft, 2023).

context embeddings, whereby context features less in assisted reality, as the user may be required to update when a specific task has been completed. On the other hand, context-awareness is higher in mixed reality, as it recognises and identifies objects in the surroundings and tracks the user's progress.

1.4.3 XR IN PEDESTRIAN-VEHICLE INTERACTION STUDIES

Both VR and AR have been utilised in the study of pedestrian-vehicle interactions. VR experiments can be conducted in a number of ways. Some setups, such as a Cave Automatic Virtual Environment (CAVE) are mainly geared toward research applications and are costly to build. CAVEs are comprised of a large space with projections on all sides, which allow the participant to move around the space naturally (Cruz-Neira et al., 1992). The participant's head movements are tracked and the visual projections of the artificial environment change accordingly to provide the correct perspective.

Despite a number of studies investigating pedestrian-vehicle interactions in a CAVE environment (Kaleefathullah et al., 2022; Schneider et al., 2021), most experiments have been conducted using HMDs (Ackermans et al., 2020; Colley et al., 2021; Deb et al., 2017; De Clercq et al., 2019; Hou et al., 2020; Mahadevan et al., 2019; Mok et al., 2022; Schmitt et al., 2022), or other mediums such as video, images, or audio (Pijnenburg, 2017), presumably because they are faster to set up and are more cost-effective. Despite offering a less realistic experience than the real world, VR provides the researcher with increased experimental control (Parsons, 2015). However, there is a risk that participants' behaviour may also be influenced by the absence of actual risk (Mahadevan et al., 2019; Palmeiro et al., 2018).

On the other hand, AR is able to mimic real-world conditions more closely, enhancing ecological validity. In fact, HMD AR has been employed in trial studies to investigate cyclist behaviour (Matviienko et al., 2022), and also to superimpose virtual vehicles in a real

environment (Hartmann et al. 2018; Maruhn 2020). The utilisation of HMD AR ensures experimental repeatability and increased safety for participants since there are no collision risks.

Moreover, AR has also been applied for vehicle-pedestrian interactions to provide crossing advice. Examples include the presentation of the concepts as superimposed layers on a photo of the streetscape (Hesenius et al., 2018), as concepts (Li et al., 2022) and as part of a VR simulation (Pratticò et al., 2021; Tran et al., 2022; Tran et al., 2023 Malik et al., 2023). However, these concepts were either only described, or evaluated in a VR environment without an implementation or demonstration in a real environment. Furthermore, other studies did not explore how such AR interfaces would affect attention, which location of the environment is best suited to display the interfaces, and what these factors are contingent upon.

Therefore, considering the limitations in previous simulator studies, in the design of AR interfaces, the limitations of eHMIs, and the potential of AR technology, this thesis presents the following motivation to fill these various lacunas.

1.5 RESEARCH GAPS AND MOTIVATION

CONSIDERING the limitations of eHMIs and the potential of pervasive and context-aware AR, there are several reasons why the latter technology can serve as a solution to overcome the aforementioned limitations. This thesis is primarily motivated by a novel approach that involves presenting interfaces aiding in AV-pedestrian interactions as an augmented layer accessible to pedestrians through AR-enabled wearables. Consequently, the eHMI is no longer confined to the automated vehicle or street infrastructure but instead becomes integrated to the VRU.

Information pertaining to the AV's status and intentions could be displayed as augmented information on the user's virtual layer. It is envisioned that such an approach would alleviate various issues mentioned in subsection 1.3.4, which negatively impact eHMIs related to VRU-AV interaction. Since the AR interface would be displayed to every pedestrian individually, each pedestrian would have a separate communication detailing whether a particular AV has issued a communication directly to them. Such a system would alleviate ambiguous situations where pedestrians are uncertain whether the AV's communication was exclusive to them and initiate an erroneous crossing decision that may prove detrimental to the actor in question.

Secondly, the issue of pedestrians not deciphering visual eHMIs in time (Clamann et al., 2017), and the language barrier problem (Ackermann et al., 2019; Colley et al., 2019) would also potentially be alleviated with this approach since the system would allow for various customisations to be made to the HUD, such as the user's preferred information modality (eg. text or iconography) and language. Furthermore, this approach may alleviate the cognitive load on pedestrians when encountering varying eHMI designs according to car brand since an individual pedestrian would experience the same interface for each AV. Such an approach complies with the vision for standardising explicit communication such as eHMIs and having a universal system for every car (Rasouli and Tsotos, 2019).

Further to the designing of new interfaces, there are a number of questions surrounding the use of AR in this domain, which have not yet been tackled by previous studies. Mainly, *it is still unclear whether AR would offer any tangible benefits in increasing pedestrian*

safety and trust in interacting with AVs as they cross a road. Moreover, it is also unclear *which kind of AR technology is best suited for the challenge, and which design elements and anchoring positions (ie. head-locked, or world-locked) are preferred (ie. which locations in the environment are best to integrate information on)*. In the latter case, it is still to be investigated on whether anchoring of the information is contingent, if at all, on the location of the traffic, and the pedestrian's attention.

Lastly, the thesis is also motivated to *investigate the use of various simulation and evaluation methodologies to validate any developed AR interfaces, and the evaluation methods themselves with increased levels of ecological validity*. Research in the domain has been conducted in both lower-cost online studies, and also in relatively expensive multi-million dollar immersive VR setups, such as CAVEs. However, the notion of what are the different requirements for the choice between the different approaches is still unclear. Hence, is it always necessary to go for a higher ecological validity to evaluate? A further element which merits investigation is *the impact of Large Language Models (LLMs) as a tool to facilitate the analysis of data collected in human-robot interaction (HRI) and human-computer interaction (HCI) research*, such as data collected from interviews, questionnaires, think-aloud protocols etc.

Therefore, the thesis is mainly motivated by three aspects; to develop novel AR interfaces for pedestrian-vehicle interactions, to evaluate the interfaces with increased ecological validity, and to validate and contrast the validation methods themselves. Hence, the following objectives and research questions are presented.

1.6 OBJECTIVES OF THIS THESIS

1. **Design augmented reality interfaces which enhance trust in automated vehicles.**
   In particular, *What are the user-preferred design elements for AR interfaces which facilitate communication between an AV and vulnerable road users?*

2. **Develop virtual and augmented reality simulation methods to investigate the interaction between automated vehicles and vulnerable road users.**
   Specifically, *What are the different use cases for different simulation methods?*
   And *Does the ecological validity of evaluation methods affect trust?*

3. **Identify the best way to present and anchor the information in AR, and determine which kind of AR design is best suited for interfaces related to pedestrian-vehicle interaction.**
   Therefore, *Which interface types, and anchoring positions are preferred?*

4. **Assess the suitability of augmented reality technology for the development of interfaces which facilitate automated vehicle-vulnerable road user communication.**
   Hence, *What is the potential of augmented reality in future mixed traffic?*
   And, *Do humans trust indirect communication from a machine?*

1.7 THESIS OUTLINE

THE main content of the thesis is split into four parts, which reflect the main processes undertaken in order to reach the objective set in the preceding section. Each part

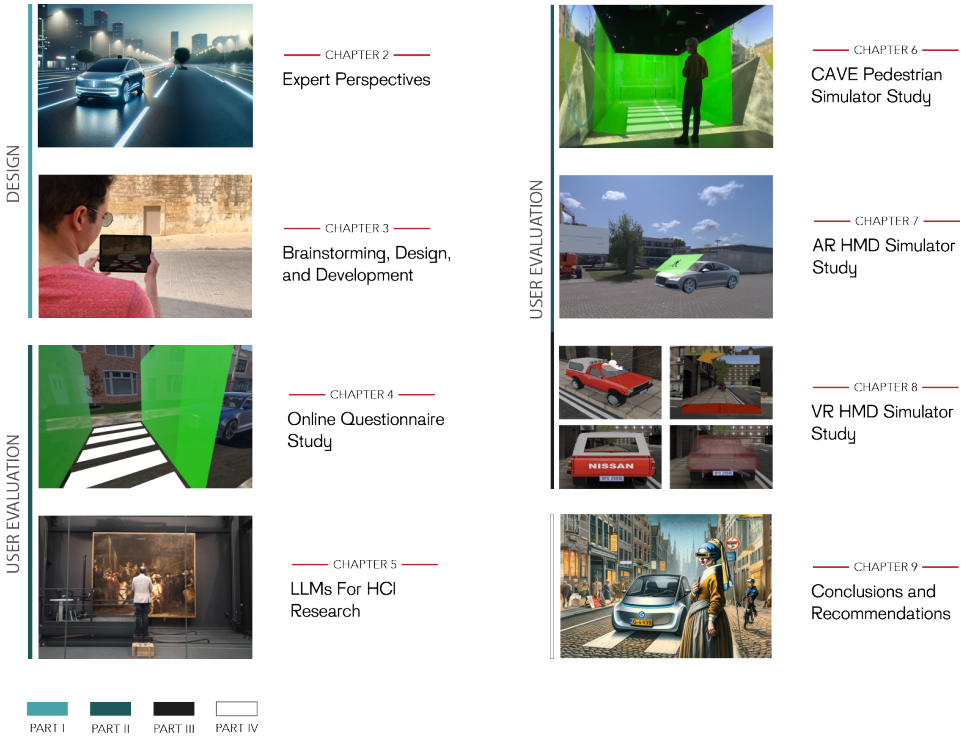


Figure 1.6: A diagram illustrating how the work presented in this thesis is split into four major parts, each containing relevant chapters related to a process. The coloured bars to the left of the chapters represent the different parts, while text labels to the left of the bars highlight the process.

groups a number of relevant chapters according to the theme of the process. Part I contains chapters related to the design process, Part II groups chapters related to the user evaluation process, Part III contains a further user evaluation chapter, which treats Diminished Reality, and finally, Part IV presents the final chapter containing recommendations and concluding reflections. Figure 1.6 illustrates the thesis outline as described in this section.

PART I: DESIGN

The first part of the thesis covers the design aspects of the overall investigation, where an exploratory interview study was conducted with experts in the field to gather requirements. Following the requirements gathering, a prototyping exercise was initiated in order to design novel augmented reality interfaces for pedestrian-vehicle interactions.

CHAPTER 2: VULNERABLE ROAD USERS AND THE COMING WAVE OF AUTOMATED VEHICLES: EXPERT PERSPECTIVES

Chapter 2 presents the results from an interview study in which 16 researchers were invited to give their perspectives on automated vehicles (AVs), and the interaction with vulnerable road users (VRUs) in the future urban environment. Aspects such as smart infrastructure, external human-machine interfaces (eHMIs) and the potential of augmented reality were

addressed during the interviews. The aim of this chapter was to examine the opportunities and challenges in these respective domains from the viewpoint of the scientific experts and to focus on the potential of utilising AR in future mixed traffic (RQ1).

CHAPTER 3: TOWARDS FUTURE PEDESTRIAN-VEHICLE INTERACTIONS: INTRODUCING THEORETICALLY-SUPPORTED AR PROTOTYPES

In Chapter 3, the process behind designing novel AR interfaces for pedestrian-vehicle interactions is presented. A genius-based (experience based) design method was used for the overall design process, split into a four-phase process. In these phases, concept ideas were extracted from the expert interviews, and then further expanded with previous literature and established theories, through which nine AR interfaces emerged and were designed, implemented and demoed in a real environment as part of the final two phases.

PART II: EVALUATION

The second part of the thesis presents the different methods which were utilised in order to evaluate the use of augmented reality interfaces for pedestrian-vehicle interactions. Each subsequent chapter introduces a higher level of ecological validity. In some of the evaluation methods, large amounts of qualitative data were collected. In an effort to evaluate this data efficiently, and extract key elements, novel evaluation methods that make use of Large Language Models (LLMs) were utilised. Hence, an intermediate chapter (Chapter 5) was written as a preamble to Chapters 6 and 7, to serve as a validation of the method prior to its use in the subsequent chapters.

CHAPTER 4: AUGMENTED REALITY INTERFACES FOR PEDESTRIAN-VEHICLE INTERACTIONS: AN ONLINE STUDY

Chapter 4 presents an online questionnaire study which was conducted in order to get a first impression on respondent acceptance (Objectives 1 and 4) of the AR interfaces for pedestrian-vehicle interactions which were developed as part of the experiment presented in Chapter 3, and investigate their preferences on the aesthetics and behaviour of the interfaces amongst other metrics. The questionnaire was shared across five European countries pertaining to the locations of the main SHAPE-IT project consortium member institutions (Germany, The Netherlands, Norway, Sweden, and The United Kingdom), with each country having its own customs and behaviours in traffic interactions. The study was successfully completed by 992 respondents, equally distributed across the five countries; such a number of responses allowed for higher statistical power in the results obtained.

CHAPTER 5: USING CHATGPT FOR HUMAN-COMPUTER INTERACTION RESEARCH: A PRIMER

Chapter 5 presents a number of use cases for Large Language Models (LLMs) in evaluating qualitative data to extract salient conclusions. In the chapter, an LLM was applied to questionnaire textbox data, interview transcript data, and think-aloud transcribed data. As LLMs became more widely accessible following the conclusion of the online questionnaire experiment presented in Chapter 4, in Chapter 5, a comparison between the methods used in the previous chapter and the application of an LLM to the free-text qualitative data captured in the questionnaire is presented. Furthermore, as qualitative interview data was

envisioned as part of any further analysis methods with higher ecological validity than the online questionnaire study, this chapter serves as a validation of the method and examines the potential benefits of applying LLMs to the evaluation stack. Hence, it can be considered as a prelude to Chapters 6 and 7 which both made use of LLMs in the analysis of qualitative data capture.

CHAPTER 6: IMMERSIVE INSIGHTS - EVALUATING AUGMENTED REALITY INTERFACES FOR PEDESTRIANS IN A CAVE-BASED EXPERIMENT

In Chapter 6, results from a pedestrian simulator experiment are presented. The interfaces were evaluated by 30 participants in a Cave automatic virtual environment (CAVE) to examine whether the prior online evaluation of the AR interfaces presented in Chapter 4 could be replicated in an immersive virtual environment, and if AR design effectiveness depends on pedestrian attention allocation. Furthermore, the study explored the different use cases for diverse simulation methods (Objective 2), and the effects of participant distraction, and multiple interface anchoring positions (Objective 3).

CHAPTER 7: AUGMENTED REALITY FOR SUPPORTING THE INTERACTION BETWEEN PEDESTRIANS AND AUTOMATED VEHICLES: AN EXPERIMENTAL OUTDOOR STUDY

Chapter 7 is the culmination of increasing the ecological validity in the evaluation process of the designed AR interfaces. In this chapter, the evaluation of the AR interfaces for pedestrian-vehicle interactions goes a step further from evaluating using online or VR methods, by user testing a subset of the interfaces in the real-world using an AR HMD. Participants were tasked with pressing a button to indicate whether they felt safe to cross whilst wearing an AR HMD which displayed a driving AV that did or did not yield. The interfaces were also displayed in certain conditions, together with a baseline condition with no interface. Participants were then tasked to rate the interfaces based on intuitiveness, convincingness, trustworthiness, clarity/understandability, and visual attractiveness, as well as a 9-item acceptance scale, which are similar to the scales used in Chapter 4. Once again, the Chapter explores the different use cases for different simulation methods (Objective 2), interface anchoring positions (Objective 3), as well as the question of trust in such interfaces by participants (Objective 4).

PART III: DIMINISHED REALITY

The third and final part of the thesis looks beyond the main research questions and objectives and offers an additional study which also explores the effect of removing information through the use of Diminished Reality (a point raised in Chapter 2) from the real world, in contrast to the previous chapters which only added information to the real world layer.

CHAPTER 8: HEAD-LOCKED, WORLD-LOCKED, OR CONFORMAL DIMINISHED-REALITY? AN EXAMINATION OF DIFFERENT AR SOLUTIONS FOR PEDESTRIAN SAFETY IN OCCLUDED

Chapter 8 presents a study which explored the effectiveness of interfaces for preventing collisions between pedestrians and cars due to poor visibility, specifically when a parked vehicle obstructs the view. In the presented VR study, participants were tasked to hold a key on the keyboard to indicate whether they felt safe to cross at a particular moment. The

conditions presented were a live video feed of the occluded region presented directly on the HMD, a video feed displayed across the street, and two conformal diminished reality designs: a see-through display on the occluding vehicle, and a solution where the occluding vehicle has been made semi-transparent. A baseline condition without augmented information served as a reference. Furthermore, some of the conditions were tested with and without the addition of an on-screen guiding arrow indicating the location of the approaching vehicle. The chapter once again explores the potential of AR for such pedestrian-vehicle interactions (Objective 4), and also the preferred interface types and anchoring positions (Objectives 1 and 3).

PART IV: CONCLUSION

CHAPTER 9: CONCLUSIONS AND RECOMMENDATIONS

The final chapter of the thesis presents the key outcomes of all the chapters presented in Parts I - III, and a discussion of how their separate conclusions tie into the overall arching aims, goals, and research questions of this thesis. Finally, a number of reflections and recommendations on potential future research, which builds upon the work presented in these chapters, are presented, followed by a final, conclusive remark.

REFERENCES

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, 75, 272-282. https://doi.org/10.1016/j.apergo.2018.11.002

Ackermans, S., Dey, D., Ruijten, P., Cuijpers, R. H., & Pfleging, B. (2020, April). The effects of explicit intention communication, conspicuous sensors, and pedestrian attitude in interactions with automated vehicles. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-14). https://doi.org/10.1145/3313831.3376197

Ahn, S., Lim, D., & Kim, B. (2021). Comparative study on differences in user reaction by visual and auditory signals for multimodal eHMI design. In HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part III 23 (pp. 217-223). Springer International Publishing. https://doi.org/10.1007/978-3-030-78645-8_27

Alam, F., Mehmood, R., Katib, I., Albogami, N. N., & Albeshri, A. (2017). Data fusion and IoT for smart ubiquitous environments: A survey. IEEE Access, 5, 9533–9554. https://doi.org/10.1109/ACCESS.2017.2697839

Apple. (2023). Introducing Apple Vision Pro: Apple's first spatial computer. https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/

Apple Developer. (2023). Augmented Reality - Apple Developer. https://developer.apple.com/augmented-reality/

Aukstakalnis, S. (2016). Practical augmented reality: A guide to the technologies, applications, and human factors for AR and VR. Addison-Wesley Professional.

Bazilinskyy, P., Dodou, D., & De Winter, J. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. Transportation Research Part F: Traffic Psychology and Behaviour, 67, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

Bimber, O., & Raskar, R. (2005). Spatial augmented reality: merging real and virtual worlds. CRC press.

Bourdon, J. (2020). From correspondence to computers: A theory of mediated presence in history. Communication Theory, 30(1), 64-83. https://doi.org/10.1093/ct/qtz020

Chang, C. M., Toda, K., Sakamoto, D., & Igarashi, T. (2017, September). 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 (pp. 65-73). https://doi.org/10.1145

/3122986.3122989

Charisi, V., Habibovic, A., Andersson, J., Li, J., & Evers, V. (2017). Children's views on identification and intention communication of self-driving vehicles. In Proceedings of the 2017 conference on interaction design and children (pp. 399-404). https://doi.org/10.1145/3078072.3084300

Chou, F., Shladover, S. E., & Bansal, G. (2016). Coordinated merge control based on V2V communication. https://doi.org/10.1109/vnc.2016.7835933

Clamann, M., Aubert, M., & Cummings, M. L. (2017). Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles (No. 17-02119).

Colley, M., Li, S., & Rukzio, E. (2021). Increasing pedestrian safety using external communication of autonomous vehicles for signalling hazards. In Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (pp. 1-10). https://doi.org/10.1145/3447526.3472024

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. Communications of the ACM, 35(6), 64-73.

Colley, M., Walch, M., Gugenheimer, J., & Rukzio, E. (2019). Including people with impairments from the start: External communication of autonomous vehicles. In Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: adjunct proceedings (pp. 307-314). https://doi.org/10.1145/3349263.3351521

Daimler. (2017). Autonomous concept car smart vision EQ fortwo.https://media.mbusa.com/releases/release-80848dccd3f3680a764667ad530987e9-autonomous-concept-car-smart-vision-eq-fortwo

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., & Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. Applied ergonomics, 65, 449-460. https://doi.org/10.1016/j.apergo.2017.03.007

De Clercq, K., Dietrich, A., Núñez Velasco, J. P., De Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. Human factors, 61(8), 1353-1370. https://doi.org/10.1177/0018720819836343

De Winter, J., & Dodou, D. (2022). External human-machine interfaces: Gimmick or necessity? Transportation Research Interdisciplinary Perspectives, 15. https://doi.org/10.1016/J.trip.2022.100643

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., & Terken, J. (2020). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives, 7, 100174. https://doi.org/10.1016/j.trip.2020.100174

Dietrich, A., Willrodt, J. H., Wagner, K., & Bengler, K. (2019). Projection-based external human machine interfaces-enabling interaction between automated vehicles and pedestrians. In Proceedings of the DSC 2018 Europe VR.

Directive 2010/40. The framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport. European Parliament, Council of the European Union. Retrieved from http://data.europa.eu/eli/dir/2010/40/oj

Fuest, T., Sorokin, L., Bellem, H., Bengler, K. (2018). Taxonomy of Traffic Situations for the Interaction between Automated Vehicles and Human Road Users. In: Stanton, N. (eds) Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing, vol 597. Springer, Cham. https://doi.org/10.1007/978-3-319-60441-1_68

Global Mobility Report. (2017). Sum4all. https://www.sum4all.org/publications/global-mobility-report-2017

Gordon, S. M., Kozolski R. J., Kundu, A., Malkin, K. P., Pickover, A. C. (2016). Automated control of interactions between self-driving vehicles and pedestrians (U.S. Patent 9483948).

Grubert, J., Langlotz, T., Zollmann, S., & Regenbrecht, H. (2016). Towards pervasive augmented reality: Context-awareness in augmented reality. IEEE transactions on visualization and computer graphics, 23(6), 1706-1724. https://doi.org/10.1109/TVCG.2016.2543720

Hartmann, M., Viehweger, M., Stolz, M., Watzenig, D., Spitzer, M., & Desmet, W. (2018). "Pedestrian in the Loop": An approach using augmented reality (No. 2018-01-1053). SAE Technical Paper. https://doi.org/10.4271/2018-01-1053

Hasan, R., & Hasan, R. (2022). Pedestrian safety using the Internet of Things and sensors: Issues, challenges, and open problems. Future generation computer systems, 134, 187-203. https://doi.org/10.1016/j.future.2022.03.036

Hesenius, M., Börsting, I., Meyer, O., & Gruhn, V. (2018). Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. In Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct (pp. 261-268). https://doi.org/10.1145/3236112.3236148

Hou, M., Mahadevan, K., Somanath, S., Sharlin, E., & Oehlberg, L. (2020). Autonomous vehicle-cyclist interaction: Peril and promise. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-12). https://doi.org/10.1145/3313831.3376884

Hoyer, W. D., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the customer experience through new technologies. Journal of interactive marketing, 51(1), 57-71. https://doi.org/10.1016/j.intmar.2020.04.001

Hudson, C. R., Deb, S., Carruth, D. W., McGinley, J., & Frey, D. (2019). Pedestrian perception of autonomous vehicles with external interacting features. In Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2018 International Conference on Human Factors and Systems Interaction, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9 (pp. 33-39). Springer International Publishing. https://doi.org/10.1007/978-3-319-94334-3_5

Jaekl, P. M., Allison, R. S., Harris, L. R., Jasiobedzka, U. T., Jenkin, H. L., Jenkin, M. R., & Zikovitz, D. C. (2002, March). Perceptual stability during head movement in virtual reality. In Proceedings IEEE virtual reality 2002 (pp. 149-155). IEEE. https://doi.org/10.1109/VR.2002.996517

Jayaraman, S. K., Creech, C., Robert Jr, L. P., Tilbury, D. M., Yang, X. J., Pradhan, A. K., & Tsui, K. M. (2018). Trust in AV: An uncertainty reduction model of AV-pedestrian interactions. In Companion of the 2018 ACM/IEEE international conference on human-robot interaction (pp. 133-134). https://doi.org/10.1145/3173386.3177073

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & Winter, J. D. (2022). External human–machine interfaces can be misleading: An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human factors, 64(6), 1070-1085. https://doi.org/10.1177/0018720820970751

Lee, Y. M., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C., ... & Merat, N. (2021). Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. Cognition, Technology & Work, 23, 367-380. https://doi.org/10.1007/s10111-020-00635-y

Lee, Y. M., Sidorov, V., Madigan, R., de Pedro, J. G., Markkula, G., & Merat, N. (2023). Hello, is it me you're stopping for? The effect of eHMI familiarity on pedestrians' crossing behaviour in an ambiguous situation. https://doi.org/10.31234/osf.io/2eyz4

Li, Y., Kamalasanan, V., Batista, M., & Sester, M. (2022). Improving pedestrian priority via grouping and virtual lanes. arXiv. https://arxiv.org/abs/2205.08783

Litman, T., 2021. Autonomous vehicle implementation predictions. Victoria Transport Policy Institute. https://www.vtpi.org/avip.pdf

Löcken, A., Golling, C., & Riener, A. (2019). How should automated vehicles interact with pedestrians? A comparative analysis of interaction concepts in virtual reality. In Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications (pp. 262-274). https://doi.org/10.1145/3342197.3344544

Lundgren, V. M., Habibovic, A., Andersson, J., Lagström, T., Nilsson, M., Sirkka, A., Fagerlönn, J., Fredriksson, R., Edgren C., Krupenia, S., & Saluäär, D. (2017). Will there be new communication needs when introducing automated vehicles to the urban context?. In Advances in Human Aspects of Transportation: Proceedings of the AHFE 2016 International Conference on Human Factors in Transportation, July 27-31, 2016, Walt Disney World, Florida, USA (pp. 485-497). Springer International Publishing. https://doi.org/10.1007/978-3-319-41682-3_41

Magic Leap. (2023). Magic Leap 2. https://www.magicleap.com/magic-leap-2

Mahadevan, K., Sanoubari, E., Somanath, S., Young, J. E., & Sharlin, E. (2019). AV-Pedestrian interaction design using a pedestrian mixed traffic simulator. In Proceedings of the 2019 on designing interactive systems conference (pp. 475-486). https://doi.org/10.1145/3322276.3322328

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-12). https://doi.org/10.1145/3173574.3174003

Mann, S. (1994). Wearable, tetherless, computer-mediated reality (with possible future applications to the disabled). Technical report 260.

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing pedestrian behavior in augmented reality—proof of concept. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 313-321). IEEE. https://doi.org/10.1109/VR46266.2020.00051

Malik, J., Kim, N. Y., Parr, M. D., Kearney, J. K., Plumert, J. M., & Rector, K. (2023). Do simulated augmented reality overlays influence street-crossing decisions for non-mobility-impaired older and younger adult pedestrians? Human factors. https://doi.org/10.1177/00187208231151280

Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. arXiv preprint arXiv:1804.08386. https://doi.org/10.48550/arXiv.1804.08386

Mantovani, G., & Riva, G. (1999). "Real" presence: how different ontologies generate different criteria for presence, telepresence, and virtual presence. Presence, 8(5), 540-550. https://doi.org/10.1162/105474699566459

Matthews, M., Chowdhary, G., & Kieson, E. (2017). Intent communication between autonomous vehicles and pedestrians. arXiv preprint arXiv:1708.07123. https://doi.org/10.48550/arXiv.1708.07123

Matviienko, A., Müller, F., Schön, D., Seesemann, P., Günther, S., & Mühlhäuser, M. (2022, April). BikeAR: Understanding cyclists' crossing decision-making at uncontrolled intersections using Augmented Reality. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-15). https://doi.org/10.1145/3491102.3517560

Meta. (2023). Meta Quest Pro. https://www.meta.com/nl/en/quest/quest-pro/

Microsoft. (2023). Microsoft Hololens - Mixed Reality technology for Business. https://www.microsoft.com/en-us/hololens

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE TRANSACTIONS on Information and Systems, 77(12), 1321-1329.

Mok, C. S., Bazilinskyy, P., & de Winter, J. (2022). Stopping by looking: A driver-pedestrian interaction study in a coupled simulator using head-mounted displays with eye-tracking. Applied ergonomics, 105, 103825. https://doi.org/10.1016/j.apergo.2022.103825

Nuñez Velasco, P., Farah, H., van Arem, B., & Hagenzieker, M. (2017). Interactions between vulnerable road users and automated vehicles: A synthesis of literature and framework for future research. In Proceedings of the road safety and simulation international conference (Vol. 2017, pp. 16-19).

Palmeiro, A. R., van der Kint, S., Vissers, L., Farah, H., de Winter, J. C., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. Transportation research part F: traffic psychology and behaviour, 58, 1005-1020. https://doi.org/10.1016/j.trf.2018.07.020

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Frontiers in human neuroscience, 9, 660. https://doi.org/10.3389/fnhum.2015.00660

Pijnenburg, J. (2017). Naturalism: Effects of an intuitive augmented reality interface property in the display of automated driving status. Master's thesis. Delft University of Technology, Delft, The Netherlands. http://resolver.tudelft.nl/uuid:92d6b140-56f7-440a-b9ba-ff997c00ab4b

Poushneh, A., & Vasquez-Parraga, A. Z. (2017). Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. Journal of Retailing and Consumer Services, 34, 229-23. https://doi.org/10.1016/j.jretconser.2016.10.005

Pratticò, F. G., Lamberti, F., Cannavò, A., Morra, L., & Montuschi, P. (2021). Comparing state-of-the-art and emerging augmented reality interfaces for autonomous vehicle-to-pedestrian communication. IEEE Transactions on Vehicular Technology, 70(2), 1157-1168. https://doi.org/10.1109/TVT.2021.3054312

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017). Understanding pedestrian behavior in complex traffic scenes. IEEE Transactions on Intelligent Vehicles, 3(1), 61-70. https://doi.org/10.1109/TIV.2017.2788193

Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE transactions on intelligent transportation systems, 21(3), 900-918. https://doi.org/10.1109/TITS.2019.2901817

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. Computers in human behavior, 133, 107289. https://doi.org/10.1016/j.chb.2022.107289

Ridel, B., Reuter, P., Laviole, J., Mellado, N., Couture, N., & Granier, X. (2014). The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. Journal on Computing and Cultural Heritage (JOCCH), 7(2), 1-18. https://doi.org/10.1145/2611376

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. Frontiers in Psychology, 10, 2757. https://doi.org/10.3389/fpsyg.2019.02757

SAE International. (2021). SAE Levels of Driving AutomationTM Refined for Clarity and International Audience. https://www.sae.org/blog/sae-j3016-update

Schmitt, P., Britten, N., Jeong, J., Coffey, A., Clark, K., Kothawade, Sunil S., Grigore C. E., Khaw A., Konopka C., Pham L., Ryan K., Schmitt C., & Frazzoli, E. (2021). Can cars gesture? A case for expressive behavior within autonomous vehicle and pedestrian interactions. IEEE Robotics and Automation Letters, 7(2), 1416-1423. https://doi.org/10.1109/LRA.2021.3138161

Schneemann, F., & Gohl, I. (2016). Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. In 2016 IEEE intelligent vehicles symposium (IV) (pp. 38-43). IEEE. https://doi.org/10.1109/IVS.2016.7535361

Schneider, S., Maruhn, P., Dang, N. T., Pala, P., Cavallo, V., & Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and

Head-Mounted Displays. Human Factors. https://doi.org/10.1177/001872082098744

Schoettle, B. (2017). Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles. University of Michigan.

Semcon (2016). The Smiling Car - Self driving car that sees you. https://semcon.com/smiling car/

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. Presence Teleoperators Virtual Environ., 1(1), 120-125. https://doi.org/10.1162/pres.1992.1.1.120

Tahmasbi-Sarvestani, A., Mahjoub, H. N., Fallah, Y. P., Moradi-Pari, E., & Abuchaar, O. (2017). Implementation and evaluation of a cooperative vehicle-to-pedestrian safety application. IEEE Intelligent Transportation Systems Magazine, 9(4), 62-75. https://doi.org/10.1109/MITS.2017.2743201

Tran, T. T. M., Parker, C., Wang, Y., & Tomitsch, M. (2022). Designing wearable augmented reality concepts to support scalability in autonomous vehicle–pedestrian interaction. Frontiers in Computer Science, 4, 866516. https://doi.org/10.3389/fcomp.2022.866516

Tran, T. T. M., Parker, C., Hoggenmüller, M., Hespanhol, L., & Tomitsch, M. (2023). Simulating wearable urban augmented reality experiences in VR: Lessons learnt from designing two future urban interfaces. Multimodal Technologies and Interaction, 7, 21. https://doi.org/10.3390/mti7020021

Troel-Madec, M., Boissieux, L., Borkoswki, S., Vaufreydaz, D., Alaimo, J., Chatagnon, S., & Spalanzani, A. (2019). eHMI positioning for autonomous vehicle/pedestrians interaction. In Adjunct Proceedings of the 31st Conference on l'Interaction Homme-Machine (pp. 1-8). https://doi.org/10.1145/3366551.3370340

Verma, H., Pythoud, G., Eden, G., Lalanne, D., & Evéquoz, F. (2019). Pedestrians and visual signs of intent: towards expressive autonomous passenger shuttles. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(3), 1-31. https://doi.org/10.1145/3351265

Wedel, M., Bigné, E., & Zhang, J. (2020). Virtual and augmented reality: Advancing research in consumer marketing. International Journal of Research in Marketing, 37(3), 443-465. https://doi.org/10.1016/j.ijresmar.2020.04.004

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). Engineering psychology and human performance. Routledge. World Health Organisation. (2020). The top 10 causes of death. https://www.who.int/news- room/fact-sheets/detail/the-top-10-causes-of-death

Yu, X., & Marinov, M. (2020). A study on recent developments and issues with obstacle detection systems for automated vehicles. Sustainability, 12(8), 3281. https://doi.org/10.3390/su12083281

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things journal, 1(1), 22-32. https://doi.org/10.1109/JIOT.2014.2306328

I

Part I: Design

2

Vulnerable road users and the coming wave of automated vehicles: Expert perspectives

This chapter was published as: *Tabone, W.*, De Winter, J. C. F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock. P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B.,& Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transportation Research Interdisciplinary Perspectives , 9, 100293

ABSTRACT

Automated driving research over the past decades has mostly focused on highway environments. Recent technological developments have drawn researchers and manufacturers to look ahead at introducing automated driving in cities. The current position paper examines this challenge from the viewpoint of scientific experts. Sixteen Human Factors researchers were interviewed about their personal perspectives on automated vehicles (AVs) and the interaction with VRUs in the future urban environment. Aspects such as smart infrastructure, external human-machine interfaces (eHMIs), and the potential of augmented reality (AR) were addressed during the interviews. The interviews showed that the researchers believed that fully autonomous vehicles will not be introduced in the coming decades and that intermediate levels of automation, specific AV services, or shared control will be used instead. The researchers foresaw a large role of smart infrastructure and expressed a need for AV-VRU segregation, but were concerned about corresponding costs and maintenance requirements. The majority indicated that eHMIs will enhance future AV-VRU interaction, but they noted that implicit communication will remain dominant and advised against text-based and instructive eHMIs. AR was commended for its potential in assisting VRUs, but given the technological challenges, its use, for the time being, was believed to be limited to scientific experiments. The present expert perspectives may be instrumental to various stakeholders and researchers concerned with the relationship between VRUs and AVs in future urban traffic.

2.1 GLOSSARY

- **Anthropomorphic eHMIs:** eHMIs that feature elements that borrow from human behaviour or appearance. Examples are artificial eyes in the area of the headlamps that can follow a crossing pedestrian (Chang et al., 2017), a car with a 'smiling' display on the bumper (Semcon, 2016), and a physically actuated hand attached to the roof of a vehicle that can signal a pedestrian to cross (Mahadevan et al., 2018).

- **Artificial General Intelligence (AGI):** hypothetical machine intelligence that is able to understand any task that a human can.

- **Augmented reality (AR):** AR allows the user to see the real world with virtual objects overlaid or embedded in it. AR supplements reality rather than replacing it (Azuma, 1997).

- **Automated Vehicle (AV):** a vehicle capable of driving itself but which requires human intervention at certain points. Automated vehicles are not to be confused with autonomous vehicles, which are vehicles capable of sensing their environment and moving safely without human input, also known as connected and autonomous vehicles (CAVs), driverless vehicles, robotic vehicles, or vehicles that exhibit SAE Level 5 automation (Taeihagh, & Lim, 2019).

- **Detached eHMIs:** eHMIs that are not attached to the vehicle, but which may be projected on the ground or visible elsewhere in the environment or on a wearable.

- **Explicit communication:** "behaviour signaling perception and/or movement without at the same time achieving either of these" (Schieben et al., 2019a). Examples are hand gestures, vocal communication, and eye contact (Schmidt, 2000).

- **Extended reality (XR):** is an umbrella term for immersive technologies, where 'X' can stand for various spatial computing technologies (e.g., VR, AR).

- **External human-machine interfaces (eHMIs):** communication devices located on the outside of the vehicle that can communicate to surrounding road users. An example is an electronic display on the front of the car (Colley, Häkkilä, Pfleging, & Alt, 2017; Schieben et al., 2019b).

- **Implicit communication:** "behaviour which is at the same time both achieving and signaling movement and/or perception" (Schieben et al., 2019a). Examples are vehicle speed and gap size (see also Bazilinskyy, Dodou, & De Winter, 2019; Dey & Terken, 2017; Schmidt, 2000).

- **SAE levels of automation:** classification of different levels of vehicle automation from no automation (Level 0) to full automation (Level 5). From Level 0 to 2, the driver has to be in control of the vehicle, and the automated driving system provides limited assistance. At Level 3 (conditionally automated driving), the driver is not anymore required to monitor the road permanently but has to take over control when the system requests so. At Level 4, the vehicle may operate itself without human intervention on certain types of roads, whereas at Level 5, it can drive itself anywhere and under all conditions (SAE International, 2018).

- **Segregated traffic:** refers to the separation of traffic streams, for example through the subdivision of towns and cities into certain units where road traffic is restricted, and pedestrians predominate (Mayhew, 2015).

- **Shared control:** a situation in which human and computer are performing the same task at the same time (Sheridan & Verplank, 1978). However, the term shared control is often used more loosely to describe a situation in which a human and computer are performing distinct aspects of a task (e.g., monitoring vs control) at the same time.

- **Smart infrastructure:** a traffic system which can monitor, measure, analyse, communicate, and act based on sensor captured information (adapted from Royal Academy of Engineering, 2012).

- **Vehicle-to-Everything (V2X) communication:** an umbrella term for a vehicle communication system where information from on-board sensors and other sources travels via high-bandwidth wireless links. V2X encompasses vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-pedestrian (V2P), and vehicle-to-network (V2N) communication. V2X may be part of future automated driving systems, where vehicles will be able to communicate with other vehicles, to pedestrians with wearables, data centres, and infrastructure such as signage, dynamic lane markers, cameras, parking metres, and street lighting.

- **Virtual reality (VR):** a computer-generated simulation of a realistic experience. VR usually blocks out the real world and replaces it with a virtual synthetic environment (Azuma, 1997).

- **Vulnerable Road User (VRU):** non-motorised road users, such as pedestrians and cyclists as well as motorcyclists and persons with disabilities or reduced mobility and orientation (Directive 2010/40).

- **Wearables:** devices that can be worn and which contain computer technology or can connect to the internet.

2.2 INTRODUCTION

AUTOMATED driving is a topic that has been discussed for over 80 years. Already in the first half of the 20th century, futuristic plans were created to roll out an automated highway system (AHS) in the United States (Geddes, 1940; Kröger, 2016). These futuristic visions have turned out to be inaccurate. Automated driving can better be described as an evolutionary process in which more and more computer systems have appeared in cars. Cruise control has been around since the 1950s, adaptive cruise control is available for more than 20 years (Bärgman & Victor, 2020; Bengler et al., 2014; Jurgen, 2006; Stanton, Young, & McCaulder, 1997), and there now is Level 2 automation (SAE International, 2018) where the driver can be intermittently hands-free, but not mind-free (Banks, Stanton, & Harvey, 2014; Dikmen, & Burns, 2016).

Over the past decade, there have been a large number of research projects that have focused on automated driving for relatively simple road environments such as highways. In particular, since about 2010, there has been a surge of research on highway automation involving SAE Level 3 automation, where the driver is free to perform a non-driving task, such as watching a movie. The research so far has covered topics such as human-machine interface design (as reviewed in Carsten & Martens, 2019), transitions of control (e.g., Eriksson & Stanton, 2017; Forster, Naujoks, Neukum, & Huestegge, 2017; Körber, Prasch, & Bengler, 2018; Zeeb, Buchner, & Schrauf, 2015), and driver state monitoring to track the driver's level of stress and visual attention (e.g., Cabrall, Janssen, & De Winter, 2018; Dong, Hu, Uchimura, & Murayama, 2011; Kim & Yang, 2017).

Research is now entering a new phase, with researchers and manufacturers examining automated driving technology in more complex environments such as cities. This direction concerns new research projects, such as the EU-funded project Supporting the Interaction of Humans and Automated Vehicles: Preparing for the Environment of Tomorrow (SHAPE-IT) (European Commission, 2019), which investigates this challenge from a Human Factors perspective. City automation raises various new questions, such as when AVs will be truly self-driving (SAE Levels 4 and 5) or whether such vehicles will have a form of shared control in which continuous input from the human operator is required.

With the advent of vehicle automation in cities, research topics will switch from drivers in AVs towards VRUs and their interactions with AVs (e.g., Applin, Riener, & Fischer, 2015; Hagenzieker, 2015). Foremost, it should be examined whether AVs will be beneficial to the safety and efficiency of VRUs. Moreover, the role of smart infrastructure in future traffic and whether such infrastructure will be used in the communication between AVs and VRUs must

be explored (Banks, Stanton, Burnett, & Hermawati, 2018; Sewalkar, & Seitz, 2019; Toh, Sanguesa, Cano, & Martinez, 2020). The question here is whether the road infrastructure should communicate with road users, and how the communication between AVs and VRUs should take place.

One of the research areas dealing with communication between AVs and VRUs concerns external human-machine interfaces (eHMIs). A large variety of eHMIs designs exist in academia (see Dey et al., 2020a; Rouchitsas & Alm, 2019) and industry (see Bazilinskyy et al., 2019). These come in different modalities, including LED strips and screens, robotic attachments, projections on the road, and auditory signals, amongst others. The question remains, however, whether AVs will communicate with VRUs via eHMIs or whether AVs will only use traditional signalling methods such as turn indicators and brake lights (see Norman, 2014). The importance of this topic may rise if the behaviour of AVs becomes indistinguishable from their manually driven counterparts (see Emuna, Borowsky, & Biess, 2020; Oliveira, Proctor, Burns, & Birrell, 2019; Stanton, Eriksson, Banks, & Hancock, 2020).

Anthropomorphism could be introduced in future traffic, and some researchers have already proposed human-like eHMIs (e.g., eyes on the car by Chang et al., 2017), to enhance user acceptance and safety. Anthropomorphic communication is popular in robotics (see Duffy, 2003; Fink, 2012), but it is still an open question whether anthropomorphism would be beneficial in eHMIs. Another fundamental topic which is still open for debate is whether communication between AVs and VRUs should be egocentric or allocentric in nature. The former entails that the AV instructs the VRU on their next action (e.g., “cross now”), which may be regarded as clear and unambiguous (Bazilinskyy et al., 2019), whereas the latter communicates the state of the AV and leaves the decision to the VRU, an approach taken by various research groups (e.g., Cefkin, Zhang, Stayton, & Vinkhuyzen, 2019).

There are other open questions such as whether eHMIs could or should be detached from the AV (as discussed in Eisma et al., 2020, and Mahadevan, Somanath, & Sharlin, 2018). More specifically, the question is whether receiving information that is separate from the AV would prove advantageous over eHMIs that are attached to the body of the AV. Furthermore, an important challenge is how eHMIs should operate in a traffic situation consisting of multiple road users. In some cases, the eHMI may want to send a message to just a single VRU (e.g., ‘I have seen you’), which could confuse other road users. These factors, together with VRU workload, cultural and language barriers, as well as the particular user requirements of older persons and children, are among the challenges to be explored.

A solution to these challenges could be the introduction of augmented reality (AR). The effectiveness of AR technology, a subset of extended-reality (XR), has already been explored and demonstrated in many domains, including driving (Schall et al., 2013) and navigation (Narzt et al., 2006), but also arts, education, entertainment, medicine, tourism, military applications, and marketing (e.g., Chang et al., 2014; Sanna, & Manuri, 2016; Tabone, 2020; Van Krevelen, & Poelman, 2010). Recent research has recommended the use of XR technology in AV-VRU communication and pedestrian simulation testing (Perez, Hasan, Shen, & Yang, 2019). Although it is still speculative whether AR will be used in future AV-VRU communication, the recognition of AR in other domains and the fact that AR technology is likely to improve in the coming years suggest that it is a worthwhile area of scientific endeavour: Could AR be a useful addition to daily life and VRU-AV

communication specifically?, what information should an AR device show to the VRU?, could AR offer a solution to the above-mentioned problem of multi-agent interactions, and how to address the issue that not all VRUs may have access to AR devices?

Extended Reality technology has already been employed in the testing of eHMIs through VR simulation and head-mounted displays (Bazilinskyy, Kooijman, Dodou, & De Winter, 2020; Deb, Carruth, Fuad, Stanley, & Frey, 2020), screen-based setups (Schwebel, Gaines, & Severson, 2008), mixed-reality setups (Maruhn, Dietrich, Prasch, & Schneider, 2020), and Cave Automatic Virtual Environment (CAVE) simulation (e.g., Kaleefathullah et al., in press) (see Feldstein, Lehsing, Dietrich, & Bengler, 2018; Schneider, & Bengler, 2020, for more examples). It is still, unclear, however, whether these methods are as valid as naturalistic testing.

Various research groups are currently studying the design of the future city, and dozens of concepts for VRU-AR interaction have been proposed so far, a situation which Dey et al. (2020a) recently characterised as an "eHMI jungle". Given the current state of the field, it seems worthwhile to perform a high-level survey among leading academics in the field. More specifically, the aim of this paper is to shed light on the above questions and uncertainties by soliciting the independent input of renowned Human Factors scientists who are experienced in this area. The idea is to assimilate the views of these researchers and generate an overview of topics of agreement and disagreement. This paper is organised as follows: firstly, the methods used in this study are described (Section 2), followed by 14 narratives that summarise the researchers' views (Section 3), and a discussion that reflects, using academic literature, on the convergences and divergences of the major points examined in the study (Section 4). Key takeaways and concluding remarks are presented at the end (Section 5).

2.3 METHODS

HUMAN FACTORS researchers were invited to share their views on the topic of AV-VRU communication in future traffic. The researchers were selected based on their publication record in the area of Human Factors of automated driving and based on their eminence as judged from the number of citations in their Google Scholar profile. Furthermore, preference was given to principal investigators in the SHAPE-IT project, because of its high relevance to the study objective. For diversity of views, it was ascertained that a maximum of one researcher was recruited from the same institution, with the exception of the TU Delft which featured two researchers from a different faculty, and with the exception of two researchers who themselves invited a colleague to join them in the same interview. The researchers were included as authors for their intellectual contribution.

In order to obtain input from the researchers in equivalent forms in terms of the topics covered, text length, and terminology used, it was decided to solicit the input from the researchers through semi-structured interviews, which were subsequently summarised by the first author, with the help of the second author, into a concise narrative.

The interview questions (see the Appendix), were split into four main themes: (1) general questions on AVs, (2) eHMIs, (3) AR and AR eHMIs, and (4) VR and AR for experiments. More specifically, the first section explored the researchers' views on the arrival of SAE Level 4 and 5 vehicles, shared control, the future of pedestrian-to-vehicle interaction, and smart road infrastructure. This was followed by a section that explored the researchers'

views on the usefulness of eHMIs, design considerations and future design directions such as eHMI detachment from the vehicle. The third section of the interview concerned the exploration of the viability of AR technology for eHMIs. Lastly, the interview concluded with questions related to the effectiveness of VR simulations for investigating pedestrians' behaviour around AVs and the potential use of AR technology in such experiments.

The researchers were interviewed via video conference and recorded by consent during the months of May and June 2020, with the interview durations varying between 45 to 120 minutes. Each video interview was transcribed and summarised into a narrative of approximately 700 words. The narratives reflect the researchers' statements during the interviews and do not include citations to academic references unless the researchers verbally referred to literature.

The narrative was sent to each researcher for approval or further adjustment from their end. Each researcher adjusted their contribution to clarify various points, respecting the thematic structure of the original contribution. The fourth theme of the interview (i.e., VR and AR for experiments) was omitted from the individual researcher narratives because the researchers provided similar comments on the positive and negative aspects. The major points made for the fourth part are included in the Discussion.

In summary, the aim was to create a position paper containing the current perspectives of independent researchers. Similar exercises have been conducted previously, in the form of multi-author guideline papers, working groups, expert consensus papers, and Delphi surveys, (e.g., Leiter et al., 2011; Lorenzon et al., 2018; Potapov et al., 2019; and see Kyriakidis et al., 2019, which explored the distinctive perspectives regarding Human Factors challenges in the development of AVs focusing on highway driving and the driver inside the AV).

2.4 RESEARCHERS' VIEWS

2.4.1 JONAS BÄRGMAN

Automated Vehicles: Because of all the complexities involved, including vehicle performance in different weather conditions, it will be a long time before we see SAE Level 5 vehicles on the roads, if ever. Level 4 vehicles will be available on motorways within five years or so. However, as they would require hand-over at, for example, off-ramps, many may still call it Level 3 automation. So, the SAE levels can be quite confusing. There will be some low-speed whole-trip Level 4 vehicles in cities in the same time frame. Pilots are already ongoing, but still on a relatively small scale. We will see fewer crashes with AVs in urban environments as all human problems related to perception and distraction will be avoided. However, there will likely be other types of crashes where the human would have been better at preventing the crash. There will also be new behavioural risks, where pedestrians will take the opportunity to cross the roads as they know that the AVs would stop. Therefore, AVs will not, at least not initially, be as mobile as manually driven vehicles.

Communication will play an important role in future traffic. However, it will take some time until communication equipment such as transponders in vehicles is pervasive. A complete shift to 5G communication on our mobile phones would greatly help in accomplishing this. Following the creation of standards, 5G would also be useful for vehicles to see beyond objects that obstruct camera and LIDAR vision.

eHMIs: eHMIs will make it easier to communicate the status of the vehicle to VRUs

and will allow building predictions on how AVs will perform. The communication of eHMIs should be language independent and focus on auditory, visual, and perhaps tactile signals. There must be standardisation, as it would be highly problematic to learn the eHMI for every car manufacturer. Standardisation would also help cross-cultural communication.

We will see context-aware AR in the distant future but not the kind where there are perfect overlays on the real-world image a user sees. There are too many components right now that still need improvement, and prices need to reduce. I see a lot of challenges for a communication-based AR system.

As an information provider, AR glasses can give the user auditory or textual feedback on objects in their sight, but great care needs to be taken not to overwhelm wearers with information. If at all, indications should only be given when pedestrians are about to commit an action, such as crossing. The interface should be simple and able to handle situations when multiples cars are approaching reliably. There should be no direct instruction to pedestrians about when they can cross. Rather, the pedestrian symbol should disappear, similar to showing only the red stoplight. In this way, the decision is left to the pedestrian, and only the AV's intent is communicated.

Augmented Reality: We cannot assume that pedestrians will be wearing AR glasses. AR design should be guided by the assumption that at least one of the pedestrians will not have it. Hence, we should still adhere to all the normal rules, and kinematic cues such as keeping speed, braking, or accelerating, will still be the main means for AVs to communicate their intentions. In other words, pedestrians may infer that an AV is automated through an eHMI and base their action on the vehicle's kinematic cues. Kinematic cues would be the most efficient and safe way for one-to-many or many-to-many AV-VRU communication scenarios. It is likely that eHMIs will not be needed anymore when AVs are ubiquitous, and humans have learned to interact with them. While AR information could be relied upon and helpful for pedestrians to navigate traffic situations, I do not think that we are going to see this being used in safety or time-critical situations. It must still be safe without them.

2.4.2 MARTIN BAUMANN

Automated Vehicles: There are currently SAE Level 5 people movers that drive at 15 km/h, but it will take 50 to 75 years for Levels 4 and 5 automation to occur on urban roads sharing the space with non-automated traffic participants. VRU-AV communication will mainly consist of AV behaviour, supported by basic eHMIs that communicate the intentions of the AV. I think in the future this can be complemented by wearables on pedestrians that would enable communication with the AVs.

Maybe we will see infrastructure such as lightning on the ground to support the interaction between pedestrians and AVs. Since smart infrastructure is expensive, the question is: who will pay for it? Smart infrastructure might therefore be introduced only in dedicated and selected shared spaces, which could become more efficient with the investment. Smart infrastructure will allow the AV to increase its perceptual horizon and see around corners. Wearables probably will enhance pedestrians' perception of smart infrastructure elements. I think this is something that will happen and will be probably accepted by traffic participants if the privacy issues are treated well.

eHMIs: I am positive that eHMIs will effectively support the interaction between AVs and pedestrians. However, it depends on their design and the information presented. The

current ways of communication and interaction between drivers and pedestrians will need to be transferred to AV-pedestrian communication. For example, in case of a deadlock, you would need to communicate explicitly, and so eHMIs would be essential. I think that eHMIs that communicate the intention of the vehicle are more effective than those that give advice to the interaction partner. AVs cannot be culture-independent, and there must be adaptations to the cultural context. For example, when I am standing at a zebra crossing, I would expect the car to stop for me in Germany, but this may not be the case in France or Italy. This means that I will have to communicate and interact in a way that depends on the cultural context. These differences in behaviour and habits must be respected by AVs.

Most of the communication should happen through legacy behaviours such as the movement of the AV. If this is not enough and there is still ambiguity, then simple explicit signals should factor in. Wearables might support this process, but if they cannot provide an unambiguous reference to a vehicle outside, then VRUs may not trust and use them. Receiving a message that a vehicle from the right will stop to let me cross without specifying which of the vehicles, does not seem very helpful. Thus, eHMI that are placed elsewhere than on the vehicle itself may not be effective and acceptable.

Augmented Reality: AR glasses could be of help here to solve this problem. They have the potential to make the situation more transparent if they do not overload the user with all possible information but present the relevant things only. Pervasive and context-aware AR could be available soon with a multitude of applications, if it is not already there. An example would be navigation advice through AR.

AR glasses will have to be socially acceptable. One of the problems of Google Glass was that people disliked the thought of being recorded. AR glasses offer the possibility to display more information than an AV-mounted eHMI could, and to identify the AV that is the source of communication and the interaction partner. The AR system should provide information related to safety. For example, the pedestrian could be presented with safety corridors related to which vehicles will stop for them. The advantage of this safety corridor concept is that it is integrating information from several vehicles. In doing so, it is clear to whom the vehicles are communicating, in contrast to the undirected communication of traditional eHMIs. Nevertheless, there will still be a need for eHMIs on the car or infrastructure as a backup since there will not be 100% market penetration for these glasses.

2.4.3 SHUCHISNIGDHA DEB

Automated Vehicles: SAE Level 3 is not safe because it splits the monitoring and control roles between the system and the driver, which confuses drivers. Therefore, SAE Level 4 should be introduced instead of SAE Level 3. Level 4 will be on the roads in 10 years as it is not safe yet. We must rebuild our infrastructure to support such vehicles. SAE Level 5 vehicles will be introduced in urban environments in the very distant future, following their introduction in dedicated lanes. For the time being, I foresee shared control and smart infrastructure solutions, with operator insides the vehicle and in control rooms.

Pedestrians are cautious when told that there will be AVs on the road, but this caution subsides quickly once correct behaviour is observed from the AVs. There is a risk that pedestrians will be jaywalking in front of AVs. Similarly, children may assume that the AV will stop for them, or they could just run out into the middle of the road. This would be very concerning. We cannot simply make AVs very conservative because it means that anybody

could obstruct them and create a traffic jam. Road users should be educated on AVs and their capabilities and limits.

eHMIs: Substantial efforts have been conducted to enhance AV-VRU interaction via eHMIs using LED lights on the vehicle. However, we have not seen such eHMIs on current SAE Level 2 vehicles, and it is unclear how it will work out in the future. I believe that eHMIs are essential on Level 4 and 5 AVs since pedestrians need a signal of some sort. People prefer AVs with eHMIs, and confusion arises when there is no such interface or driver present. If elements are borrowed from current pedestrian signals, we can develop an interface that could become a new standard that pedestrians can learn. Research shows that people usually understand eHMI signals, such as a stop sign or upraised hand, the way they should. However, anthropomorphic interfaces such as a smiley face are not useful as they may elicit surprise and curiosity, which can be counterproductive. Especially children may huddle around the vehicle to explore the eHMI. Text always works well, but it depends on literacy, and it is problematic in cross-country scenarios.

I think it is reasonable to trust eHMIs that are detached from the AV. If we want to connect everything, we will have a mobile application that will give users a signal based on acoustics or other signals like vehicle speed. We are currently exploring whether providing information to cyclists on mobile phones would be overwhelming or useful. AR glasses will probably be useful for cyclists in the future. I am not sure how useful these would be for pedestrians since they move slowly and hence have enough time to decide.

Augmented Reality: Pervasive and context-aware AR will be of great benefit to humans, especially for the training of workforce, where it is already being used. For example, in the USA, AR is already being used in the Navy and for construction safety, and agriculture. However, we need to do a lot of research before we put such devices to people's eyes. AR should not overwhelm the user. Information should include highlighting of hazards such as specific alerts that a vehicle is approaching. While this can be a good solution for vulnerable road users like pedestrians, it can also be overwhelming. For situations where not everyone has an AR device, I would like to see information from the smart infrastructure. If there is no such infrastructure, the signal should come from the car itself. There could also be a hybrid approach depending on the location.

AR could be used to alleviate ambiguous situations where an AV is communicating to multiple pedestrians. Such a situation is confusing and may lead to accidents, but it is still better than not having a signal at all. The solution here is once again standardisation and training.

2.4.4 AZRA HABIBOVIC

Automated Vehicles: SAE Level 5 vehicles will be driving in urban environments only in the very distant future. For Level 4, it would depend on what is meant by that level. We can have Level 4 in different operational design domains. We will probably first have Level 4 automation in specific operational design domains, such as highways. Vehicles with this potential may be available by 2023–2024, while Level 4 vehicles on a wider scale may surface in 10–15 years. Vulnerable road user safety will be improved if AVs are designed properly, but I think there are still many uncertainties around AVs in general. I do not think that automated driving would eliminate all possible accidents and incidents in traffic, but hopefully, it will make the situation better.

Smart infrastructure will be part of future transport. Although AVs should be able to operate by using the information provided through on-board sensors only, their operation could be improved with the help of digital infrastructure and other vehicles. As it looks right now, information exchange will not necessarily occur using V2V but rather through different cloud systems. However, making use of safety-critical information from other sources is currently challenging.

eHMIs: It is uncertain if eHMIs would be needed or implemented. eHMI might be necessary when automated driving is not very mature and not accepted in society yet. Another point which is not studied enough is how eHMIs affect traffic flow and efficiency. Also, long-term behavioural effects of eHMIs are largely unknown. It will take many years to be able to decide on the perfect eHMI modality since we would need a large enough fleet of AVs to test this. eHMIs must be based on different modalities and not be based exclusively on light or text. Text-based eHMIs are challenging because they would need to be translated into many different languages, whereas anthropomorphic eHMIs pose a challenge of cost and durability. The vehicle should communicate its intent and status without instructing people as the latter could be the cause of legal incidents. You can never be sure that neighbouring vehicles will also stop for the pedestrian. One should also consider eHMIs in terms of vehicle motion and vehicle appearance, as many important signals are today communicated to other users via such implicit means.

Requirements of people with different needs in society, such as children and people with visual or auditory impairments, must be addressed. The prototypes that exist now and are used in research studies are usually based on one modality, and if they are evaluated, they are not evaluated with children and older persons. We need to have a much wider approach to research and development. Education should ensure that children learn the meaning of an eHMI just as they learn the meaning of traffic lights. Another approach to make eHMIs suitable to children would be using multimodality and existing colour conventions, such as green for 'go' and red for 'stop'. However, with current regulations, this might be difficult as certain colours are already in use in traffic, and we are therefore limited in what colours to use in eHMIs.

Augmented Reality: Augmented reality has potential, but for some reason, it has not reached a breakout yet. Pervasive and context-aware AR has been under development for many years, but my impression is that AR devices are not moving forward at high speed. It could mean that the technology has not reached an acceptable level of maturity or that we have not found a good application yet. But this could change drastically, as happened with the smartphone.

AR has the potential to amplify information that is already available but is unseen. Accordingly, AR has the potential to be useful for children and people with impairments. I do not think that the information in the AR environment should be any different compared to communication by the AV itself. It should be directed and minimalistic in nature as we do not want to overload the user. In any case, there must be a combination of communication modalities, as not all VRUs will be using AR. AR would be interesting to explore for cases when the AV has to communicate to individual pedestrians.

2.4.5 MARJAN HAGENZIEKER

Automated Vehicles: SAE Level 4 AV shuttles have already been implemented on urban roads on dedicated lanes and sometimes mixed with cyclists and pedestrians on short stretches of road. It can be debated whether they are really Level 4 because there are stewards on-board. The steward needs to intervene often because of technological failure or due to objects on the road. Level 4 and 5 passenger cars are decades away, if they will ever even materialise for large scale use. It will be particularly challenging to get them implemented on a large scale on urban roads. Since SAE Level 5 is difficult to achieve, I believe that we will probably have shared control for a long time. I am not too optimistic about pedestrian safety as I am afraid that technology is promising more than it can deliver. Although we see that advanced sensors that can detect pedestrians are being developed, it is unclear whether they can predict what pedestrians will do.

It is difficult to say which role smart infrastructure will have in the future because there will be a need for standards and numerous authorities need to collaborate. The European view is that there will be such infrastructure in the future whereas, in other parts of the world and certain industries, there are wishes of independence, where vehicles communicate with each other without the need of specific communication with the infrastructure.

eHMIs: It is not clear yet to what extent eHMIs will be helpful. It is still a new area of research, and most research has been done for a single person interacting with a single vehicle. The research so far shows that it is difficult to convey a message that is understood by everyone in the same way. Road users base their decisions not necessarily on the eHMI but rather on AV speed and distance. The need for communication with pedestrians will decrease as pedestrians build up experience with AVs. The most effective eHMIs appear to be those that communicate the vehicle's intentions not only to pedestrians but to all traffic around it.

I am curious as to whether anthropomorphic eHMIs will be effective. We should look at other fields such as health and robotics, to see how they use anthropomorphic ideas. Moreover, the issue of different cultures needs to be solved. Text messages work best when the receiver knows the language. We have the same problem in traditional traffic, where foreigners do not always understand road signs. There will probably be a multitude of solutions just as has happened with route guidance. It is good to have standards. In this way, we can build up expectations and schemas, and accordingly make fewer errors and become safer. However, we are so much at the beginning; it is too early to work on standardisation. For the time being, we should be open-minded and do out-of-the-box research.

Augmented Reality: Pervasive and context-aware AR will become available soon. Some solutions are already available for specific groups of people, but we are a long way from having it on a larger scale. The information that is useful in AR is not any different from what pedestrians currently need. That is, it would be good to know if there are obstacles on the road or moving objects that the pedestrian should be aware of. The AR system could communicate what behaviour is allowed, similar to how road users currently gather such information through visual and auditory feedback of the road environment. However, one must consider which tasks should stay with humans because humans may be better at certain tasks or can do them naturally without external aids.

AR could be a solution to language barriers: If you know the user, you can implement the right language. Moreover, AR can be helpful in AV-VRU communication, but it is still

something that has to be investigated. Maybe one could stimulate road users to use AR equipment even if they do not like it. However, if it is pleasurable to use, they may gain a lot more from it. AR technology will be rejected, or it will be adopted out of free will because it serves lots of purposes, just like the smartphone. As with any automation, technology is one thing and what humans want is another. Sometimes developments go very fast, and sometimes it becomes disastrous and dies out.

2.4.6 P. A. HANCOCK

Automated Vehicles: AVs will be restricted to freeways in the coming years and then move on to arterial roads and urban areas within ten years. We will asymptote towards SAE Level 5 as automated control penetrates all forms of transportation, including ocean transport such as container ships. The latter will occur silently compared to ground transportation, and once manufacturers realise this, they will progress even more rapidly towards the development of SAE Level 5 AVs. For such vehicles to be deemed safe by pedestrians, we need to show empirically that the AV does no greater harm than the manned vehicle. AVs will first be used in niche areas, where they can best be employed for making a profit.

Pedestrians are highly vulnerable, and high-density pedestrian areas are not a good place for AVs to operate, as a conservative algorithm would drastically slow down the vehicle's movement. Since the intelligence may not necessarily have to be on the AV itself, such pedestrian dense environments can especially benefit from smart infrastructure. However, this presents both political and technical challenges, which include making the intelligent roadway work for non-AVs and the limitation of sensor capacities in different geographical locations. A further issue is the funding model. Infrastructure, since it is a public good, would presumably need a public-private transformation for improvement.

eHMIs: There may be value in using eHMIs for informing pedestrians about the AV's intentions. It would be ideal to provide a unified affordance of the intentions of the vehicle, rather than just a visual colouring. For example, it would be interesting to alter the AV's perceived surface to make it appear as more or less threatening, contingent on need. The message communicated should attempt to resolve as much ambiguity with the shortest possible signal. With regard to AR-based HMIs, text would be too slow and constrained, and therefore, there must be more focus on graphical, aural, and tactile message representations. The focus would be more on collision avoidance than on information transmission per se.

It is unknown how an eHMI should communicate to multiple individual pedestrians; this would also raise the question of liability in case of a collision. The AV's algorithm would need to demonstrate that it did not explicitly discriminate one pedestrian over another. In terms of the Haddon hierarchy (Haddon, 1970), the easiest solution is to keep pedestrians and AVs separated as much as possible and have only limited locations where interaction can occur. If interaction were to occur with detached eHMIs or wearables, it should not impose the cognitive load that is freed from the AV driver onto the pedestrian, as this could lead to overload very quickly. The idea of having these capabilities is not to aid you while you are walking but to permit you to do something else while you are walking.

Augmented Reality: AR will be helpful in VRU-AV communication. Humans prefer accessing information anywhere and anytime as required. This is being strongly demonstrated during the present COVID-19 lockdown period. There is much potential for future use too, such as the integration into spacesuits for use on Mars missions. Even if most people wear

AR glasses, some people will still be excluded. Also, there are several Human Factors and Ergonomics issues that need to be solved, including for how long a user could wear AR glasses. Ideally, the user should not know that they are wearing it since the whole point about interfaces is that they disappear. If you push on the idea of Gibsonian affordances (Gibson, 1979), the affordances are not a conscious experience; affordances are an implicit process of the environment around the perceiver. Another important issue is what would happen if you remove or leave the AR off. In this case, one may expect negative transfer, and a possible solution would be to invest in the infrastructure.

2.4.7 RIENDER HAPPEE

Automated Vehicles: Automation in passenger cars will gradually progress from current SAE Level 2 to Levels 3 and 4, with an operational design domain limited to highways and other roads where pedestrians and cyclists are not to be expected. The perception capabilities of these vehicles will enhance safety on all types of roads through increasingly effective intervention systems, such as automated emergency braking. Level 5 is decades away, but driverless Level 4 shuttles already operate on public roads with a safety steward on board, and several sites aim to drive without steward in the near future. The presence of a steward is reassuring for occupants but often not visible or known to other road users. Hence, driverless shuttles provide a suitable basis for investigating and improving AV interaction with pedestrians and cyclists. Surveys show that pedestrians and cyclists encountering driverless shuttles are willing to accept such AVs on public roads. However, they also want AVs to express intentions through visual and auditory interfaces.

For milestone improvements in urban road safety and AV acceptance, a bigger change is needed. I believe that AVs and vulnerable road users such as pedestrians and cyclists should not encounter vehicles driving at high speeds and violating traffic rules. This can be achieved through traffic separation, banning manually driven cars and with speed limits that are enforced through communication systems. Pedestrians and cyclists must also do their bit to improve safety, by changing their behaviour. AVs will aim to prevent all accidents, even if they result from other road users' misbehaviour. Therefore, misbehaviours may induce deadlock in dense pedestrian areas. Cities may have to be redesigned to contain well-designed static infrastructure and communication systems. Traffic lights may be replaced by or connected to communication systems in cars and systems worn by pedestrians and cyclists.

eHMIs: Similar to smart infrastructure, eHMIs can contribute to safety and AV acceptance. Our research indicates eHMIs are especially useful at low speeds where pedestrians have time to interpret and react to eHMI signals. At longer distances, recognition of eHMIs is problematic. Our experiments showed surprisingly small differences between fundamentally different types of eHMIs in terms of acceptance and effect on behaviour, and participants learned to use eHMIs quickly. Possibly, our participants simply reacted to the changing eHMI colour, text, or symbol. This eHMI change was always coupled with implicit communication, which remains an important factor.

Overtrust in eHMIs may cause accidents. This ties in with the issue of directing the message to the appropriate actor: pedestrians may see a message that was not intended for them and cross erroneously. One solution is that AVs communicate status rather than instruct. After further research in a preferably worldwide population, we must harmonise eHMIs,

similar to current traffic signs and vehicle lights.

Augmented Reality: Regarding signals received wirelessly, the smartphone is too unreliable. AR is promising, but for prolonged use, comfort should be improved in terms of resolution, image stability, and smoothness in order to prevent eyestrain or even motion sickness. AR will enable personalised information to VRUs early during an AV interaction. AR would be ideal for conveying if a vehicle is automated, whether an AV has seen the VRU, and which action the AV will take. An interesting concept would be an augmented 3D traffic light in the form of a virtual fence to stop pedestrians from crossing a vehicle lane. It would be ideal for tram lanes as well, and simple enough for a child to understand even in ambiguous situations when multiple pedestrians are present.

Safe and acceptable AV interaction may be feasible without eHMIs or AR, but I do expect substantial benefits of such systems. To compensate for not everyone having access to the technology, AVs should be designed to be understandable even for those who do not have the technology or have other limitations.

2.4.8 JOSEF KREMS AND CLAUDIA ACKERMANN

Automated Vehicles: Level 4 automation will be available on urban roads in 10 to 20 years, while Level 5 automation may appear 5 years later. However, some people will still like to have the pleasure to drive by themselves. These people may have to be forced to use automation by law to reduce accidents. AVs' communication to pedestrians should be through implicit behaviour from the vehicle, while eHMIs should be used in ambiguous situations.

Another option is to use the infrastructure to communicate to the pedestrian, so the car does not have to be used as a communication entity. Smart infrastructure will play a pivotal role in future electromobility. Silent cars should not be equipped with additional noise as it would counteract the idea of silent cars. A better solution is to use infrastructure to warn people via devices such as smartphones. Infrastructure will also play a key role in traffic separation. Segregating different modalities such as AVs from manually driven cars, cyclists, and pedestrians is very costly, however. There might be a lot of warning signals in the future, but there is a limit to the number of signals that humans can handle. At the same time, information processing capabilities will increase as people will come better at understanding how the connected world works.

eHMIs: eHMIs should be avoided unless we can show that their benefits are large. It is a tricky thing to have new signals on the road as we know that signals tend to add workload and confusion. eHMIs should make crossing easier rather than more difficult. There are still numerous questions related to how the right message should be communicated to the right person. There will not be a need for eHMIs for every single interaction for every pedestrian. There are a lot of technological, regulatory, and standardisation challenges such as regarding projections on the road at daylight, the limited colour options for eHMIs, designing in the appropriate size to make the interfaces visible from 30–40 m ahead, and making sure that eHMIs work across different crossing cultures and language barriers. This was one of the main benefits of the UN Vienna Convention (United Nations, 1968). Unfortunately, for the time being, all OEMs are creating their own distinct designs.

Driver gestures do not play a significant role except for very short distances; a pedestrian's decision to cross is usually taken before that. Therefore, there will not be much of a difference

between an AV or a manually driven car approaching from a certain distance, and so the same signals and interaction principles apply. Kinematic cues such as vehicle deceleration would still signify that the vehicle has acknowledged the pedestrian and intends to stop. Hence, an eHMI will only be useful when the pedestrian is still unsure whether the car has acknowledged them.

Visual eHMIs are effective, but there is no clear answer as to whether to use symbols or text-based messages. Another issue is that current eHMIs are very artificial and have no cultural backup. They would require a lot of training or time for people to get used to them. While there are eHMI designs that utilise anthropomorphic elements to enhance understandability, future AVs will allow a whole new design for the AV and eHMIs. The inside and outside of the AV will probably not be humanlike but designed so that pedestrians can recognise the AVs and adapt their behaviour. Our research has shown that people want to have instructions from the car about what to do. Participants found information about the status of the AV to be ambiguous and less trustworthy.

Augmented Reality: AR-based eHMIs will not be much different from regular eHMIs. Most people do not like additional tools and will have a problem wearing AR glasses. However, acceptance would be higher if it is something that does not require continuous interaction. It would be important to support situation awareness. We need to find out what kind of information has to be made artificial or augmented. Context information such as regarding the dynamics from approaching vehicles could prove useful. For cases where not everyone has access to AR, intelligent infrastructure such as a flashing zebra crossing or a special traffic light that addresses pedestrians, could be used.

Although AR already exists in research environments and at OEMs, it will not have a role in the market of the near future due to its cost. For the time being, this technology will be used for some special applications and users such as firefighters, police, or maybe rescue personnel. A designed-for-all approach will take a long time.

2.4.9 JOHN D. LEE

Automated Vehicles: SAE Level 5, by definition, will never happen. It is defined as being able to drive under any condition, which is impossible even for manually driven cars. There will always be limitations, such as during a snowstorm. At the same time, Level 4 shuttles that have no driver, steering wheel, or brake pedals already exist in urban environments. As these vehicles become more advanced in the next ten years, they might be perceived as Level 5 vehicles because their passengers feel that they can go anywhere. For the time being, shared control is viable, because people can pair with automation to accommodate unanticipated variability of the driving environment. Autonomous taxis that serve multiple people are really the promise of vehicle automation. But this promise cannot be achieved through shared control. Therefore, to get to the promise of AVs there needs to be Level 4 automation.

Interaction with pedestrians is going to be challenging in several ways. Firstly, the unpredictability of pedestrians makes it almost impossible for people or algorithms to avoid collisions. Secondly, pedestrians interact in a social and culture-specific manner, which makes it challenging to create algorithms with the requisite culturally-specific driving expertise. Lastly, pedestrians may perceive the risk of AVs differently compared to manual vehicles. Pedestrians may reject AVs since pedestrians have no incentive to accept them into

their space.

I am not optimistic about smart road infrastructure as this presents a challenge of cost and backward compatibility. A virtual traffic signal might work well for properly equipped vehicles but will be invisible to those without. So, it would be difficult to get people to invest in this. I think the most challenging part would be how to communicate with pedestrians. If it is virtual, then that mandates equipment on the pedestrian to signal to the infrastructure and display the received signals. This may be possible in a country with a high standard of living where everybody can be outfitted with smart glasses. But in other countries, you have economic disparities, which will leave large parts of the population wandering about without instrumentation and displays. Non-smart infrastructure might be a more productive way forward. Best practices of infrastructure design that currently helps drivers and pedestrians negotiate the roadways might also help pedestrians and AVs interact safely.

eHMIs: eHMIs will almost certainly be implemented in the future. They will be useful in building trust among pedestrians whilst giving them a feeling that the AV is polite. However, eHMI design should be considered as a secondary communication channel. eHMIs should be paired with vehicle motion cues since these can communicate the intention of the automation more effectively than say a text-based interface because people have evolved to communicate through motion. The motion of the vehicle can appear menacing or safe, depending on the deceleration profile. This approach would also be helpful to make such communication cross-cultural, since motion cues are language independent. Motion also addresses a deeper anthropomorphic level based on a fundamental perception of motion and the meaning of that motion. On the other hand, the surface anthropomorphic level is based on elements such as putting eyes on the vehicle. I am uncertain how effective that would be but, work in robotics suggest that it might work as people seem to be sensitive to eye gaze. eHMIs that instruct the pedestrian can be risky as people may over-rely on the automation. Instead, it should inform the pedestrian of its intent, which would hopefully prime the pedestrian to check for other cues in the environment. A final important aspect for eHMI development is to include children, older people, and other vulnerable populations as a priority test cases because what is explainable to them in a few words would hopefully be understandable to the rest of the population.

With regards to detached eHMIs, I think that physical proximity is important, or at least the illusion of it. With detached eHMIs, there may be an additional burden of mental rotation and mapping of the image to the object it refers to. More generally, detached eHMIs need to consider the frame of reference of the information that the person must use to interpret the information. The motion of the vehicle itself is likely the most direct and interpretable cue.

Augmented Reality: AR smart glass technology is here, but its penetration is still low. It is a technology that may be on the cusp of widespread use, but whether people will find it useful and adopt it broadly is the question. This technology will offer both a benefit and disbenefit to people, similar to the smartphone, but it will have even greater power to attract and guide attention.

AR could prove beneficial for both pedestrians and drivers. It could help drivers understand what the automation is doing and why they might have to take back control. On the pedestrian side, AR will enrich the communication between the pedestrian and the driver. Pedestrians could see the vehicle's intent more directly. Another thing that AR can help with is giving directions and public transit coordination, where information is overlaid on the

world. However, relying on AR as a safety feature for interacting with AVs is problematic. AR provides more flexibility and opportunity for personalization than physical reality for the design of eHMIs, but personalisation may not be a good idea since users oftentimes do not understand what the best solution is.

Most importantly, AR facilitates communication but only for those who have it and are wearing it, and we should design for those who are not. Consequently, it should be considered a secondary information source that complements non-AR sources of information. Working out the necessary communication in the absence of AR seems important to me. I am inclined to say that the vehicle cues should be sufficient on its own to ensure safety. In other words, the non-augmented layer would have to be sufficient for those who are not seeing the augmented layer.

2.4.10 MARIEKE MARTENS

Automated Vehicles: There is a lot of confusion about what the levels actually mean. When thinking this through, there is, for instance, a lot of confusion about Level 4, with a huge difference between SAE Level 4 public transport and SAE Level 4 personal cars. For public transport or robot taxis, you can train a vehicle to drive on specific routes. For the general public, this will be interpreted as fully automated or autonomous driving. However, for personal vehicles, Level 4 means the car can drive itself for a specific amount of time, offering the driver the chance to do something else and requesting the driver to take back control in various conditions. The only difference in this case between Level 3 and Level 4 is that, if the person does not take back control, there is a safety-backup, without specification of how safe this option actually is. Level 5 refers to fully automated or autonomous driving, under all conditions. I do not understand this rush for Level 5; the most important thing is how we can improve traffic safety and avoid confusion for other road users.

eHMIs: The concept of the eHMI has been introduced to support the interaction with other road users and prevent confusion about what the vehicle is doing when in automated mode. The eHMI may need to show that a dual-mode vehicle is in automated mode. Depending on the surroundings, other signs and signals may be needed. One must distinguish between what is absolutely necessary to communicate–to improve traffic safety or at least not cause accidents–and what is not absolutely required but may still be wise, for instance to improve acceptance. Some signals may be the same as manually driven vehicles, such as turn signals and stop lights in the rear of the vehicle. Secondly, there is the question of liability, which also guides the way manufacturers will present messages to pedestrians. For example, it is unlikely that an AV will tell the pedestrian what to do. Even though people often indicate that they want to know if the AV has detected them, I believe that we should not do this because it could give rise to misunderstandings and unsafe situations. It makes more sense to indicate what the AVs plans to do, the so-called communication of intent. When there are multiple AVs, pedestrians, cyclists, and other road users, it will be difficult to indicate who the message is for. An issue is that eHMIs may distract VRUs from looking at non-AVs, which is undesirable (as also pointed out by International Organization for Standardization [ISO], 2018).

Augmented Reality: Smartphones or wearables could be used to predict the influx of pedestrians such as at schools or at the end of big events, and reroute AVs accordingly. However, I do not believe that the solution is that everybody should be connected and that

VRUs are warned for AVs on their phones.

I do not believe in smiling cars, strange designs, text, or voice messages. eHMI messages should be simple and always in relation to the vehicle's behaviour as vehicle motion is really the strongest cue. The largest benefits of eHMIs are in shorter-distance communication, since when a vehicle is further away, people pay attention to other cues such as speed, movement, and distance. It would be interesting to be notified not just about the AV's intention to stop but also about where it intends to stop; this is what we found in research.

With regard to eHMIs not being on the car, I think this is too risky since different cars may project different things and view may be blocked or changed by the presence of others. And let's not forget that the movements of the car are the primary means for conveying vehicle intent.

2.4.11 NATASHA MERAT

Automated Vehicles: The deployment of SAE Level 5 vehicles in an urban environment will probably never happen, whereas SAE Level 4 vehicles may be deployed around 2040. Until that time, it would be a good idea to keep the human in the loop, and shared control may be one of the ways to accomplish this.

In the InterACT project (interACT, 2017), we have observed that, while technological developments have made AVs great at obstacle detection, they still have limitations when it comes to seeing around other vehicles and anticipating the future movements of other road users such as cyclists and pedestrians. We must go a step further than object detection and have the AV communicate to different traffic actors so that there is a decent flow of traffic. The general feeling is that people want the AV to communicate with them, as was made very clear in the CityMobil2 project.

eHMIs: We are currently looking at whether eHMIs can replace interactions by drivers. It has become clear that the presence of an eHMI generally translates to quicker pedestrian crossing decisions. Although there is no set formula on how to design these interfaces, we do not think that text is useful, as it is not very international or decipherable at a distance. We are, therefore, testing lights, but there are still uncertainties regarding the choice of colour and presentation methods, and of course, this is still an issue for visually impaired road users. Research on anthropomorphic messaging has suggested that personalised messaging, which, for example, makes use of a family member's voice, can be effective. It would be interesting to investigate whether similar results could be achieved for an AV-based eHMI.

Pedestrians report that knowing that the AV has detected them is important. However, a problem that remains is how to communicate between an AV and multiple actors simultaneously. So far, most studies have been conducted between only two actors: an AV and a pedestrian. The use of infrastructure and wearables have been mentioned as possible solutions. For smart infrastructure to be part of the equation, there must be much more investment in reliable communication technology.

Augmented Reality: Pervasive and context-aware AR is already being developed and implemented inside the car. So why not outside of the car? AR would help us see things that we cannot see and, for research, investigate traffic situations in which human presence would not be safe. There is a role for psychologists in the design of AR systems; the design should ensure that users direct their attention to the right information at the right time. The use of AR glasses can be powerful as it can allow communication with different people at the same

time. Accordingly, we can move forward from the present one-to-many communication, which presently is a problematic situation.

Most eHMI research has been conducted in the Western world, with only a few exceptions. Introducing eHMIs everywhere would be challenging. AR technology may contribute to solving the issue of communicating to multiple cultures as a different interface would be used for each, similar to having different voices for your satnav. In a way, AR will also allow pedestrians to walk around while consuming media and be prompted when a vehicle communicates. This is a similar analogy to the AV, allowing the driver to do other tasks. In a way, AR will free the pedestrian, just like the AV frees up the driver.

2.4.12 DON NORMAN AND COLLEEN EMMENEGGER

Automated Vehicles: We do not like the levels. They mislead because they use the wrong dimensions to characterise the complex and subtle distinctions that are necessary. Note that fully autonomous vehicles already exist in mining, agriculture, and factory floors, and there are attempts to make home deliveries autonomous. It will take a very long time until commercial versions of these vehicles are released onto urban roadways without changing the infrastructure and without separating the humans from the AVs.

As regards to shared control, we must first look at its definition. It means that both the operator and the machine are performing a task together. What this often means is that a person is supposed to sit and do nothing for hours, but is expected to respond in a tenth of a second in cases of an emergency. That is not sharing but monitoring, and it is an impossible situation. We have long argued that excellent but partial autonomy should be skipped in favour of full autonomy. Complete autonomy in any situation, especially off-road in difficult terrain will not happen for a very long time, if ever.

There are ways of dividing up the job so that the autonomy does what it is good at and people can do what they are good at, both doing it together, much as a rider and a horse coordinate their activities. Riding a horse is a shared activity, where the horse does all the low-level detailed stuff while the rider controls the goals and the pace. But where necessary, the rider can force the horse to do things it would otherwise not do, and if the rider falls asleep or is otherwise incapable of supervision, the horse can take over, either returning to its home base or simply stopping at a safe location, waiting for assistance.

It is the AV's responsibility to make its intentions clear to all other traffic participants. However, signalling the intention does not mean that everyone who needs to know it receives or understands the message. The AV's intention could be conveyed in many ways, but requiring everyone to carry a special device to read that intention is not a viable solution.

The safest approach is to separate the means of travel, removing the priority over all other users of a shared infrastructure that has been given to the automobile. Safety requires separation of the many different modes of transportation. Fundamentally, the problem is that mixed modes of almost anything are dangerous.

It is incredible how well the current traffic system works. Our research has indicated that pedestrians and other road users pay attention to the movements of cars. A car that stops is signalling to others. Similarly, a car that travels very slowly, far slower than is permitted, is also signalling. Where a car stops at a crosswalk signals whether the car will wait for pedestrians and other users or whether it is anxious to get moving, which means other road users should be cautious. Any form of explicit communication should be built on top of this

baseline. The vehicle should communicate its intention so that other road users can chart their best course of action.

eHMIs: The approach of having the car signal to other road users that it is safe for them to proceed is very dangerous. Drivers do it today with hand signals, sometimes leading to accidents. This type of signalling only works when there are so few road users that the intended recipient is unambiguous and there are no other possible vehicles that might suddenly appear, negating the signal given by the first vehicle.

Our studies show that vehicles should convey intentions and not tell others how to behave. Movement is one way of doing this. The nuances of gestures through vehicle motion cues may be different from one country to the next, but the notion of using motion to communicate is still a valuable one. Using motion would be ideal in a multicultural setting. If messages are placed outside of the car, human attention becomes more scattered, especially when there are many vehicles, so they may miss a critical signal. The use of sound may help because hearing picks up sounds from all directions. However, if many vehicles are simultaneously signalling through sound, the result might be confusion and chaos. These problems need to be explored deeply. And here is where standardisation is essential.

Augmented Reality: Developing AR-based eHMIs is a wrong perspective as it puts the responsibility on the pedestrian to act on the presented information. Instead, it should be a pedestrian-centred approach where the signalling comes from the pedestrian and the AV will drive safely, rather than the other way around. Furthermore, we do not know if people will adapt to AR devices. A safety concern is about what will happen when the technology does not operate properly. AR glasses are effective and widely used in specialised activities, but they may not be appropriate for solving traffic problems.

Concerning the problem of an AV communicating to multiple pedestrians, we must remember that there is a lot of following behaviour in traffic. A pedestrian decides to cross not only by watching the traffic but also by observing what other pedestrians are doing. All vehicles will stop if a group of pedestrians cross the road. This happens because the mob of people have signalled its intention to the automobiles. Such communication of intentions goes both ways, and the car may therefore not have to signal to multiple pedestrians.

2.4.13 THOMAS B. SHERIDAN

Automated Vehicles: There are already AVs out there, but it does not mean that people will accept them. Furthermore, there is much to driving that is social. Pedestrians can interpret the face of a driver and their hand signals, which is a difficult task for artificial intelligence to pick up. I believe that, with driving, you need to be in the loop, alert, and attentive. Some of my previous work stated that one should not expect a driver who is not attentive to come back in the loop after an instant take-over signal. It takes maybe 10 to 20 seconds for attention to be restored. Therefore, you must jump over that Level (SAE 3) and be smarter. I am a bit sceptical that automated driving will be accepted quickly since it all boils down to trust. In the area of Human Factors, trust of automation has become very important. I believe that the driver should be in the loop with at least some of the functions of an AV, so either the driver is doing all or most of the tasks, or the computer is doing all or most of the tasks. So, there could be a scenario where the driver handles steering, and the computer handles braking. This is known as traded control and not what people are referring to as shared control.

eHMIs: Smart road infrastructure will have many potential uses, but the problem is

that it is expensive, requires maintenance, and is vulnerable to damage by weather. I am worried about the safety and reliability of that infrastructure. On another note, since I am retired, I have not been familiar with the eHMI concept. However, I find it an interesting and wonderful idea. I believe that they would make a difference if they are a substitute for the driver's communication. If the AV is confused and not sure about itself or another actor, then communication, similar to how drivers do with honking or hand movement, would be important to pedestrians. It would be ideal for communicating what is expected of the pedestrian or the driver. Using anthropomorphic communication would be a good thing. In fact, anthropomorphism is common across cultures, and hence it could be instrumental in aiding in the breaking down of cultural barriers for communication with eHMIs.

Augmented Reality: Pervasive and context-aware AR is possible of course and is already being used in various settings such as by mechanics, assembly line workers, and storage facilities, where the glasses help the user navigate towards a package with the queried ID. AR-based eHMIs would be most useful in low-speed scenarios where the vehicle and/or the pedestrian have stopped or are moving very slowly and there is uncertainty about who would go first. This is usually sorted out socially between two people, and in this case, eHMIs would be instrumental in replicating such interaction. AI-based communication through AR could be done even when there is no driver in the car.

For cases when a pedestrian is not wearing the glasses, there would be other cues in addition to the AR eHMIs. The vehicle could honk its horn, edge, and also move towards the centreline. This would also be useful in cases where the AV has to communicate with multiple pedestrians. In a manually driven scenario, the vehicle is slowly edged ahead at a crossing, and the driver would check if any pedestrian is moving or not. In case of no movement, the vehicle will be edged beyond and continue driving. Such communication could be replicated by an AV. AR eHMIs would also be useful when travelling to other countries with unfamiliar cultures to the traveller. It would be interesting to have a future scenario where hotels would rent out AR glasses that would help me cross the streets of that particular place.

I do not believe that receiving information from a secondary source such as AR glasses would cause cognitive overload, as the overload is caused by the uncertain situation in the first place. There is more such overload caused by the uncertainty than there is by the certainty of having some kind of a clear signal. A clear signal reduces the cognitive load irrespective of the communication technology that is used. Having everything clear is better than having some degree of uncertainty in a potential accident situation.

2.4.14 NEVILLE A. STANTON

Automated Vehicles: SAE Level 3 vehicles are currently being simulated on the roads through SAE Level 2 vehicles with their early warnings switched off. There are many problems that still need solving and this pushes the release of Level 4 out to 10–20 years from now. Truly SAE Level 5 that can operate on all roads (urban, rural, multi-carriageway), in all conditions, requires Artificial General Intelligence (AGI) and would, therefore, launch in the distant future beyond 50 years from now. With current AVs, it is still unclear who is in control, and this is causing collisions. For the time being, we should use automation only for motorway driving. It is unlikely that there will be a time when VRUs will be completely safe, as the AV would require humanlike AGI to handle all situations in a complex urban

setting. This requires a lot of time and training. However, current test vehicles operate at very slow speeds, nearly walking pace, and are ultra-cautious, so I do not necessarily see an immediate risk to people should these be implemented into urban environments.

V2I communication would be useful in promoting safety, but we must be careful not to remove all current infrastructure, since there will be people who would still enjoy driving and riding manually, for example, sports cars and motorcycles as well as classic and vintage vehicles. Smart infrastructure may direct pedestrians to cross anywhere on the road by relying on mapping technology and the AV's current position. Once again, the challenge to solve here are the manually-driven/ridden vehicles and how the infrastructure should handle them. In the future, when everything is connected, infrastructure will have an important role in changing the way we operate, and it may enable us to move away from vehicle ownership.

eHMIs: The premise of knowing the intention of the vehicle through eHMIs is a good one, and it is commendable to try and make them anthropomorphic. These interfaces can help but may also confuse and mislead due to cultural bounds. Something as simple as flashing the headlamps may be interpreted differently across cultures and contexts. For example, a signal from the host vehicle to give way to the oncoming vehicle or a signal from the host vehicle to the oncoming vehicle to say that I am coming through so you need to give-way, are polar opposite messages. Therefore, eHMIs must be simple and tested for all cultures and multiple scenarios since even basic symbols have the potential to be misinterpreted. Moreover, as Human Factors teaches us, we must design for all. It is my belief that the design should be the same for all pedestrians irrespective of their age; else we risk confusion.

Augmented Reality: AR can be introduced into the equation since we already have near-field communication based on location. AR glasses may allow for hands-free navigation and assist pedestrians with speed estimations by projecting the AV's trajectory. Moreover, I believe that AR could be beneficial to motorcyclists, as it is essential for them to know other road users' intentions. A problem arises in a mixed manual and automated environment where the VRU might not get the manual car driver's signal as they are looking for other communication. Providing individual context information to all actors around the AV seems very complex and difficult to achieve successfully. It would be much simpler to convey the AV's intentions and let each VRU interpret those intentions with reference to their own context. In contemporary urban design, mode separation of traffic such as by means of tunnels and bridges is the safest option. It would be interesting to see how the separation of pedestrians and vehicles could be done electronically through AR and smart infrastructure.

2.5 DISCUSSION

2.5.1 COMPARISON OF THE RESEARCHERS' VIEWS ON THE FUTURE OF AVS

The consensus among the researchers' views was that it will take decades for SAE Level 4 vehicles to be introduced on the roads, first on highways and later in urban environments. The consensus was that SAE Level 5 vehicles are a long way off, with some of the researchers stating that these would never come to fruition, highlighting that it is difficult for AVs to engage in social interaction in mixed traffic (see also Müller, Risto, & Emmenegger, 2016), drive in particular road types and weather conditions, and anticipate the behaviour of VRUs.

While Level 4 transport shuttles have already been introduced, they are still accompanied by a steward (see Heikoop, Nuñez Velasco, Boersma, Bjørnskau, & Hagenzieker, 2020). Some researchers noted that fully autonomous vehicles (SAE Level 5 automation) are already available in various industries, but that it would be very difficult to introduce their commercial equivalents in urban environments without changing the infrastructure or segregating VRUs from AVs. Of note, the researchers seemed coherent and did not appear to disagree about the state of current and future automation technology. However, the SAE levels themselves were regarded as open to different interpretations. Several researchers critiqued the ambiguity of the SAE levels (see also Hancock, 2020; Inagaki, & Sheridan, 2019), and one researcher even noted that SAE Level 5 is impossible by definition. Furthermore, the researchers expressed nuance and explained that the answer as to when SAE Level 4 and 5 automation will be available depends on the specific use of automation (e.g., public versus private transport).

Various researchers noted that shared control is a viable option in the short term and the horse metaphor was mentioned as an ideal type of shared control (Abbink, Mulder, & Boer, 2012; Flemisch et al., 2003; Norman, 2009). However, it was pointed out that shared control is not regarded as desirable in the long term, as shared control requires the human to be in the loop, which counteracts the idea of being able to benefit from a robot taxi (e.g., being able to work inside one's vehicle). Strikingly, many researchers used the words 'shared control' rather loosely (e.g., referring to automation that requires some human involvement) or in fact meant traded control (per Sheridan & Verplank, 1978). Traded control, that is, automation that requires the driver to take over immediately, was generally regarded as a bad idea as an operator cannot be expected to return into the loop quickly (Stanton et al., 1997; Mok et al., 2015, for further critical review, see Banks, Eriksson, O'Donoghue, & Stanton, 2018; Casner, Hutchins, & Norman, 2016). In the same vein, some of the researchers suggested that intermediate levels of automation should be skipped in favour of full autonomy.

AV-pedestrian interaction in the future urban environment was deemed challenging by the researchers, with various concerns focused on the behaviour of pedestrians themselves (see Domeyer, Lee, & Toyoda, 2020; Markkula et al., in press). More specifically, there was a worry that pedestrian-heavy areas will become more susceptible to jaywalking due to the pedestrians' expectation that AVs will always stop (see also Liu, Du, Wang, & Da Young, 2020; Millard-Ball, 2018). Such a scenario could render AVs immobile or make them drive very slowly under caution. It was suggested that pedestrians should change their behaviour and put safety first. Furthermore, the researchers recommended the segregation of traffic participants but noted that this would be a costly and resource-intensive solution.

A suggested solution for improved interaction was the inclusion of smart infrastructure in urban environments. In fact, the potential of smart infrastructure was recognised by many, but through different interpretations. It was mentioned that future infrastructure could receive state signals from AVs, communicate to VRUs wirelessly, or provide feedback to VRUs via the road surface, smart traffic lights, smartphones, and cloud systems. It was mentioned that smart infrastructure would be beneficial in pedestrian-heavy environments and could enhance the AVs's capabilities, such as by increasing their perceptual horizon. At the same time, it was argued by most of the researchers that cost, maintenance, and the reliability of wireless communication of smart infrastructure are major concerns. For these reasons, several researchers were outright negative about smart infrastructure, including V2I communication, especially for low-income countries. They saw a greater potential in

AVs that act independently from infrastructure and recommended further investment that direction.

2.5.2 COMPARISON OF VIEWS ON EHMIS AND AR FOR EHMIS

The majority of researchers pointed out that future AVs should or will be equipped with eHMIs. They noted that VRUs would like the AV to communicate to them (see also Habibovic et al., 2018; Nordhoff, Stapel, Van Arem, & Happee, 2020) and would tend to base their actions on the eHMI message received. It was also noted that car manufacturers would like to ensure that their AV communicates at least some cues for resolving confusion and for liability reasons. The researchers recognised the potential of anthropomorphic eHMIs and found anthropomorphism an interesting area of research. However, they tended to be critical towards superficial forms of anthropomorphism such as artificial eyes and smiles on the AV. It was generally mentioned that eHMIs should convey their state (allocentric communication) and not instruct VRUs what to do (egocentric communication). Furthermore, the majority mentioned that text-based eHMIs should not be used because they require translations to different languages, may be hard to read from a distance, and take time to read. Instead, eHMIs should be simple, perhaps signify a change of state and not much else, so that children can understand it too. The rejection of text-based eHMIs poses a dilemma because several empirical studies have found such eHMIs to be effective or preferred (see Ackermann, Beggiato, Schubert, & Krems, 2019; Bazilinskyy et al., 2019; Fridman et al., 2017). Implicit communication (i.e., vehicle speed, trajectory, and distance) was regarded as dominant and easiest to interpret (see Ackermann, Beggiato, Bluhm, & Krems, 2018; Moore, Currano, Strack, & Sirkin, 2019). According to the researchers, eHMIs are at best a secondary cue: An eHMI should support and confirm the existing implicit communication and not be detached from the AV (such as via infrastructure or projections on the road) because of issues of stimulus-response compatibility.

The researchers pointed out that AR is already successfully used in certain industries (e.g., manufacturing and agriculture) as well as in smartphones. However, they were generally critical towards AR in future traffic, noting challenges of privacy, invasiveness, user-friendliness, technological feasibility (brightness, image stability, and wireless communication reliability), and inclusiveness (i.e., not everybody having access to such devices). Participants were considering different types of AR in their answers, including basic vibrotactile feedback and feedback on smartphones, but also advanced types of visual feedback presented in head-mounted glasses. Furthermore, AR can be non-conformal and conformal, where the latter is defined as feedback that is embedded in the real world. Conformal AR allows for the most innovative feedback opportunities. The researchers noted AR possibilities such as holographic traffic lights and signs, the removal of irrelevant information in the world, an indicator nudging the user's attention towards an AV, the projection of safety zones or coloured road surfaces, or a fence or barrier indicating that one should not cross (see Eriksson et al., 2019 for a similar barrier concept for drivers). Several researchers noted that AR feedback is only useful for tasks that involve longer time constants, such as navigation and wayfinding. They noted that AR for short-term tasks such as collision avoidance is not feasible or useful, since implicit communication is dominant and the temporal window for possible interaction is short. The researchers concurred that AR could resolve the one-to-many problem in eHMIs, or resolve language barriers by providing person-specific feedback, although some

doubted this notion. The researchers noted that AR should be a secondary cue to implicit communication and eHMIs in the real layer because not everyone can be expected to wear an AR device.

The importance of standardisation of wireless communication protocols and eHMI designs were emphasised by many. The researchers indicated that the current proliferation of eHMI concepts is problematic (see also Dey et al., 2020a; Emmenegger, Risto, Bergen, Norman, & Hollan, 2016; Merat, Louw, Madigan, Wilbrink, & Schieben, 2018). There was an emphasis on the importance of standardising colours for use in eHMIs, similar to how there are current international standard colours for various traffic signals, such as red for stop (see Dey, Habibovic, Pfleging, Martens, & Terken, 2020b; Faas, & Baumann, 2019). Furthermore, training and education of the meaning of eHMI/AR-based feedback were regarded as essential.

It was interesting that the researchers, although often in agreement, differed in their degree of conservatism. Some participants were conservative and noted that much more research is needed about what types of eHMIs are needed if at all, and expressed concerns about VRU workload and reliability, cost, and maintenance of technology. Others were liberal in their thoughts and saw great promise in new technology. For example, they mentioned that AR and eHMIs have the potential to reduce VRU workload and confusion. One researcher mentioned that, given the fact that we develop automation that offloads drivers in AVs so that they can engage in non-driving tasks (e.g., infotainment, working), then why not develop similar support for VRUs, allowing VRUs to consume media and be informed about traffic via AR only when needed?

A number of researchers made enlightening parallels with current traffic. For example, it was mentioned that the car's horn and turn indicators are in fact eHMIs, that blind-spot warning systems and HUDs are existing forms of AR for drivers, that current pedestrian traffic lights are already an example of one-to-many communication (thereby suggesting that this problem does not need to be resolved), and that the issue of cross-cultural interpretation of texts and messages is present also in current traffic.

2.5.3 COMPARISON OF VIEWS ON VR AND AR SIMULATION

There were mixed opinions on the usefulness of VR simulation testing. Some stressed the importance of VR simulation in creating controlled, repeatable, and affordable environments, enabling experiments with large numbers of participants quickly and effectively. VR also allows for testing with vulnerable groups in a less risky environment (e.g., Deb et al., 2020). Some of the researchers who had worked with VR found that their results were in agreement with similar experiments that were carried out in the real world (see also Deb, Carruth, Sween, Strawderman, & Garrison, 2017; Fuest, Schmidt, & Bengler, 2020; Kaplan et al., in press; Klüver, Herrigel, Heinrich, Schöner, & Hecht, 2016). However, others stressed the importance of real-life studies, with VR testing being an intermediate step.

Several researchers mentioned that participants in a VR simulator study might not behave naturally due to their awareness that they are inside a VR environment and part of an experiment. It was noted that in experiments with human participants, the participants tend to please the investigator and attempt to do what the investigator wants them to do. Another reported downside to simulation testing is that most VR studies so far do not involve complicated scenarios but focus on one-to-one interactions without traffic, thus offering

only a slice of the full experience. Other critical notions on VR simulation were about the difficulty and cost required in creating a physical movement sensation. However, there was a compromise in this argument concerning CAVE simulators, which allow pedestrians to walk through the actual space and encounter cars driven by other humans.

While the researchers were somewhat negative about AR in future traffic, 15 of the 16 researchers regarded AR as a suitable or excellent research tool for simulating and testing eHMIs and training in a realistic context. It was pointed out that in AR-based experiments, there would be more natural triggers, situations, risk, and real decision-making as compared to VR. Whether the promise of AR as a research tool will be developing into a commercial product in future traffic remains to be seen.

2.6 CONCLUSIONS

THIS study invited 16 Human Factors researchers to share their views on the topic of AV-VRU communication in future traffic with discussion points ranging from their views on the future of AVs, smart infrastructure, eHMIs, AR for eHMIs, and simulation. The researchers agreed that SAE Level 5 automation is still far away from on-the-road implementation and that intermediate solutions such as shared control could be viable. At the same time, there was substantial heterogeneity on the definition of automated driving, as it can be used in different contexts such as public transport and specialised services. It was also believed that automation will improve safety; however, for that, rigorous measures such as segregated roads would be needed. The majority of researchers expressed concerns about smart infrastructure because of the cost and maintenance issues involved. They therefore felt that AVs that can move independently from infrastructure are the way forward.

The majority of researchers agreed that eHMIs will form part of the future interaction process between VRUs and AVs. However, they noted that there are still several open research questions that should be addressed before moving on to standardisation. The consensus appears to be that text-based eHMIs and eHMIs that provide instructions to VRUs should be avoided. The importance of experience/training and mental model formation was highlighted by several of the researchers. They emphasised that the long-term effects of eHMIs should be studied, and examine whether eHMIs are important relative to implicit communication, which is likely dominant in AV-VRU interaction.

AR technology and wearables were enthusiastically received by the researchers who, however, also highlighted various practical limitations such as user-friendliness, invasion of privacy, and information overload in case multiple streams of information compete for the wearer's attention. The notion that the one-to-many communication problem of current eHMIs could be solved through AR technology was positively welcomed. Various design concepts were provided, such as the use of virtual fences, the use of AR as a secondary cue to implicit communication, and person-specific feedback. However, at the same time, it was recognised that, for the time being, AR would be more of a research tool than something ready for public roads.

2.6.1 LIMITATIONS

A limitation of this paper is that it contains personal views that are not always backed up by empirical evidence. The researchers were asked to share their predictions on AVs decades

into the future, and these predictions may turn out to be inaccurate. For example, one of the researchers referred to the widespread future use of 5G technology for V2I and V2V communication. Although 5G is indeed regarded to have strong growth potential (Andrews, Buzzi, Choi, Hanly, Lozano, Soong, & Zhang, 2014; Li, Da Xu, & Zhao, 2018), it remains to be seen whether 5G will be used in future AVs on a broad scale. Similar remarks can be made about AR, which is a technology that is still in a nascent phase.

A second limitation is that although their views were solicited independently, the selected researchers for this study are not entirely independent because they know each other through the academic network. A third limitation is the lack of industry representation, a decision taken that was taken to not favour one industry entity over another. The industry is likely to bring different sort of topics into the fray, such as regarding standardisation, regulations, and commercial viability (Emmenegger & Norman, 2019). A mixture of perspectives from Human Factors experts and industry could yield interesting insights that did not come across in the present paper.

2.6.2 RECOMMENDATIONS FOR FUTURE RESEARCH

From the findings in this study, it is recommended that there be a push for standardisation of various eHMI elements. Also recommended are longitudinal studies to learn about the long-term effects of eHMIs (see also by Faas, Kao, & Baumann, 2020). Thirdly, the present interviews revealed that there is a lot to learn from current traffic and available research about anthropomorphism in robotics, HUDs in cars and aircraft, and existing types of communication such as brake lights and turn indicators, etc. Accordingly, we recommend that Human Factors scientists carefully familiarize themselves with the existing literature base before embarking on new empirical research. Finally, there should be a push towards standardisation of terminology in the research field. The present study highlighted differences between researchers in employed definitions, such as regarding shared control, smart infrastructure, and the SAE levels of automation. This semantic issue highlights that more communication is needed across the research field to homogenise such interpretations.

REFERENCES

Abbink, D. A., Mulder, M., & Boer, E. R. (2012). Haptic shared control: smoothly shifting control authority? Cognition, Technology & Work, 14, 19–28. https://doi.org/10.1007/s10111-011-0192-5

Ackermann, C., Beggiato, M., Bluhm, L. F., & Krems, J. (2018). Vehicle movement and its potential as implicit communication signal for pedestrians and automated vehicles. Proceedings of the 6th Humanist Conference. The Hague, Netherlands.

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, 75, 272–282. https://doi.org/10.1016/j.apergo.2018.11.002

Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C. K., & Zhang, J. C. (2014). What Will 5G Be? IEEE Journal on Selected Areas in Communications, 32, 1065–1082. https://doi.org/10.1109/jsac.2014.2328098

Applin, S. A., Riener, A., & Fischer, M. D. (2015). Extending driver-vehicle interface research into the mobile device commons: Transitioning to (nondriving) passengers and their vehicles. IEEE Consumer Electronics Magazine, 4, 101–106. https://doi.org/10.1109/MCE.2015.2463372

Azuma, R. T. (1997). A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6, 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. Applied Ergonomics, 68, 138–145. https://doi.org/10.1016/j.apergo.2017.11.010

Banks, V. A., Stanton, N. A., Burnett, G., & Hermawati, S. (2018). Distributed cognition on the road: Using EAST to explore future road transportation systems. Applied Ergonomics, 68, 258–266. https://doi.org/10.1016/j.apergo.2017.11.013

Banks, V. A., Stanton, N. A., & Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind' free driving. Safety Science, 62, 505–514. https://doi.org/10.1016/j.ssci.2013.10.014

Bärgman, J., & Victor, T. (2020). Holistic assessment of driver assistance systems: how can systems be assessed with respect to how they impact glance behaviour and collision avoidance? IET Intelligent Transport Systems, 14, 1058–1067. https://doi.org/10.1049/iet-its.2018.5550

Bazilinskyy, P., Dodou, D., & De Winter, J. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. Transportation Research Part F: Traffic Psychology and

Behaviour, 67, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2020). Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. Driving Simulation Conference Europe. Antibes, France.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. IEEE Intelligent Transportation Systems Magazine, 6, 6–22. https://doi.org/10.1109/mits.2014.2336271

Cabrall, C. D., Janssen, N. M., & de Winter, J. C. (2018). Adaptive automation: automatically (dis) engaging automation during visually distracted driving. PeerJ Computer Science, 4, e166. https://doi.org/10.7717/peerj-cs.166

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. Cognition, Technology & Work, 21, 3–20. https://doi.org/10.1007/s10111-018-0484-0

Casner, S. M., Hutchins, E. L., & Norman, D. (2016). The challenges of partially automated driving. Communications of the ACM, 59, 70–77. https://doi.org/10.1145/2830565

Cefkin, M., Zhang, J., Stayton, E., & Vinkhuyzen, E. (2019). Multi-methods research to examine external HMI for highly automated vehicles. In H. Krömker (Eds.), HCI in Mobility, Transport, and Automotive Systems. HCII 2019. Lecture Notes in Computer Science, vol 11596 (pp. 46–64). Cham: Springer. https://doi.org/10.1007/978-3-030-22666-4_4

Chang, C. M., Toda, K., Sakamoto, D., & Igarashi, T. (2017). Eyes on a car: An interface design for communication between an autonomous car and a pedestrian. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 65–73). Oldenburg Germany. https://doi.org/10.1145/3122986.3122989

Chang, K., Chang, C., Hou, H., Sung, Y., Chao, H., & Lee, C. (2014). Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. Computers & Education, 71, 185–197. https://doi.org/10.1016/j.compedu.2013.09.022

Colley, A., Häkkilä, J., Pfleging, B., & Alt, F. (2017). A design space for external displays on cars. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct (pp. 146–151). Oldenburg Germany. https://doi.org/10.1145/3131726.3131760

Deb, S., Carruth, D. W., Fuad, M., Stanley, L. M., & Frey, D. (2020). Comparison of child and adult pedestrian perspectives of external features on autonomous vehicles

using virtual reality experiment. In N. Stanton (Ed.), Advances in Human Factors of Transportation. AHFE 2019. Advances in Intelligent Systems and Computing, vol 964 (pp. 145–156) Cham: Springer. https://doi.org/10.1007/978-3-030-20503-4_13

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., & Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. Applied Ergonomics, 65, 449–460. https://doi.org/10.1016/j.apergo.2017.03.007

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., ... Terken, J. (2020a). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives, 7, 100174. https://doi.org/10.1016/j.trip.2020.100174

Dey, D., Habibovic, A., Pfleging, B., Martens, M., & Terken, J. (2020b). Color and animation preferences for a light band eHMI in interactions between automated vehicles and pedestrians. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, HI. https://doi.org/10.1145/3313831.3376325

Dey, D., & Terken, J. (2017). Pedestrian interaction with vehicles: roles of explicit and implicit communication. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 109–113). Oldenburg Germany. https://doi.org/10.1145/3122986.3123009

Dikmen, M., & Burns, C. M. (2016). Autonomous driving in the real world: Experiences with tesla autopilot and summon. Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 225–228). Ann Arbor, MI. https://doi.org/10.1145/3003715.3005465

Directive 2010/40. The framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport. European Parliament, Council of the European Union. Retrieved from http://data.europa.eu/eli/dir/2010/40/oj

Domeyer, J. E., Lee, J. D., & Toyoda, H. (2020). Vehicle automation – Other road user communication and coordination: Theory and mechanisms. IEEE Access, 8, 19860–19872. https://doi.org/10.1109/access.2020.2969233

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. IEEE Transactions on Intelligent Transportation Systems, 12, 596–614. https://doi.org/10.1109/tits.2010.2092770

Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems, 42, 177–190. https://doi.org/10.1016/s0921-8890(02)00374-3

Eisma, Y. B., Van Bergen, S., Ter Brake, S. M., Hensen, M. T. T., Tempelaar, W. J., & De Winter, J. C. F. (2020). External human–machine interfaces: The effect of display location on crossing intentions and eye movements. Information, 11, 13. https://doi.org/10.3390/info11010013

Emmenegger, C., & Norman, D. (2019). The challenges of automation in the automobile: Commentary on Hancock (2019) Some pitfalls in the promises of automated and autonomous vehicles. Ergonomics, 62, 512–513. https://doi.org/10.1080/00140139.2019.1563336

Emmenegger, C., Risto, M., Bergen, B., Norman, D., & Hollan, J. (2016). The critical importance of standards for the communication between autonomous vehicles and humans. Poster presented at the Automated Vehicles Symposium. San Francisco, CA.

Emuna, R., Borowsky, A., & Biess, A. (2020). Deep reinforcement learning for human-like driving policies in collision avoidance tasks of self-driving cars. Retrieved from https://arxiv.org/pdf/2006.04218.pdf

Eriksson, A., Petermeijer, S. M., Zimmermann, M., De Winter, J. C. F., Bengler, K. J., & Stanton, N. A. (2019). Rolling out the red (and green) carpet: Supporting driver decision making in automation-to-manual transitions. IEEE Transactions on Human-Machine Systems, 49, 20–31. https://doi.org/10.1109/thms.2018.2883862

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. Human Factors, 59, 689–705. https://doi.org/10.1177/0018720816685832

European Commission. (2019). Supporting the interaction of humans and automated vehicles: Preparing for the environment of tomorrow – Project description. Retrieved from https://cordis.europa.eu/project/id/860410

Faas, S. M., & Baumann, M. (2019). Light-based external human machine interface: Color evaluation for self-driving vehicle and pedestrian interaction. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63, 1232–1236. https://doi.org/10.1177/1071181319631049

Faas, S. M., Kao, A. C., & Baumann, M. (2020). A longitudinal video study on communicating status and intent for self-driving vehicle–pedestrian interaction. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–14). Honolulu, HI. https://doi.org/10.1145/3313831.3376484

Feldstein, I. T., Lehsing, C., Dietrich, A., & Bengler, K. (2018). Pedestrian simulators for traffic research: state of the art and future of a motion lab. International Journal of Human Factors Modelling and Simulation, 6, 250–265. https://doi.org/10.1504/ijhfms.2018.096128

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: S. S. Ge, O. Khatib, J. J. Cabibihan, R. Simmons, & M. A. Williams (Eds.), Social Robotics. ICSR 2012. Lecture Notes in Computer Science, vol 7621 (pp. 199–208). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34103-8_20

Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., & Schutte, P. C. (2003). The H-Metaphor as a guideline for vehicle automation and interaction (Report No. NASA/TM-2003-212672). Hampton: NASA, Langley Research Center.

Forster, Y., Naujoks, F., Neukum, A., & Huestegge, L. (2017). Driver compliance to take-over requests with different auditory outputs in conditional automation. Accident Analysis & Prevention, 109, 18–28. https://doi.org/10.1016/j.aap.2017.09.019

Fridman, L., Mehler, B., Xia, L., Yang, Y., Facusse, L. Y., & Reimer, B. (2017). To walk or not to walk: Crowdsourced assessment of external vehicle-to-pedestrian displays. Retrieved from https://arxiv.org/abs/1707.02698

Fuest, T., Schmidt, E., & Bengler, K. (2020). Comparison of methods to evaluate the influence of an automated vehicle's driving behavior on pedestrians: Wizard of Oz, virtual reality, and video. Information, 11, 291. https://doi.org/10.3390/info11060291

Geddes, N. B. (1940). Magic motorways. New York: Random House.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., … Larsson, P. (2018). Communicating intent of automated vehicles to pedestrians. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.01336

Haddon, W. (1970). On the escape of tigers: an ecologic note. American Journal of Public Health and the Nations Health, 60, 2229–2234. https://doi.org/10.2105/ajph.60.12.2229-b

Hagenzieker, M. (2015). Dat paaltje had ook een kind kunnen zijn [That bollard could have been a child]. Inaugural lecture. Delft University of Technology.

Hancock, P. A. (2020). Driving into the future. Frontiers in Psychology, 11. https://doi.org/10.3389/fpsyg.2020.574097

Heikoop, D. D., Nuñez Velasco, J. P., Boersma, R., Bjørnskau, T., & Hagenzieker, M. P. (2020). Automated bus systems in Europe: A systematic review of passenger experience and road user interaction. Advances in Transport Policy and Planning, (pp. 51–71)

Elsevier. https://doi.org/10.1016/bs.atpp.2020.02.001

Inagaki, T., & Sheridan, T. B. (2019). A critique of the SAE conditional driving automation definition, and analyses of options for improvement. Cognition, Technology & Work, 21, 569–578. https://doi.org/10.1007/s10111-018-0471-5

interACT. (2017). InterACT project. https://www.interact-roadautomation.eu.

International Organization for Standardization. (2018). Road Vehicles – Ergonomic Aspects of external visual communication from automated vehicles to other road users (ISO/TR Standard No. 23049).

Jurgen, R. K. (2006). Adaptive cruise control (No. PT-132). SAE Technical Paper.

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & De Winter, J. (in press). External Human-Machine Interfaces can be misleading. An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human Factors.

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (in press). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. Human Factors. https://doi.org/10.1177/0018720820904229

Kim, H. J., & Yang, J. H. (2017). Takeover requests in simulated partially autonomous vehicles considering human factors. IEEE Transactions on Human-Machine Systems, 47, 735–740. https://doi.org/10.1109/thms.2017.2674998

Klüver, M., Herrigel, C., Heinrich, C., Schöner, H., & Hecht, H. (2016). The behavioral validity of dual-task driving performance in fixed and moving base driving simulators. Transportation Research Part F: Traffic Psychology and Behaviour, 37, 78–96. https://doi.org/10.1016/j.trf.2015.12.005

Körber, M., Prasch, L., & Bengler, K. (2018). Why do I have to drive now? Post hoc explanations of takeover requests. Human Factors, 60, 305–323. https://doi.org/10.1177/0018720817747730

Kröger, F. (2016). Automated driving in its social, istorical and cultural contexts. Autonomous Driving, (pp. 41–68) Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-48847-8_3

Kyriakidis, M., De Winter, J. C. F., Stanton, N., Bellet, T., Van Arem, B., Brookhuis, K., ... Happee, R. (2019). A human factors perspective on automated driving. Theoretical Issues in Ergonomics Science, 20, 223–249. https://doi.org/10.1080/1463922x.2017.12931

87

Leiter, L. A., Fitchett, D. H., Gilbert, R. E., Gupta, M., Mancini, G. J., McFarlane, P. A., ... Ur, E. (2011). Identification and management of cardiometabolic risk in Canada: A position paper by the cardiometabolic risk working group (Executive Summary). Canadian Journal of Cardiology, 27, 124–131. https://doi.org/10.1016/j.cjca.2011.01.016

Li, S., Xu, L. D., & Zhao, S. (2018). 5G Internet of Things: A survey. Journal of Industrial Information Integration, 10, 1–9. https://doi.org/10.1016/j.jii.2018.01.005

Liu, P., Du, Y., Wang, L., & Da Young, J. (2020). Ready to bully automated vehicles on public roads? Accident Analysis & Prevention, 137, 105457. https://doi.org/10.1016/j.aap.2020.105457

Lorenzon, L., Biondi, A., Carus, T., Dziki, A., Espin, E., Figueiredo, N., ... D'Ugo, D. (2018). Achieving high quality standards in laparoscopic colon resection for cancer: A Delphi consensus-based position paper. European Journal of Surgical Oncology, 44, 469–483. https://doi.org/10.1016/j.ejso.2018.01.091

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 (pp. 1–12). Montréal, Canada. https://doi.org/10.1145/3173574.3174003

Markkula, G., Madigan, R., Nathanael, D., Portouli, E., Lee, Y. M., Dietrich, A., ... Merat, N. (in press). Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic. Theoretical Issues in Ergonomics Science. https://doi.org/10.1080/1463922x.2020.1736686

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing pedestrian behavior in augmented reality—proof of concept. Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 313–321). Atlanta, GA.. https://doi.org/10.1109/vr46266.2020.00051

Mayhew, S. (2015). A dictionary of geography. Oxford Paperback Reference.

Merat, N., Louw, T., Madigan, R., Wilbrink, M., & Schieben, A. (2018). What externally presented information do VRUs require when interacting with fully Automated Road Transport Systems in shared space? Accident Analysis & Prevention, 118, 244–252. https://doi.org/10.1016/j.aap.2018.03.018

Millard-Ball, A. (2018). Pedestrians, autonomous vehicles, and cities. Journal of Planning Education and Research, 38, 6–12. https://doi.org/10.1177/0739456x16675674

Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 2458–2464). Las Palmas, Spain. https://doi.org/10.1109/itsc.2015.396

Moore, D., Currano, R., Strack, G. E., & Sirkin, D. (2019). The case for implicit external human-machine interfaces for autonomous vehicles. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 295-307). Utrecht, Netherlands. https://doi.org/10.1145/3342197.3345320

Müller, L., Risto, M., & Emmenegger, C. (2016). The social behavior of autonomous vehicles. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (pp. 686–689). Heidelberg, Germany. https://doi.org/10.1145/2968219.2968561

Narzt, W., Pomberger, G., Ferscha, A., Kolb, D., Müller, R., Wieghardt, J., … Lindinger, C. (2006). Augmented reality navigation systems. Universal Access in the Information Society, 4, 177–187. https://doi.org/10.1007/s10209-005-0017-5

Nordhoff, S., Stapel, J., Van Arem, B., & Happee, R. (2020). Passenger opinions of the perceived safety and interaction with automated shuttles: A test ride study with 'hidden' safety steward. Transportation Research Part A: Policy and Practice, 138, 508–524. https://doi.org/10.1016/j.tra.2020.05.009

Norman, D. (2009). The design of future things. New York: Basic books.

Norman, D. (2014). Turn signals are the facial expressions of automobiles. New York: Diversion Books.

Oliveira, L., Proctor, K., Burns, C. G., & Birrell, S. (2019). Driving style: How should an automated vehicle behave? Information, 10, 219. https://doi.org/10.3390/info10060219

Perez, D., Hasan, M., Shen, Y., & Yang, H. (2019). AR-PED: A framework of augmented reality enabled pedestrian-in-the-loop simulation. Simulation Modelling Practice and Theory, 94, 237–249. https://doi.org/10.1016/j.simpat.2019.03.005

Potapov, E. V., Antonides, C., Crespo-Leiro, M. G., Combes, A., Färber, G., Hannan, M. M., … Gustafsson, F. (2019). 2019 EACTS expert consensus on long-term mechanical circulatory support. European Journal of Cardio-Thoracic Surgery, 56, 230–270. https://doi.org/10.1093/ejcts/ezz098

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.02757

Royal Academy of Engineering. (2012). Smart infrastructure: the future. Royal Academy of Engineering, London, UK. http://www.raeng.org.uk/publications/reports/smart-infrastructure-the-future.

SAE International. (2018). SAE International releases updated visual chart for its 'Levels of Driving Automation' standard for self-driving vehicles. Retrieved from https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-"levels-of-driving- automation"-standard-for-self-driving-vehicles.

Sanna, A., & Manuri, F. (2016). A survey on applications of augmented reality. Advances in Computer Science: An International Journal, 5, 18–27.

Schall, M. C., Rusch, M. L., Lee, J. D., Dawson, J. D., Thomas, G., Aksan, N., & Rizzo, M. (2013). Augmented reality cues and elderly driver hazard perception. Human Factors, 55, 643–658. https://doi.org/10.1177/0018720812462029

Schieben, A., Wilbrink, M., Kettwich, C., Dodiya, J., Weber, F., Sorokin, L., ... & Dietrich, A. (2019a). Testing external HMI designs for automated vehicles–An overview on user study results from the EU project interACT. In 9. Tagung Automatisiertes Fahren.

Schieben, A., Wilbrink, M., Kettwich, C., Madigan, R., Louw, T., & Merat, N. (2019b). Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations. Cognition, Technology & Work, 21, 69–85. https://doi.org/10.1007/s10111-018-0521-z

Schmidt, A. (2000). Implicit human computer interaction through context. Personal Technologies, 4, 191–199. https://doi.org/10.1007/bf01324126

Schneider, S., & Bengler, K. (2020). Virtually the same? Analysing pedestrian behaviour by means of virtual reality. Transportation Research Part F: Traffic Psychology and Behaviour, 68, 231–256. https://doi.org/10.1016/j.trf.2019.11.005

Schwebel, D. C., Gaines, J., & Severson, J. (2008). Validation of virtual reality as a tool to understand and prevent child pedestrian injury. Accident Analysis & Prevention, 40, 1394–1400. https://doi.org/10.1016/j.aap.2008.03.005

Semcon. (2016). The Smiling Car - Self driving car that sees you. Retrieved from https://semcon.com/smilingcar

Sewalkar, P., & Seitz, J. (2019). Vehicle-to-pedestrian communication for vulnerable road users: Survey, design considerations, and challenges. Sensors, 19, 358. https://doi.org/10.3390/s19020358

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Massachusetts Institute of Tech Cambridge Man-Machine Systems Lab.

https://doi.org/10.21236/ada057655

Stanton, N., Young, M., & McCaulder, B. (1997). Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. Safety Science, 27, 149–159. https://doi.org/10.1016/s0925-7535(97)00054-4

Stanton, N. A., Eriksson, A., Banks, V. A., & Hancock, P. A. (2020). Turing in the driver's seat: Can people distinguish between automated and manually driven vehicles? Human Factors and Ergonomics in Manufacturing & Service Industries. https://doi.org/10.1002/hfm.20864

Tabone, W. (2020). The Effectiveness of an augmented reality guiding system in an art museum. D. Seychell & A. Dingli (Eds.), Rediscovering Heritage Through Technology. Studies in Computational Intelligence, vol 859 (pp. 197–214) Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-36107-5_10

Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. Transport Reviews, 39, 103–128. https://doi.org/10.1080/01441647.2018.1494640

Toh, C. K., Sanguesa, J. A., Cano, J. C., & Martinez, F. J. (2020). Advances in smart roads for future smart cities. Proceedings of the Royal Society A, 476, 20190439. https://doi.org/10.1098/rspa.2019.0439

United Nations. (1968). Convention on road traffic. Done at Vienna on 8 November 1968. Amendment 1. Retrieved from http://www.unece.org/fileadmin/DAM/trans/conventn/crt1968e.pdf

Van Krevelen, D., & Poelman, R. (2010). A survey of augmented reality technologies: Applications and limitations. International Journal of Virtual Reality, 9, 1–20. https://doi.org/10.20870/ijvr.2010.9.2.2767

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident Analysis & Prevention, 78, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

APPENDIX

QUESTIONS FOR INTERVIEW WITH HUMAN FACTORS RESEARCHERS

THIS interview investigates researcher opinion and vision on the relationship between automated vehicles (AVs) and vulnerable road users (VRU) from a human factors perspective. It will discuss expectations on the introduction of automated vehicles on urban roads (using SAE levels of automation) and their safety from an external user's perspective. Following this is an exploration of current AV-VRU communication technologies, their effectiveness and potential solutions for interactions using an extended reality approach. The interview concludes with questions on simulation technology, its effectiveness and future potential.

GENERAL QUESTIONS ON AV

1. When are you expecting SAE level 4 and SAE level 5 vehicles to be introduced to urban roads?
2. Do you think we should still have shared control in AVs? Or should we opt for a complete SAE 5 vision?
3. What are your safety expectations towards pedestrians in a future scenario where AVs are on the roads?
4. Where do you think pedestrian-to-vehicle interaction is heading?
5. Do you think smart road infrastructure would have a pivotal role in the future urban traffic environment?
6. How do you think multiagent interactions and information sharing would occur in future traffic? i.e., : How will infrastructure communicate with cars/pedestrians?

QUESTIONS ON EHMIS

7. Do you think that external human-machine interfaces will be implemented into AVs in the future, and is it a good idea?
8. Do you think eHMIs would enhance pedestrian trust and feeling of safety in a future urban environment full of autonomous vehicles?
9. What kind of eHMIs do you think are more effective? Why?
10. Anthropomorphic or non-anthropomorphic eHMIs? Why?
11. Egocentric or allocentric eHMIs? Why?
12. Do you think an AV equipped with an eHMI communicating to multiple pedestrians in an urban setting would give rise to an ambiguous and confusing scenario?
13. Would you trust an eHMI which is not attached directly to the AV? (think of this as second-hand source information). Why?
14. Do you think such an approach may lead to cognitive overload? Why?

15. There is the issue of different crossing cultures and language barriers. How do you think this communication problem could be solved for eHMIs?

16. How should such technology adapt to children and older people?

QUESTIONS ON AR AND AR EHMIS

17. Do you think pervasive and context-aware AR technology is possible in the near future and how far away do you think we are from such technology?

18. Do you think this technology will be of benefit to humans?

19. How can AR be used in the future?

20. Would you be comfortable wearing AR wearables (glasses/lenses) when this technology becomes pervasive in the future?

21. Do you think AR would be helpful for VRU-AV communication? Why?

22. Do you think there is potential in using AR for eHMI design? Why?

23. What information do you think is essential for VRUs in an AR environment?

24. There are obviously several problems with this approach, namely not everyone having AR wearables, similar to not everyone being in possession of a smartphone. A possible solution to this would be to combine other eHMIs on the car or infrastructure which would still convey information to a pedestrian. What are your thoughts on this? What do you think would be the best possible combination?

25. With current solutions, it is still difficult to interpret to whom the AV is communicating to if there are multiple VRUs, and this could potentially cause accidents. Do you think that this approach would alleviate the ambiguous situation when there is an AV attempting to communicate with a group of VRUs?

26. There have also been occurrences where VRUs were confused in situations where textual eHMIs were utilised especially in cases where the text was written in a language unfamiliar to them. Do you think a customised AR eHMI approach would solve such a problem? Why?

QUESTIONS ON VR AND AR EXPERIMENTS

27. Current VR simulation studies investigate the behaviour of pedestrians when interacting with AVs. Do you think such studies replicate the behaviour of pedestrians in real-life traffic conditions? If not, how can this be solved?

28. Do you think there is potential in using AR cars for in situ eHMI simulation testing? Why?

29. Are there any advantages of AR over VR in this context? What are they?

3

Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes

This chapter was published as: *Tabone, W.*, Lee, Y. M., Merat, N., Happee, R., & De Winter J. C. F. (2021). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 209–218).

ABSTRACT

The future urban environment may consist of mixed traffic in which pedestrians interact with automated vehicles (AVs). However, it is still unclear how AVs should communicate their intentions to pedestrians. Augmented reality (AR) technology could transform the future of interactions between pedestrians and AVs by offering targeted and individualized communication. This paper presents nine prototypes of AR concepts for pedestrian-AV interaction that are implemented and demonstrated in a real crossing environment. Each concept was based on expert perspectives and designed using theoretically-informed brainstorming sessions. Prototypes were implemented in Unity MARS and subsequently tested on an unmarked road using a standalone iPad Pro with LiDAR functionality. Despite the limitations of the technology, this paper offers an indication of how future AR systems may support future pedestrian-AV interactions.

3.1 INTRODUCTION

FUTURE automated vehicles (AVs) have to be able to drive in complex environments containing many interaction partners, including vulnerable road users (VRUs). In recent years, the 'control loop', in which the automated vehicle locates itself, perceives its surroundings, and decides upon the best trajectory, has changed into an 'interaction loop', where multiple road users cooperate through the wireless exchange of information (Troel-Madec et al., 2019). To close the interaction loop, the AV may need to communicate its intentions to VRUs, who traditionally received such information explicitly from the driver and implicitly through vehicle kinematics (Lee et al., 2020). Current solutions are smart infrastructure and vehicle-mounted external human-machine interfaces (eHMIs) (De Clercq et al., 2019; Dietrich et al., 2019; Mahadevan et al., 2018; Sorokin et al., 2019), and see Dey et al., 2019 for a review of 70 eHMI concepts. However, these interfaces may be hard to read from a distance and potentially ambiguous, especially when encountering multiple pedestrians, as it could be unclear to whom the AV is communicating (Rasouli & Tsotos, 2019).

AR has already been used to support operators in many domains, including the military (Livingston et al., 2011; TechCrunch 2021), museums (Tabone 2020), industry (Ma et al., 2011), and entertainment (Von Itzstein et al., 2017). As AR technology becomes more context-aware and pervasive (Grubert et al., 2016), AR may become of benefit to pedestrians. Through AR, it becomes possible to remove the display from the vehicle and place it anywhere in the world. Since each pedestrian would receive individualised communication, AR could solve the one-to-many communication problem of eHMIs (Tabone et al., 2021). Additionally, AR may prove advantageous in terms of costs, as it may be cheaper to build virtual interfaces than it would be to build their physical counterparts such as eHMIs.

AR has already been employed in traffic research, for example, to provide drivers with navigation advice in the form of augmented road signs (Rusch et al., 2014) or to inform the driver about safe and unsafe slots via green and red coloured road surfaces (Wang et al., 2020). Previous research on AR for VRUs has focused on handheld devices (Lakehal et al., 2020; Tang & Zhou, 2020) and AR glasses (Montuwy et al., 2018) for pedestrian route guidance and navigation. Furthermore, a 'pedestrian in the loop' system that uses the HoloLens to augment vehicles in the real environment for testing safety-critical situations

has been developed by Hartmann et al., 2018, whereas Kamalasanan et al. 2021 conducted a HoloLens experiment to investigate the effect of an AR traffic light on pedestrian behaviour. Additionally, AR concepts that aim to provide crossing advice to VRUs have been presented as superimposed layers on a photo of the streetscape (Hesenius et al., 2018) and as part of a VR simulation (Pratticò et al., 2021). However, to date, the literature presents only few AR prototypes for pedestrian-AV interactions in a real environment. Furthermore, the literature offers little guidance regarding the question: how should such prototypes function and look? This paper attempts to address this question by presenting AR prototypes for pedestrian-AV interactions that function in a real environment. The prototypes were designed using a theoretically informed approach that draws upon fundamental principles and concepts such as affordances (Gibson, 1986)], the field of safe travel (Gibson & Crooks, 1938), and risk perception (Schiff et al., 1962). In the spirit of open science, the code behind the AR concepts is provided as supplementary material (doi: https://10.4121/14933082) to facilitate reproducibility and encourage further development.

In this paper, we present nine AR concepts that aim to support pedestrians that want to cross the road. The concepts were derived by building upon interviews with human factors experts (Tabone et al, 2021) and supported by theories of perception, human factors, and spatial computing.

3.2 METHODS

A four-phase process was employed to create the AR prototypes. In Phase 1, expert perspectives were analysed to extract AR design ideas. Phase 2 entailed identifying literature that served as theoretical background and inspiration for AR design. The next phase included an iterative design (Nielsen, 1994) brainstorming process that involved discussions on design considerations and sketching, followed by a heuristic evaluation. Phase 4 covered the implementation of the sketched concepts.

3.2.1 PHASE 1: COLLECTION OF IDEAS

In Tabone et al., 2021, experts in human factors provided their views on automated driving in future urban environments, eHMIs, and the use of AR in future traffic. For the present study, the experts' statements were extracted from Tabone et al., 2021 and its corresponding interview transcripts to identify concepts for pedestrian-AV interaction using AR. The analysis yielded seven quotes from the experts that could be turned into implementable concepts:

1. "It would be ideal to provide a unified affordance of the intentions of the vehicle, rather than just a visual colouring. For example, it would be interesting to alter the perceived surface of the AV to make it appear as more or less threatening, contingent on need." (P. A. Hancock, [Tabone et al., 2021, p. 6]).

2. "Make objects stand out using subjective contours." (P. A. Hancock, personal correspondence with J. de Winter based on Hancock, 2019).

3. "The AR system should provide information related to safety. The pedestrian could be presented with safety corridors related to which vehicles will stop for them. The advantage of this safety corridor concept is that it is integrating information from

several vehicles. In doing so, it is clear to whom the vehicles are communicating, in contrast to the undirected communication of traditional eHMIs." (Martin Baumann, [Tabone et al., 2021, p. 4]). "AR glasses may allow for hands-free navigation and assist pedestrians with speed estimations by projecting the AV's trajectory." (Neville A. Stanton, [Tabone et al., 2021, p.11]).

4. "AR should not overwhelm the user. Information should include highlighting of hazards such as specific alerts that a vehicle is approaching." (Shuchisnigdha Deb, [Tabone et al., 2021, p. 5]).

5. "Augmented 3D traffic light in the form of a virtual fence to stop pedestrians from crossing a vehicle lane. It would be ideal for tram lanes as well, and simple enough for a child to understand even in ambiguous situations when multiple pedestrians are present." (Riender Happee, [Tabone et al., 2021, p. 7]).

6. "It would be interesting to be notified not just about the AV's intention to stop but also about where it intends to stop." (Marieke Martens, [Tabone et al., 2021, p. 8]).

7. "The design should ensure that users direct their attention to the right information at the right time." (Natasha Merat, [Tabone et al., 2021, p. 9]).

These quotes served as the foundation for designing the concepts during the following phases.

3.2.2 COLLECTION OF RELEVANT LITERATURE

Literature was collected to serve as a theoretical foundation for the subsequent design phase. For each idea reported above, two authors (Authors 1 and 5) identified supported theories and related concepts from a previously compiled literature folder containing 357 papers on eHMIs, AR, and other supports for VRUs. The selection of papers was based on seminal works in the field of Human Factors and Ergonomics.

3.2.3 PHASE 3: BRAINSTORMING DESIGN SESSIONS

Brainstorming design sessions were conducted amongst the authors (Authors 1, 2, 4, & 5). The same scenario with a vehicle approaching from the right was envisaged for all concepts. An unmarked two-way road in Malta was chosen as the environment on which the AR concepts would be mapped. The car was stationary in front of the pedestrian crossing area to allow all AR concepts to be demoed irrespective of technical limitations.

It was agreed that two states would be created per concept, that is, 'vehicle is yielding', and 'vehicle is not yielding'. These states would be communicated using green and red, respectively, consistent with a study that showed these colours to be most intuitive for pedestrians judging the intentions of an approaching vehicle (Bazilinskyy et al., 2020). None of the concepts would communicate instructions, as instructions may lead to erroneous crossing decisions (Lee et al., 2019). Moreover, it was decided that concepts that were not augmented on the ground would be made of translucent material not to block the user's view of the real world. It was agreed that each concept had a supporting cue, such as movement or iconography, to facilitate redundancy gain (Lee et al., 2017).

MIRO (Miro, 2021), an online collaborative environment, was used as a supporting tool. The MIRO board contained an affinity diagram and a table with 15 design principles for human-computer interaction (Lee et al., 2017). Images taken from a LiDAR scan of the chosen environment were added to the board to aid the participants.

During the brainstorm, each concept idea from Phase 1 and corresponding literature from Phase 2 was brought forward. Design elements (e.g., dimensions, form, iconography, animation) were discussed at length with reference to the literature identified for the concept (see Phase 2). For each concept, virtual post-it notes representing different design elements were added to the board. The authors also used experience-based design to reach decisions. In total, the brainstorming was spread across three separate sessions for a total of 5 hours.

Through the design process, the seven design ideas extracted from Tabone et al., 2021 evolved into the following nine AR concepts: (1) augmented zebra crossing, (2) planes on vehicle, (3) conspicuous and looming planes, (4) field of safe travel, (5) fixed pedestrian traffic lights, (6) virtual fence, (7) phantom car, (8) nudge HUD, (9) pedestrian traffic lights HUD.

A low-fidelity paper sketch was created for each concept based on the design elements identified for each concept. Finally, a heuristic evaluation was carried out based on nine AR heuristics proposed by Endsley et al., 2017. If shortcomings in the design were flagged through the heuristic evaluation, a design modification was conducted to accommodate. Examples include modifying Concepts 8 and 9 so that the messages were placed at the top of the field of view (FOV) rather than the middle to support the heuristic 'fit with the user's perceptual abilities'. Heuristics that were approved without modification to the concepts include 'alignment of physical and virtual worlds', 'accounting for hardware capabilities', and 'adaptation to user position and motion', amongst others. Once all authors were confident with the modified sketches and heuristic checklist, the concepts were moved to Phase 4.

3.2.4 PHASE 4: IMPLEMENTATION

HARDWARE AND SOFTWARE.

The implementation was carried out using Unity 3D and its new tool for AR developers: Mixed and Augmented Reality Studio (MARS) v.1.2, together with Apple ARKit (v.2.1.9) and AR Foundation (v.2.1.8) libraries. A 15" MacBook Pro (2019) with 2.3 GHz 8-Core Intel Core i9 processor, 16 GB DDR4 memory, and Radeon Pro 560X 4 GB GPU was used for development, whereas an iPad Pro 11" (2020) was used as the target device because of its inbuilt LiDAR. A Toyota Yaris Hybrid vehicle was used for demonstrating the AR concepts.

A LiDAR 3D scan of the crossing and surrounding area of the road was captured (Figure 3.1) using the 3dScanner app (3DScannerApp, 2021) on the iPad set on a lower quality setting to reduce storage requirements. The scan with 1.1 million vertices was exported as an OBJ file and imported into Blender, where texture files extracted from the 3D scanning software were added, and the final scan was exported as an FBX file. A Unity MARS simulation environment was created using the FBX to test the augmented concepts in the Unity editor before actual testing on site.

IMPLEMENTING THE AR CONCEPTS IN MARS.

To anchor the concepts to the real world, several Unity MARS tools were used (Unity, 2021). Proxies, representing real-world objects (e.g., a flat surface, vehicle) that the final application



Figure 3.1: *Left*: The chosen unmarked crossing in Malta. *Right*: The captured LiDAR 3D scan on location.

could detect, were used to define regions where augmented elements could appear. Some proxies included semantic tags (i.e., to specify that they represent the user in the real world or the floor), whereas others included *Conditions*, *Rules*, and *Forces*, which affect the behaviour of the proxies.

Our implemented app, built through Unity MARS, used the camera and LiDAR of the iPad to scan the real environment, generate planes of the world, and spawn objects if proxy conditions were met (e.g., horizontal surface of a certain size). During the first visit to the testing location, the Unity MARS companion app (Unity, 2021b) was used to capture the environment dimensions, image markers, and proxy data of the crossing region and vehicle, to be used in the implementation. For each implementation, the objects that comprise the AR concept were added as children to their Proxy object. *ShowChildrenOnTrackingAction* was applied, so that child objects were spawned, oriented, and anchored with the proxy object once the proxy's conditions were met.

For concepts that were augmented on the ground (i.e., Concepts 1 and 6), the crossing region's horizontal proxy data was used together with a number of conditions. More specifically, *IsPlaneCondition*, *FlatFloorCondition*, and a *'floor'* semantic tag were added to specify that the object to augment upon had to be a flat surface and floor. An *AlignmentCondition* was added to align objects in a *HorizontalUp* orientation, whereas a *PlaneSizeCondition* specified that a 1.5 m x 1.5 m flat plane had to be scanned and mapped for augmentation to occur. To make the AR concept align with the user, the objects augmented on the proxy were aligned to a *ProxyObjectReferencePerson* (representing the user's position) through a *ProxyAlignmentForce*.

Image markers (i.e., doorway across the street, photo of the car) were used for instances where a crossing region (i.e., Concepts 5 and 7) or a car (i.e., Concepts 2 and 3) had to be recognised. In these cases, a *MarkerCondition* was added to the proxy, and the image to be recognised was selected from the Marker Library imported through the companion app. The child objects would spawn and anchor to the real-world region that was mapped through the image marker.

For Concept 4, a direct placement script by the Unity MARS team (Unity, 2020) was used so that an object could be spawned and mapped to any surface found by raycasting. Lastly, concepts with a HUD (Concepts 8 and 9) used the traditional Unity UI workflow in a Unity MARS session.

3.2.5 CREATING THE FINAL APPLICATION AND TESTING ON-SITE.

For each concept, a plane or point cloud visualisers were added to offer visual feedback to the user that the system was detecting the environment during operation. The shifting from the yielding to the non-yielding state was performed manually via a tap on the iPad screen.

During each iteration, the prototype was built as an iOS app onto the iPad and tested at the chosen location until its performance concerning the mapping of the visual elements to the real world was satisfactory. A single app comprising all implemented concepts and a menu was exported and executed on-site. The prototypes were recorded using the inbuilt iOS screen capture. These recordings, together with an additional number of demos involving a moving car, which yielded or maintained speed, were compiled into a single video, which is available as supplementary material. In addition, the material includes a photograph detailing how the iPad was used.

3.3 RESULTS AND DISCUSSION

3.3.1 CONCEPT 1: AUGMENTED ZEBRA CROSSING

Concepts 1 and 2 were derived from the first quote in Phase 1. For Concept 1, the part on "providing a unified affordance rather than just a visual colouring" was taken into account for the road. Hence, for Concept 1, we decided to change the surface and colour of the road (Figure 3.2). Similar non-augmented concepts have used LEDs embedded in the road to display crossings on-demand [59]. In the same vein, Eriksson et al., 2018 presented an in-vehicle augmented HMI that showed a red demarcation or green carpet on the motorway depending on whether it was unsafe or safe to change lanes.

Our prototype for Concept 1 in a non-yielding state consists of a zebra crossing with a red cross overlaid on top. For the yielding state, the red cross is replaced by two rectangular green bars that flank the zebra crossing on either vertical side. These display elements were inspired by a previously published implementation of a smart road (Löcken et al., 2019). A zebra crossing with additional elements was preferred over an entire green/red surface with overlaid arrows because the latter may offer an instructive suggestion to cross.

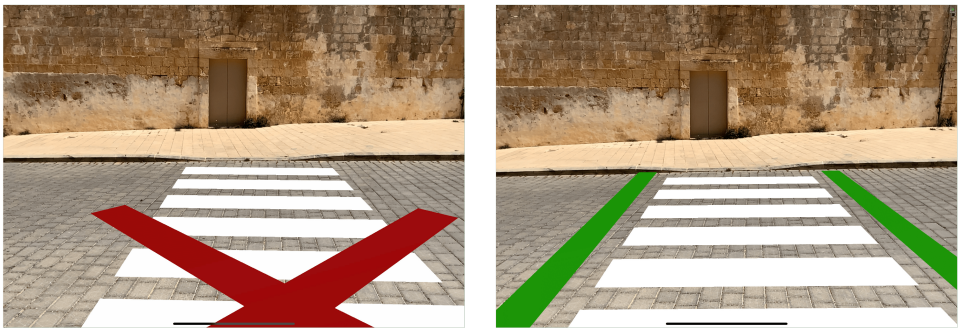


Figure 3.2: Concept 1. *Left*: non-yielding state with an 'X' placed on top of the augmented zebra crossing. *Right*: yielding state with two green bars flanking the zebra crossing.

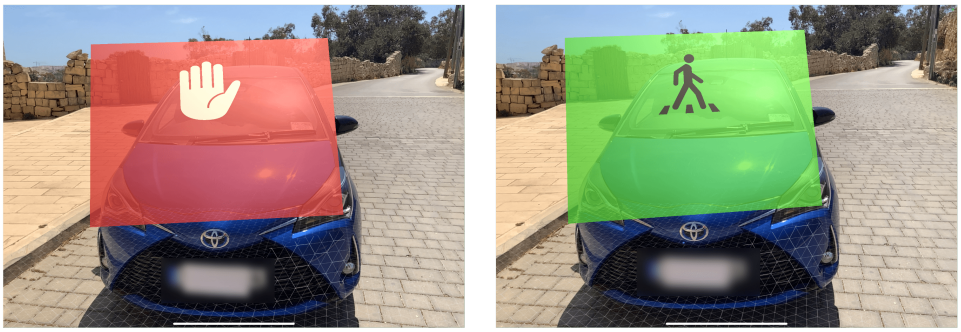


Figure 3.3: Concept 2. *Left*: non-yielding state with the red plane and hand icon augmented on the vehicle. *Right*: yielding state with the green plane and crossing figure icon augmented on the vehicle.

3.3.2 CONCEPT 2: PLANES ON VEHICLE

This AR concept entailed a change of the vehicle surface (Figure 3.3). We added a plane that would appear at the top front part of the vehicle. While vehicle kinematics combined with such an AR concept may already offer a rich cue, there is still the issue that a red car may be perceived as yielding (rather than the green car) (Bazilinskyy et al., 2020; Lee et al., 2019). Therefore, redundancy gain was employed (Wickens et al., 2015) by superimposing standard icons representing yielding (walking human figure) and non-yielding (hand palm) on the windshield, considering that pedestrians are likely to focus their visual attention on that region (Dey et al., 2019; De Winter et al., 2021).

In our prototype and supplementary video, the non-yielding red plane is spawned as soon as the image marker condition for the vehicle is satisfied (that is, the vehicle used for the image marker is detected).

3.3.3 CONCEPT 3: CONSPICUOUS LOOMING PLANES

For the third AR concept, which is based on the second quote from Phase 1, the theories of looming and conspicuity (Hancock, 2019) were drawn upon. In the study of perception, looming concerns the human response to the rapid approach of a solid body (Schiff et al., 1962). Through looming, it is possible to communicate threat and distinguish an approaching object from one that is stationary or receding (Caird & Hancock, 1994; Gibson & Carmichael, 1966). Conspicuity is defined as "something that is obvious to the eye or mind, or something that is striking and attracts attention" (Merriam-Webster, 2021). Using these principles, Concept 2 was modified so that the red plane (non-yielding) grows in size as it approaches the user and fills the entire FOV by the end, while conversely, the green plane (yielding) shrinks as the vehicle approaches. Hence, threat is mapped to the size of the plane as the vehicle approaches the crossing. In the supplementary video, the camera moves towards the vehicle as the planes grow or shrink, dependent on the state, to simulate the vehicle approaching.

3.3.4 CONCEPT 4: FIELD OF SAFE TRAVEL

The point of departure for this AR concept (Figure 3.4) was a 'safety corridor' presented to the pedestrian, as mentioned in the third quote from Phase 1. Gibson's field of safe travel



Figure 3.4: Concept 4. *Left*: non-yielding state with the red field of safe travel extending over the crossing area. *Right*: yielding state with the green field of travel shortened and truncating before the crossing area.

(Gibson & Crooks, 1938) was the theory behind this concept. The field at any given moment represents the paths that a car may take unimpeded, and it would look like a "sort of tongue protruding forward along the road" (Gibson & Crooks, 1938, Kolekar et al., 2020). The concept relates to the theory of affordances (Gibson, 1986) since the shape of the field is not related to arbitrary rules but reflects physical laws. In simpler terms, the projected field indicates where a vehicle may be in a given amount of time. Hence, if a person enters a red zone, they risk being hit by the vehicle. For the non-yielding state, the red field extends beyond the crossing point, whereas for the yielding state, the field switches to green, widens, and terminates prior to the point of crossing to emphasize the vehicle will not cross beyond this point.

In our approach, the field is represented as a flat plane emerging from the front of the vehicle. If the prototype was to be fully implemented, the plane's length would be proportional to the vehicle's speed. However, to demonstrate both states using the prototype for the supplementary video, the length of the plane was hardcoded. Hesenius et al., 2018 presented a similar concept of Safe Zones for pedestrians, highlighting regions on the road. Areas marked in green denoted zones that could be safely traversed; conversely, areas marked in red denoted potentially dangerous areas to navigate.

3.3.5 CONCEPT 5: FIXED PEDESTRIAN TRAFFIC LIGHTS

For this AR concept (Figure 3.5), based on the fourth quote in Phase 1, the familiar concept of binary pedestrian traffic lights was chosen. These lights would pop up on the other side of the crossing area to alert the pedestrian whether a vehicle is approaching and intending to yield. As an additional cue to the red and green traffic light beams used for the non-yielding and yielding states, icons of a static and walking human figure are superimposed, respectively. Moreover, since this pedestrian traffic light is a replica of its real-world counterpart, it is attached to the ground to comply with the user's mental model.

For the implementation, the 3D model of the lights was set as a child of the crossing image marker proxy. Once the iPad was pointed towards the doorway across the road, the traffic lights model was augmented on the pavement.



Figure 3.5: Concept 5. *Left*: non-yielding state with red pedestrian icon. *Right*: yielding state with green pedestrian icon.

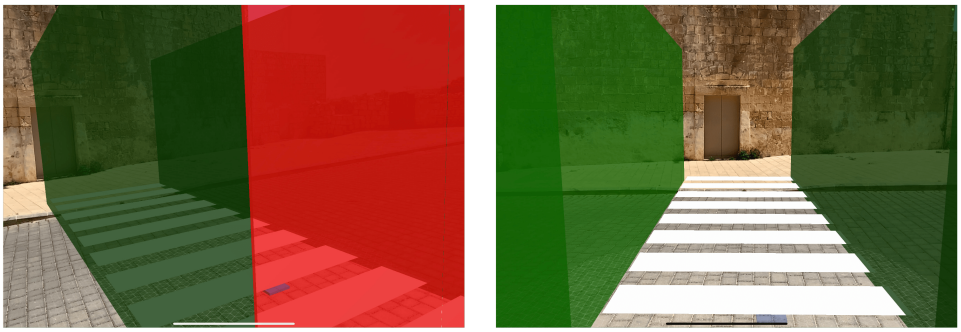


Figure 3.6: Concept 6. *Left*: non-yielding state with the closed red gate 'blocking' the pedestrian. *Right*: yielding state, where the gate is green and opened.

3.3.6 CONCEPT 6: VIRTUAL FENCE

Concept 6 was to "augment a 3D traffic light in the form of a virtual fence", as stated in the fifth quote from Phase 1. Discussions commenced with the idea of having a fence preventing pedestrians from entering the road. However, concerns arose that the fence could invite the pedestrian to walk around it, which is undesirable. Therefore, the fence concept was operationalized in the form of a tunnel (Figure 3.6).

The tunnel is created using semi-translucent green walls on either side of the crossing area, high enough to enclose an average person. A zebra crossing is displayed on the tunnel's floor to offer redundancy gain, while a semi-translucent red gate of the same height blocks the pedestrian's path to the other side of the road in the non-yielding state. The sides of the tunnel are green to indicate that cars could cross through that side of the tunnel. The gate's colour switches to green for the yielding state, and an animation of the gate swinging open plays out. For the implementation, the proxies and zebra crossing used in Concept 1 were built upon.

The 'tunnels in the sky' (Mulder, 1999; Tsang & Vidulich, 2002) AR/HUD concept from the domain of aviation was a source of inspiration for the design. Furthermore, a 'safe corridor' augmented in-vehicle HMI, which communicates to the driver whether an area



Figure 3.7: Concept 7. *Left*: non-yielding state with the red phantom car blocking the crossing area. *Right*: yielding state with the real car entering the green phantom car.

on the road should be avoided (if red) or safely steered into (green) (Lorenz et al., 2014), served as an additional source of inspiration. It has similarities to the on-the-road carpet HMI consulted for Concept 1 (Eriksson et al., 2018). These carpet metaphors inspired the addition of a zebra crossing placed on the ground between the walls.

3.3.7 CONCEPT 7: PHANTOM CAR

In this case, in addition to the vehicle's intention to yield, the concept communicates its intended stopping location (Figure 3.7). A previously published concept called 'Vehicle Intents' (Hesenius et al., 2018) highlighted the vehicle's path with an arrow and added stopping lines to communicate where the vehicle would stop if the pedestrian was to cross. The idea was further developed to include an augmented representation of a hatchback car, similar to what could be seen as part of a ghost replay system in a racing computer game or Tesla's shadow mode (Forbes, 2019).

Hence, the phantom car concept was created. A semi-translucent vehicle, which switches between green and red depending on the scenario, is augmented. The phantom car moves in front of the real vehicle to indicate where the vehicle would be after a particular time (e.g., 2 seconds into the future). In the yielding scenario, as the real vehicle slows down, the distance between itself and the green phantom car will shrink until the real vehicle has fully entered the phantom car. In the non-yielding scenario, the red phantom car will communicate its intent by moving in front of the actual vehicle at a constant distance.

In the prototype and supplementary video, the non-yielding red phantom car was augmented on the crossing to simulate that the vehicle intends to move beyond the stopping area; hence, it is unsafe to cross. In the yielding scenario, the green phantom car was augmented next to the stationary real vehicle to simulate the moment before the real vehicle had fully entered the area occupied by the phantom car.

3.3.8 CONCEPT 8: NUDGE HUD

Distracted pedestrians are the cause of many accidents (Barton et al., 2016; Mwakalonge et al., 2015). An interface that assists the pedestrian with looking towards the right direction may therefore have a benefit. This concept addresses a different challenge from the previous concepts, as the latter assume that the pedestrian is looking at the AR display anchored to the

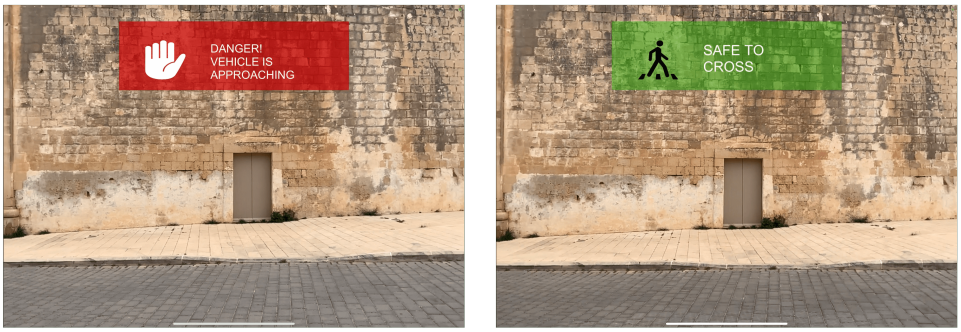


Figure 3.8: Concept 8. *Left*: the non-yielding state with the red HUD showing the hand icon and corresponding text message. *Right*: the yielding state with the green HUD, walking pedestrian icon, and corresponding text message.

environment. In Concept 8, a HUD that follows the pedestrian's FOV regardless of where they are looking, was created (Figure 3.8).

The HUD was drawn in a billboard style where text in sans serif typeface Arial was placed on a background plane of a different colour to offer good performance in terms of readability (Bernard et al., 2003; Gabbard et al., 2006; Renkewitz et al., 2008). For the non-yielding state, the text is "DANGER! VEHICLE IS APPROACHING", which changes to "SAFE TO CROSS" in the yielding state (Fridman et al., 2017). The messages use non-instructing language since there are dangers of misinterpretation in providing explicit advice (Lee et al., 2019). Icons were added to the left of the text and switched between an outstretched hand and a walking human figure, respectively.

The HUD was display-locked so that it stayed put in the same location on the display, which would be analogous to head-locked positioning on a head-mounted display such as AR glasses. This placement decision was based on empirical research recommending head-locked text placement if the real-world task involves permanent visual monitoring in the central or near-peripheral visual field (Klose et al., 2019), which in our case would be monitoring of the vehicle. Moreover, the HUD was placed in the centre top section of the FOV since it is recommended to place text in top positions in complex environments that require constant monitoring (Klose et al., 2019). Top positioning would also leave the bottom part of the FOV not occluded so that the pedestrian could still see where they are walking.

3.3.9 CONCEPT 9: PEDESTRIAN TRAFFIC LIGHTS HUD

For this concept, the pedestrian traffic light model used for Concept 5 is displayed as a HUD that follows the user's FOV (Figure 3.9). The model was modified to remove the pole that attaches the traffic light to the ground to create a 'detached' interface element. Moreover, the same interface placement location used for Concept 8 was adopted, with the difference that instead of using UI elements, the 3D traffic light model was made a child of the main camera to follow the FOV.



Figure 3.9: Concept 9. *Left*: non-yielding state with the red pedestrian icon. *Right*: yielding state with the green pedestrian icon.

3.4 GENERAL DISCUSSION

IN this paper, nine AR concepts that aim to support pedestrians in making crossing decisions when encountering future traffic in an urban environment were presented as implemented prototypes. Expert perspectives (Tabone et al., 2021) were analysed to extract AR concept ideas. Subsequently, a literature analysis was performed to identify similar approaches and theories that aided in transforming the idea into a prototype. Following this phase, brainstorming design sessions were conducted, resulting in nine designs in the form of sketches on paper. Underlying these ideas were theories of affordances (Gibson, 1986), multimodal communication (Burke et al., 2004), looming and conspicuity (Hancock, 2019; Schiff et al., 1962), and AR research in aerospace (Mulder, 1999; Tsang & Vidulich, 2002). Finally, implementation was conducted using Unity MARS, a new tool for AR developers, and the AR concepts were tested and filmed on an uncontrolled crossing.

During the brainstorming process and based on the literature consulted, it was decided that colour alone was 'not enough' and that redundancy gain had to be employed (Wickens et al., 2015). Therefore, icons were added. In the same vein, the mere presence of a zebra crossing does not guarantee safety in the real world. With the addition of the red cross and the green bars, the interface communicated that it was unsafe or safe to cross *at that moment*. Moreover, based on the literature, it was decided that messages should not be instructive (Lee et al., 2019). Hence, the concepts were made comparable in the sense that none instructed pedestrians but rather gave the suggestion that it is safe or unsafe to cross.

Occlusion (Shah et al., 2012; Witte, 2020) was a problematic factor to handle, which is why this was discussed in detail during the brainstorming session and the subsequent heuristic evaluation. The occlusion issue (Peddie, 2017) was experienced first-hand, where the gate and walls of the virtual fence (Concept 6) had to be semi-translucent so that the vehicle could be seen. The same principle was applied to the other concepts that were not augmented on the road surface.

3.4.1 LIMITATIONS AND FUTURE WORK

Several assumptions were made, and the presented concepts should not be viewed as fully finished AR interfaces that VRUs can already use. The implemented concepts and videos

take only one vehicle into account. Furthermore, although most of the prototypes were augmented using Unity MARS functionalities such as Proxies, Conditions, and Forces, for all concepts, the state change for vehicle intent (i.e., yielding, non-yielding) was manually triggered using a tap on the iPad. In real-life applications, the path planner of the AV, connected to its CAN bus, could send out a wireless signal to the VRU's AR system when the AV-VRU distance or time-to-arrival drops below a threshold value.

Concepts such as 8 and 9 were easy to implement as little context is needed for these to operate. However, because of the nascent status of the technology used, various workarounds were implemented for concepts that required knowledge of the position of the approaching vehicle (Concepts 2, 3, 4, 7) and its speed (Concepts 3, 4, and 7). These workarounds included having the real vehicle remaining static since moving real-world objects are not yet supported by the libraries used. While the road position was handled by Unity MARS through the camera and LiDAR, vehicle position was hardcoded, or image markers were relied upon. Moreover, the dimensions of the zebra crossing and tunnel for Concepts 1 and 6 were set manually, based on the 3D LiDAR scan model of the environment. For Concept 3, vehicle position readouts from LiDAR raycasts were tried to map the looming plane's size with AV-pedestrian distance. However, the desired precision was not attained, and hence an animator script was used to resize the planes. Future improvements could be accomplished through edge computer vision methods to detect the distance and speeds of approaching road users. Alternatively, IoT communication could be used, where AVs transmit their yielding intent, position, and speed wirelessly to the VRU, as pointed out above.

Although the current implementation was done on an iPad Pro (to make use of its LiDAR), the concepts could be readily exported to the head-mounted displays currently supported by Unity and Unity MARS. In that case, sensor requirements would need to be met, and issues such as outdoor luminance levels that may hinder perception (Do et al., 2020), ocular vergence-accommodation conflicts (Rolland et al., 1995; Wann et al., 1995) and latency issues leading to visually induced motion sickness (Buker et al., 2012) have to be considered. Sensor requirements vary between concepts according to complexity, as shown in Table 3.1.

Table 3.1: Sensor requirements for the concepts.

Concept	**Road position**	**Vehicle intent**	**Vehicle position**	**Vehicle speed**
(1) Augmented zebra crossing	X	X		
(2) Planes on vehicle		X	X	
(3) Conspicuous looming planes		X	X	X
(4) Field of safe travel	X	X	X	X
(5) Fixed pedestrian traffic lights	X	X		
(6) Virtual fence	X	X		
(7) Phantom car		X	X	X
(8) Nudge HUD		X		
(9) Pedestrian traffic lights HUD		X		

In the future, user studies are needed to identify which concepts are best accepted, how the concepts would function on AR glasses in the urban environment, and what triggers should be used to change between yielding and non-yielding states. The intuitiveness of the mappings used, such as the intuitiveness of the looming cues in Concept 3 and the red vs. green colours used for the tunnel walls in Concept 6, needs to be investigated as well. Furthermore, it would be important to include provisions for VRUs with visual impairments, for example, by adding tactile or auditory signals. Finally, it would be worthwhile to explore how to integrate the AR concepts with other communication modalities for VRUs, such as eHMIs and smart infrastructure.

3.5 CONCLUSION

In conclusion, this paper presents an outlook on future AV-VRU communication through AR in the form of prototypes. This paper presents nine such concepts designed using theory, expert opinion, literature, brainstorm, and implementation in Unity. In the spirit of open science, the code is provided as supplementary material to encourage further development.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Unity MARS team for their generous support. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860410.

REFERENCES

Barton, B. K., Kologi, S. M., & Siron, A. (2016). Distracted pedestrians in crosswalks: An application of the Theory of Planned Behavior. Transportation Research Part F: Traffic Psychology and Behaviour, 37, 129-137. https://doi.org/10.1016/j.trf.2015.12.012

Bazilinskyy, P., Dodou, D., & De Winter, J.C.F. (2020). External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'? In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 3721-3728). https://doi.org/10.1109/SMC42975.2020.9282998

Bernard, M. L., Chaparro, B. S., Mills, M. M., & Halcomb, C. G. (2003). Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. International Journal of Human-Computer Studies, 59(6), 823-835. https://doi.org/10.1016/S1071-5819(03)00121-6

Buker, T. J., Vincenzi, D. A., & Deaton, J. E. (2012). The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: Reducing apparent latency with predictive compensation. Human Factors, 54(2), 235-249. https://doi.org/10.1177/0018720811428734

Burke, J. L., Murphy, R. R., Rogers, E., Lumelsky, V. J., & Scholtz, J. (2004). Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(2), 103-112. https://doi.org/10.1109/TSMCC.2004.826287

Caird, J. K., & Hancock, P. A. (1994). The perception of arrival time for different oncoming vehicles at an intersection. Ecological Psychology, 6(2), 83-109. https://doi.org/10.1207/s15326969eco0602_1

De Clercq, K., Dietrich, A., Velasco, J. P. N., & De Winter, J.C.F. (2019). External Human-Machine Interfaces on automated vehicles: effects on pedestrian crossing decisions. Human Factors, 61(8), 1353-1370. https://doi.org/10.1177/0018720819836343

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., & Terken, J. (2020). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives, 7(9), 100174. https://doi.org/10.1016/j.trip.2020.100174

Dey, D., Walker, F., Martens, M., & Terken, J. (2019). Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external Human-Machine Interfaces for automated vehicles. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19) (pp. 369-378). Association for Computing Machinery. https://doi.org/10.1145/3342197.3344523

Dietrich, A., Willrodt, J.-H., Wagner, K., & Bengler, K. (2019). Projection-based external human machine interfaces-enabling interaction between automated vehicles and pedestrians. In Proceedings of the DSC 2018 Europe VR, September 5 - 7, 2018, Antibes, France (pp. 43-50).

Do, T. D., LaViola, J. J., & McMahan, R. P. (2020). The effects of object shape, fidelity, color, and luminance on depth perception in handheld mobile augmented reality. In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), November 9 - 13, Porto de Galinhas, Brazil (pp. 64-72). https://doi.org/10.1109/ISMAR50242.2020.00026

Endsley, T. C., Sprehn, K. A., Brill, R. M., Ryan, K. J., Vincent, E. C., & Martin, J. M. (2017). Augmented reality design heuristics: Designing for dynamic interactions. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61(1), 2100-2104. https://doi.org/10.1177/1541931213602007

Eriksson, A., Petermeijer, S. M., Zimmermann, M., De Winter, J.C.F, Bengler, K. J., & Stanton, N. A. (2018). Rolling out the red (and green) carpet: supporting driver decision making in automation-to-manual transitions. IEEE Transactions on Human-Machine Systems, 49(1), 20-31. https://doi.org/10.1109/THMS.2018.2883862

Forbes. (2019, April). Tesla's "Shadow" Testing Offers A Useful Advantage On The Biggest Problem In Robocars. Retrieved March 25, 2021, from https://www.forbes.com/sites/bradtempleton/2019/04/29/teslas-shadow-testing-offers-a-useful-advantage-on-the-biggest-problem-in-robocars

Fridman, L., Mehler, B., Xia, L., Yang, Y., Facusse, L. Y., & Reimer, B. (2017). To walk or not to walk: Crowdsourced assessment of external vehicle-to-pedestrian displays. arXiv preprint arXiv:1707.02698. Retrieved from https://arxiv.org/abs/1707.02698

Gabbard, J. L., Swan, J. E., & Hix, D. (2006). The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. Presence: Teleoperators & Virtual Environments, 15(1), 16-32. https://doi.org/10.1162/pres.2006.15.1.16

Gibson, J. J. (1986). The Ecological Approach to Visual Perception (Vol. 1). Psychology Press.

Gibson, J. J., & Carmichael, L. (1966). The Senses Considered as Perceptual Systems (Vol. 2, No. 1). Houghton Mifflin.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. The American Journal of Psychology, 51(3), 453-471. https://doi.org/10.2307/1416145

Grubert, J., Langlotz, T., Zollmann, S., & Regenbrecht, H. (2017). Towards pervasive augmented reality: Context-awareness in augmented reality. IEEE Transactions on Visual-

ization and Computer Graphics, 23(6), 1706-1724. https://doi.org/10.1109/TVCG.2016.2543720

Hancock, P. A. (2019). On the dynamics of conspicuity. Human Factors, 61(6), 857-865. https://doi.org/10.1177/0018720819840849

Hartmann, M., Viehweger, M., Stolz, M., Watzenig, D., Spitzer, M., & Desmet, W. (2018). "Pedestrian in the Loop": An approach using augmented reality. SAE Technical Paper, 2018-01-1053. https://doi.org/10.4271/2018-01-1053

Hesenius, M., Börsting, I., Meyer, O., & Gruhn, V. (2018). Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18) (pp. 261-268). Association for Computing Machinery. https://doi.org/10.1145/3236112.3236148

Von Itzstein, S., Billinghurst, M., Smith, R. T., & Thomas, B. H. (2017). Augmented reality entertainment: Taking gaming out of the box. In Encyclopedia of Computer Graphics and Games (pp. 1-9). Springer International Publishing. https://doi.org/10.1007/978-3-319-08234-9_81-1

Kamalasanan, V., Schewe, F., Sester, M., & Vollrath, M. (2021). Exploratory Study on the Use of Augmentation for Behavioural Control in Shared Spaces. Retrieved April 26, 2021, from ResearchGate. https://www.researchgate.net/publication/350879518_Exploratory_Study_on_the_use_of_Augmentation_for_Behavioural_Control_in_Shared_Spaces

Klose, E. M., Mack, N. A., Hegenberg, J., & Schmidt, L. (2019). Text presentation for augmented reality applications in dual-task situations. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 636-644). Osaka, Japan: IEEE. https://doi.org/10.1109/VR.2019.8797992

Kolekar, S., De Winter, J.C.F., & Abbink, D. (2020). Human-like driving behaviour emerges from a risk-based driver model. Nature Communications, 11(1), 1-13. https://doi.org/10.1038/s41467-020-18353-4

Lakehal, A., Lepreux, S., Efstratiou, C., Kolski, C., & Nicolaou, P. (2020). Investigating smartphones and AR glasses for pedestrian navigation and their effects in spatial knowledge acquisition. In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20) (pp. 1-7). New York, NY: ACM. https://doi.org/10.1145/3406324.3410722

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). Designing for People. CreateSpace.

Lee, Y. M., Madigan, R., Garcia, J., Tomlinson, A., Solernou, A., Romano, R., Markkula, G., Merat, N., & Uttley, J. (2019). Understanding the messages conveyed by automated vehicles. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19) (pp. 134-143). New York, NY: ACM. https://doi.org/10.1145/3342197.3344546

Lee, Y. M., Madigan, R., Uzondu, C., Garcia, J., Romano, R., Markkula, G., & Merat, N. (2020). Learning to interpret novel eHMI: The effect of vehicle kinematics and eHMI familiarity on pedestrians' crossing behaviour. Retrieved from https://psyarxiv.com/2xub4

Livingston, M. A., Rosenblum, L. J., Brown, D. G., Schmidt, G. S., Julier, S. J., Baillot, Y., SwanII, J. E., Ai, Z., & Maassel, P. (2011). Military applications of augmented reality. In Handbook of Augmented Reality (pp. 671-706). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-0064-6_31

Löcken, A., Golling, C., & Riener, A. (2019). How should automated vehicles interact with pedestrians? A comparative analysis of interaction concepts in virtual reality. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19) (pp. 262-274). New York, NY: ACM. https://doi.org/10.1145/3342197.3344544

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 1681-1685. Sage California, Los Angeles, CA. https://doi.org/10.1177/1541931214581351

Ma, D., Gausemeier, J., Fan, X., & Grafe, M. (2011). Virtual Reality & Augmented Reality in Industry. Berlin Heidelberg: Springer.

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating awareness and intent in autonomous vehicle-pedestrian Interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18) (pp. 1-12). New York, NY: ACM. https://doi.org/10.1145/3173574.3174003

Merriam-Webster. (2021). Dictionary by Merriam-Webster. (October 2017). Retrieved March 25, 2021, from https://www.merriam-webster.com

Miro. (2021). An Online Visual Collaborative Platform for Teamwork. (March 2021). Retrieved March 25, 2021, from https://miro.com/

Shah, M. M., Arshad, H., & Sulaiman, R. (2012). Occlusion in augmented reality. In 8th International Conference on Information Science and Digital Content Technology (ICIDT2012), 372-378. Jeju, South Korea.

Montuwy, A., Cahour, B., & Dommes, A. (2018). Older pedestrians navigating with AR glasses and bone conduction headset. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18) (pp. 1–6). New York, NY: ACM. https://doi.org/10.1145/3170427.3188503

Mulder, M. (1999). Cybernetics of Tunnel-In-The-Sky Displays. (Ph.D. Dissertation). Delft University of Technology, Delft, The Netherlands. ISBN: 90-407-1963-2.

Mwakalonge, J., Siuhi, S., & White, J. (2015). Distracted walking: Examining the extent to pedestrian safety problems. Journal of Traffic and Transportation Engineering (English edition), 2(5), 327-337. https://doi.org/10.1016/j.jtte.2015.08.004

Nielsen, J. (1994). Usability Engineering. Burlington, MA: Morgan Kaufmann.

Peddie, J. (2017). Augmented Reality: Where We Will All Live. Springer.

Pratticò, G., Lamberti, F., Cannavò, A., Morra, L., & Montuschi, P. (2021). Comparing state-of-the-art and emerging augmented reality interfaces for autonomous vehicle-to-pedestrian communication. IEEE Transactions on Vehicular Technology, 70(2), 1157-1168. https://doi.org/10.1109/TVT.2021.3054312

Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems, 21(3), 900-918. https://doi.org/10.1109/TITS.2019.2901817

Renkewitz, H., Kinder, V., Brandt, M., & Alexander, T. (2008). Optimal font size for head-mounted-displays in outdoor applications. In 12th International Conference Information Visualisation (pp. 503-508). London, UK. https://doi.org/10.1109/IV.2008.73

Rolland, J. P., Gibson, W., & Ariely, D. (1995). Towards quantifying depth and size perception in virtual environments. Presence: Teleoperators & Virtual Environments, 4(1), 24-49. https://doi.org/10.1162/pres.1995.4.1.24

Rusch, M. L., Schall Jr, M. C., Lee, J. D., & D, J. D. (2014). Augmented reality cues to assist older drivers with gap estimation for left-turns. Accident Analysis & Prevention, 71, 210-221. https://doi.org/10.1016/j.aap.2014.05.020

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". Science, 136(3520), 982-983. https://doi.org/10.1126/science.136.3520.982

Sorokin, L., Chadowitz, R., & Kauffmann, N. (2019). A change of perspective: Designing the automated vehicle as a new social actor in a public space. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19) (pp.

1-8). New York, NY: ACM. https://doi.org/10.1145/3290607.3299044

Tabone, W. (2020). The effectiveness of an augmented reality guiding system in an art museum. In Rediscovering Heritage Through Technology (pp. 197-214). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-36107-5_10

Tabone, W., De Winter, J.C.F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D., Sheridan, T. B., & Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transportation Research Interdisciplinary Perspectives, 9, 100293. https://doi.org/10.1016/j.trip.2020.100293

Tang, L., & Zhou, J. (2020). Usability assessment of augmented reality-based pedestrian navigation aid. In 22nd HCI International Conference (HCII 2020), July 19 - 24, 2020, Copenhagen, Denmark (pp. 581-591). https://doi.org/10.1007/978-3-030-49904-4_43

TechCrunch. (2021). Microsoft gets contract worth up to $22 billion to outfit US Army with 120,000 AR headsets. (March 2021). Retrieved April 1, 2021, from https://techcrunch.com/2021/03/31/microsoft-wins-contract-worth-up-to-22-billion-to-outfit-u-s-army-with-120000-ar-headsets

3DScannerApp. (2021). 3D Scanner App - LIDAR Scanner for iPad & iPhone Pro. (March 2021). Retrieved March 30, 2021, from https://www.3dscannerapp.com/

Troel-Madec, M., Boissieux, L., Borkoswki, S., Vaufreydaz, D., Alaimo, J., Chatagnon, S., & Spalanzani, A. (2019). EHMI positioning for autonomous vehicle/pedestrians interaction. In Proceedings of the 31st Conference on l'Interaction Homme-Machine: Adjunct (IHM '19) (pp. 1-8). ACM. https://doi.org/10.1145/3366551.3370340

Tsang, P. S., & Vidulich, M. A. (2002). Principles and Practice of Aviation Psychology. CRC Press.

Umbrellium. (2017). Umbrellium develops light-up crossing that only appears when needed. (October 2017). Retrieved March 25, 2021, from https://www.dezeen.com/2017/10/12/umbrellium-develops-interactive-road-crossing-that-only-appears-when-needed-technology

Unity. (2020). MARS Direct Placement Example. (June 2020). Retrieved April 26, 2021, from https://forum.unity.com/threads/mars-direct-placement-example.908381

Unity. (2021). Unity MARS Overview. (March 2021). Retrieved April 26, 2021, from https://docs.unity3d.com/Packages/com.unity.mars@1.3/manual/index.html

Unity. (2021b). Unity MARS Companion app - Open Beta announcement. (Jan 2021). Retrieved March 30, 2021, from https://forum.unity.com/threads/unity-mars-companion-app-open-beta-announcement.1037638

Wang, Z., Han, K., & Tiwari, P. (2020). Augmented reality-based advanced driver-assistance system for connected vehicles. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), October 11 - 14, 2020, Toronto, ON, Canada (pp. 752-759). https://doi.org/10.1109/SMC42975.2020.9283462

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. Vision Research, 35(19), 2731-2736. https://doi.org/10.1016/0042-6989(95)00018-U

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). Engineering Psychology and Human Performance. Psychology Press.

De Winter, J.C.F., Bazilinskyy, P., Wesdorp, D., De Vlam, V., Hopmans, B., Visscher, J., & Dodou, D. (2021). How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study. Ergonomics, 1-16. https://doi.org/10.1080/00140139.2020.1862310

Witte, T. P. (2020). Design of an Augmented Reality System for the Live Visualization of Ship Identification Data (Master's thesis). Delft University of Technology, Delft, The Netherlands.

II

Part II: Evaluation

4

Augmented reality interfaces for pedestrian-vehicle interactions: An online study

This chapter was published as: *Tabone, W.*, Happee, R., García, J., Lee, Y. M., Lupetti, M. C., Merat, N., & De Winter, J. C. F. (2023). Augmented reality interfaces for pedestrian-vehicle interactions: An online study. Transportation Research Part F: Traffic Psychology and Behaviour, 94, 170–189.

ABSTRACT

Augmented Reality (AR) technology could be utilised to assist pedestrians in navigating safely through traffic. However, whether potential users would understand and use such AR solutions is currently unknown. Nine novel AR interfaces for pedestrian-vehicle communication, previously developed using an experience-based design method, were evaluated through an online questionnaire study completed by 992 respondents in Germany, the Netherlands, Norway, Sweden, and the United Kingdom. The AR indicated whether it was safe to cross the road in front of an approaching automated vehicle. Each interface was rated for its intuitiveness and convincingness, aesthetics, and usefulness. Moreover, comments were collected for qualitative analysis. The results indicated that interfaces that employed traditional design elements from existing traffic, and head-up displays, received the highest ratings overall. Statistical results also showed that there were no significant effects of country, age, and gender on interface acceptance. Thematic analysis of the textual comments offered detail on each interface design's stronger and weaker points, and revealed unintended effects of certain designs. In particular, some of the interfaces were commented on as being dangerous or scary, or were criticised that they could be misinterpreted in that they signal that something is wrong with the vehicle, or that they could occlude the view of the vehicle. The current findings highlight the limitations of experience-based design, and the importance of applying legacy design principles and involving target users in design and evaluation. Future research should be conducted in scenarios in which pedestrians actually interact with approaching vehicles.

4.1 INTRODUCTION

FUTURE traffic, in which automated vehicles (AVs) will be driving in city environments, requires transparent communication of the intentions of the vehicle with interaction partners, such as vulnerable road users (VRUs). In traditional traffic, transparent communication between vehicles and vulnerable road users is achieved through implicit and explicit cues (Lee et al., 2021; Schieben et al., 2019). Implicit cues include vehicle speed, dynamics, and gap size, while explicit cues include the horn, hand gestures, and eye contact. VRUs base their crossing decisions primarily on implicit cues (Dey & Terken, 2017; Lee et al., 2021), whereas explicit cues tend to be used when implicit cues are ambiguous (Onkhar et al., 2021; Uttley et al., 2020). With the introduction of AVs in the urban environment, the lack of a driver or attentive passenger may require a different approach to communicating intent from the AV to the VRU (Ackermans et al., 2020; Carmona et al., 2021; Faas et al., 2020; Hensch et al., 2019). Several communication methodologies have been proposed to alleviate the problems of AV-VRU interactions. These include the use of smart road infrastructure (Löcken et al., 2019; Pompigna & Mauro, 2022; Toh et al., 2020), smart vehicle kinematics through the use of vehicle pitch, deceleration, and lateral position (Bindschädel et al., 2022; Dietrich et al., 2020; Fuest et al., 2018; Sripada et al., 2021), and external human-machine interfaces (eHMIs).

Various forms of eHMIs have been developed, including LED strips, LED screens, anthropomorphic elements, actuated robotic attachments, and projections on the road, amongst others (see Bazilinskyy et al., 2019; De Winter & Dodou, 2022; Dey et al., 2020a; Rouchitsas & Alm, 2019, for reviews of such interfaces). Despite their effectiveness in encouraging

VRUs to (not) cross in front of the AV's path, current eHMI designs have some drawbacks, namely if the eHMI needs to signal to a single pedestrian in a group, or, for text-based eHMIs, if the message is in a language unfamiliar to the pedestrian. Furthermore, so far, there has been no standardisation of eHMIs, and therefore pedestrians may encounter a variety of different eHMIs on vehicles, which could cause confusion (Rasouli & Tsotsos, 2020; Tabone et al., 2021a), with potentially dangerous consequences.

In an effort to address some of these problems, augmented reality (AR) has been proposed as a new type of communication in traffic. AR used by individual VRUs can alleviate several issues, especially the one-to-many communication problem, where multiple actors (vehicles and pedestrians) are present in the environment and it is not clear which actor is communicating to whom. Through AR, the communication signal could be sent individually and separately to each pedestrian, and does not have to be constrained to the AV itself but can be presented anywhere in the environment (Tabone et al., 2021b; Tran et al., 2022).

So far, studies on AR for pedestrian-vehicle interaction consider the driver as the AR user, by highlighting pedestrians and/or cyclists in front of the vehicle (e.g., Calvi et al., 2020; Colley et al., 2021; Currano et al., 2021; Kim et al., 2018; Pichen et al., 2020). Such solutions are becoming technologically feasible when considering that the most recent vehicle models already feature AR-based head-up displays (Volkswagen, 2020). The use of AR by VRUs themselves is still relatively rare and has mostly been constrained to route navigation tasks (e.g., Bhorkar, 2017; Dancu et al., 2015; Dong et al., 2021; Ginters, 2019), for example as an add-on to Google Maps (Ranieri, 2020). Only a small, but growing number of studies have examined the use of AR for supporting VRUs in making safe crossing decisions. Examples include road projections such as zebra crossings, safe paths, and arrows (Hesenius et al., 2018; Li et al., 2022; Pratticò et al., 2021; Tran et al., 2022), visualisation of obstructed vehicles (Matviienko et al., 2022; Von Sawitzky et al., 2020), visualisation of collision times and conflict points (Tong & Jia, 2019), warning signs (Tong & Jia, 2019; Von Sawitzky et al., 2020), and car overlays (Tran et al., 2022). Using virtual reality, Oudshoorn et al. (2021) developed bioinspired eHMIs for pedestrian-AV interaction, whereas Mok et al. (2022) developed eHMIs in the form of laser-type rays emitted from the AV. The authors noted that these types of eHMIs may be hard to physically implement on real AVs, and that AR used by pedestrians (such as through AR glasses or handheld devices) could be a viable alternative.

It should be noted that most AR concepts for VRUs are still in a conceptual stage (videos, virtual reality), while only a few AR interfaces for VRUs have been demonstrated on a real road (Maruhn et al., 2020; Tabone et al., 2021b), or in a laboratory environment (Matviienko et al., 2022; Pratticò et al., 2021; Tran et al., 2022). In Tabone et al. (2021b), novel AR interfaces for pedestrian-AV interaction were developed and demonstrated in a real crossing environment. The interfaces were designed to assist pedestrians in the decision to cross the road in front of an approaching automated vehicle which was either yielding (stopping) or non-yielding. The interfaces were based on expert perspectives extracted from Tabone et al. (2021a) and designed using theoretically-informed brainstorming sessions (see Figure 4.1 for the interfaces). In total, nine AR interfaces were designed, each with a non-yielding and yielding state, depicted in red and green respectively. These colours were selected based on their high intuitiveness rating for signalling 'please (do not) cross' (Bazilinskyy et al., 2020).

Three of the interfaces were mapped to the road, four were mapped to the vehicle, and two were head-locked to the user's field of view. The ones mapped to the road were the

augmented zebra crossing, which is a traditional zebra crossing design (1 in Figure 4.1), fixed pedestrian traffic lights (5), which depicts a familiar pedestrian traffic light design across the road, and a virtual fence (6), which includes semi-translucent walls around a zebra-crossing and a gate that opens in the yielding state. The interfaces that were mapped to the vehicle included the planes on the vehicle (2), which displays a plane on the windshield area of the vehicle, the conspicuous looming plane (3), which grows or shrinks as the vehicle approaches the pedestrian depending on the AV's yielding state, the field of safe travel (4) which projects a field on the road in front of the vehicle to communicate safety, and the phantom car (7) which projects the vehicle's predicted future motion. The final two interfaces are head-up displays: the nudge head-up display (HUD) (8), which displays text and icons, and the pedestrian lights HUD (9), which displays a head-locked version of the pedestrian traffic lights.

In Tabone et al. (2021b), the interfaces were implemented on a handheld device (iPad Pro 2020) and demonstrated in a real crossing environment (Figure 4.1), but no user study was performed. The concepts were designed using a 'genius'-based design approach (Saffer, 2010). In contrast to other design approaches, genius design does not involve users as part of the formal research phase. Instead, the design team relies on personal experience, existing knowledge of human behaviour, the problem space, and human cognition and psychology (Saffer, 2010). This approach offers the benefit of time efficiency, coherence of solutions with the original vision, and the flexibility to generate ideas quickly. Yet, such an approach could be contested as it addresses the problem space only from a designer's viewpoint without the involvement of the intended users (Nielsen, 2007).

Although a theoretical evaluation based on nine AR heuristics (Endsley et al., 2017) was performed in Tabone et al. (2021b), it is vital that AR concepts are evaluated empirically to assess whether the theoretically informed ideas are valid. Such an empirical evaluation would assess the viability of the 'genius' design approach in Tabone et al. (2021b) and whether the designers' intended effects would generalise to potential target users. Conducting a real-world study with the implemented AR prototypes would have been very difficult at the time of writing due to AR technology limitations, such as outdoor luminance levels that may hinder perception, latency issues that may lead to visually induced motion sickness, and ocular vergence-accommodation conflicts in open spaces (Buker et al., 2012; Rolland et al., 1995; Wann et al., 1995). Therefore, an online questionnaire study approach with a large number of participants was selected. A substantial number of previous works have conducted online user surveys to evaluate eHMIs for pedestrian-AV interaction (e.g., Bai et al., 2021; Bazilinskyy et al., 2020, 2021; Dey et al., 2020b; Lau et al., 2021). However, no large-sample survey of AR interfaces for VRU-AV interactions has been conducted so far.

Hence, we attempt to fill this gap and build upon the previous design work reported in Tabone et al. (2021b) by conducting an online video-based questionnaire study that investigates user acceptance of the AR interfaces across large numbers of participants, exploring key moderator variables (e.g., nationality, gender). Ratings of intuitiveness, convincingness, usefulness, aesthetics, and satisfaction with the interface were captured, which were thought to represent key dimensions of interface quality. These measures were based on previous studies which explored intuitiveness (Bazilinskyy et al., 2020), usefulness (Adell, 2010), quality of information (Lau et al., 2021), as well as aestheticism, attractiveness, and visibility (Métayer & Coeugnet, 2021). More specifically, it was reasoned

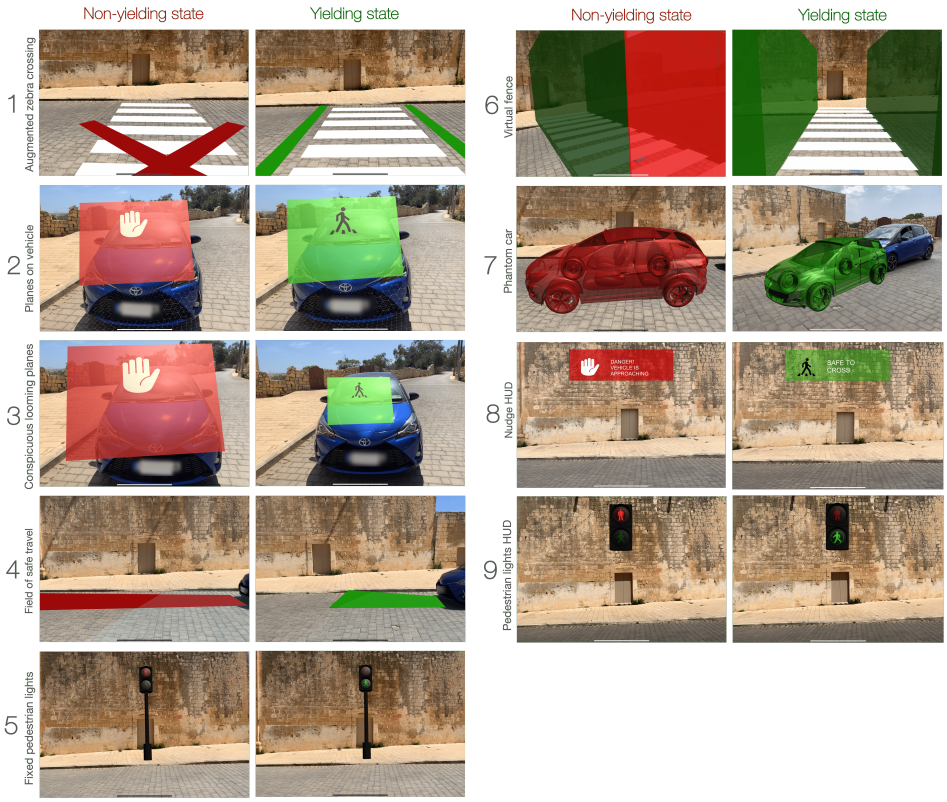


Figure 4.1: The nine AR concepts for pedestrian-vehicle interactions designed and developed by Tabone et al. (2021a). In total, nine AR interface concepts were developed, each with a yielding and non-yielding state: *1. Augmented zebra crossing, 2. Planes on vehicle, 3. Conspicuous looming planes* (i.e., planes which grew or shrank in size), *4. Field of safe travel, 5. Fixed pedestrian traffic lights, 6. Virtual fence, 7. Phantom car* (i.e., a transparent car which indicates the vehicle's predicted future position), *8. Nudge HUD* (i.e., a floating text message and icon which informed the pedestrian whether or not it was safe to cross), *9. Pedestrian traffic lights HUD*. Interfaces 1, 4, 5, 6, and 7 are projected on the road surface, while Interfaces 2 and 3 are projected on the car. Interfaces 8 and 9 are head-locked, i.e., they remain in the user's field of view.

that a high-quality AR interface should be easily understood (intuitive) and encourage people to follow up its recommendations (convincing), and be seen as useful in supporting pedestrian decision-making (usefulness). Furthermore, apart from encouraging performance, whether people like the AR interface (attractiveness, satisfaction) was seen as relevant, as when people might reject/disuse an (otherwise useful) AR interface on aesthetic grounds, it will still fail to be effective.

4.2 METHOD

IN this study, participants viewed videos in a within-subject design, with 9 AR interfaces and 2 yielding behaviours. Participants rated each video according to a number of criteria. The video content, questionnaire design and procedures, and statistical analysis methods are explained below.

4.2.1 VIDEOS

A total of 19 videos (at 30 fps) depicting an approaching AV with a representation of the AR interface in the virtual reality (VR) environment were created (Figure 4.2). More specifically, nine videos depicted a yielding AV featuring a green-coloured (RGB: 32, 244, 0) AR interface, and nine videos depicted a non-yielding AV featuring a red-coloured (RGB: 244, 0, 0) AR interface.

A 19th video was created to depict a non-yielding AV without any interface. The latter was used at the start of the questionnaire to demonstrate how confusing and dangerous a situation without any form of signal would be, especially if the vehicle does not yield, while the other 18 videos were shown to participants in the experiment section of the questionnaire.

The videos were created based on a simulation created in a Unity-built VR environment (Unity, 2022). The road environment was obtained from previous research (e.g., Kaleefathullah et al., 2020) performed in the Highly Immersive Kinematic Experimental Research (HIKER) simulator located at the University of Leeds (University of Leeds, 2022). The videos mimicked the first-person view of a stationary pedestrian considering to cross in front of an approaching vehicle and looking to the right, on a one-way street. A one-way street was selected in order to standardise the direction of traffic flow, considering that the target population of the study were from countries with different traffic systems. Other studies focusing on road crossing have also utilised a one-way street scenario (e.g., Cavallo et al., 2019; Kaleefathullah et al., 2020; Weber et al., 2019).

Trigger points and speeds were adopted from a study on pedestrian crossing in the HIKER simulator (Kaleefathullah et al., 2020). The AV, represented by the same car model in each video, spawned out of sight from the field of view (Figure 4.3, Point A) and moved at a constant speed of 30 mph (48 kph). All interfaces, irrespective of location and state, were triggered when the vehicle reached Point B, located 43 m from the participant (camera) location at Point E. For yielding AVs, the vehicle started decelerating at a rate of 2.99m/s^2 at Point C, which is located 33 m from Point E, and it came to a full stop 3 m from Point E, at Point D. In the case of a non-yielding AV, the vehicle maintained its initial speed of 30 mph throughout.

Each video started with the camera pointing towards the other end of the crossing (Figure 4.4, at time 0 s). The camera then slowly panned to the right as the vehicle approached from

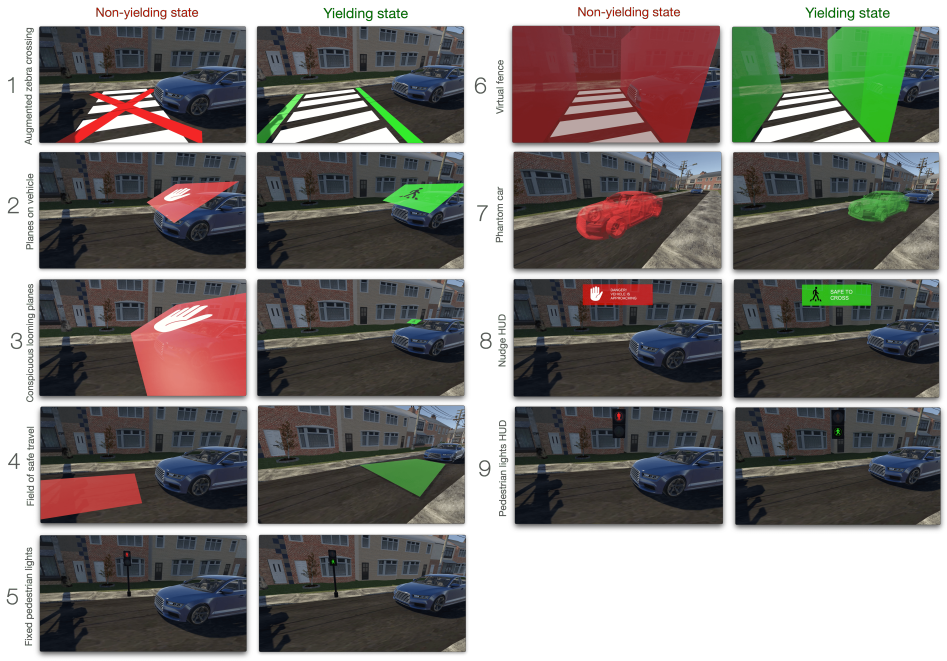


Figure 4.2: The nine AR interfaces presented in a VR environment used for this online questionnaire study. Interfaces 1, 4, 5, 6, and 7 are projected on the road surface, while Interfaces 2 and 3 are projected on the car. Interfaces 8 and 9 are head-locked. The interfaces were adapted from Tabone et al. (2021b).

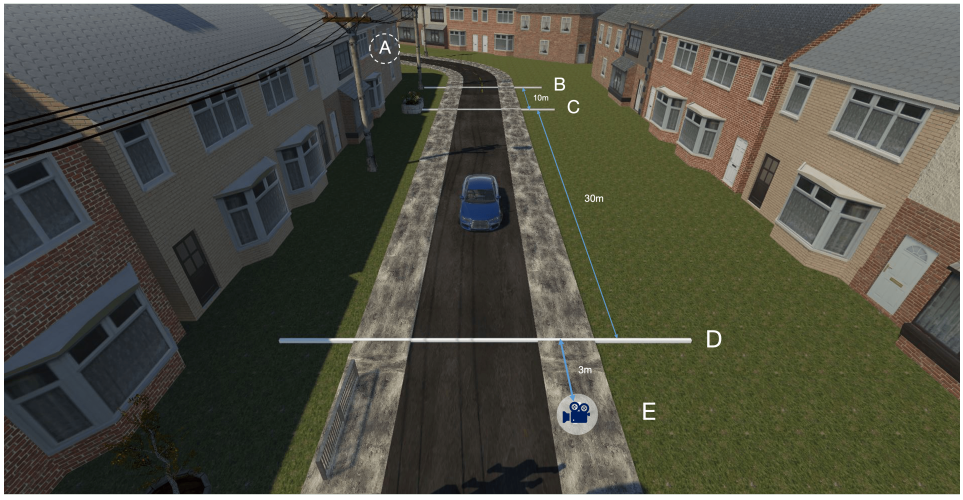


Figure 4.3: Virtual environment used in the videos. Each salient point is demarcated by a label, together with the distance (in metres) between each point. A: spawn point, B: AR interface onset, C: AV deceleration onset, D: stopping point, E: participant location. The participant position is also marked with a camera icon.

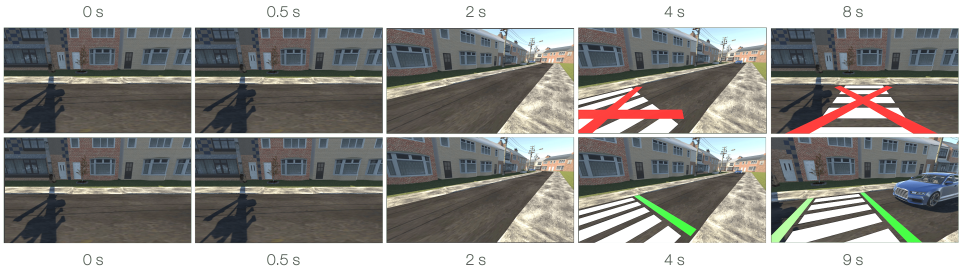


Figure 4.4: Screenshot of the camera view for Augmented Zebra Crossing at key timestamps. The screenshot at the top are for the non-yielding state, while the bottom screenshots correspond to the yielding state.



Figure 4.5: Example of side-by-side image for AR concept 1, Augmented zebra crossing. *Left*: non-yielding state, *Right*: yielding state

point A, starting at an elapsed time of 0.5 s. At an elapsed time of 2 s, the camera would have rotated by an angle of 45°, and the approaching vehicle and AR interface (regardless of type) could be seen simultaneously. At 4 s, the camera started to rotate back to the front-facing position, and it stopped rotating at 20° to the right for the yielding state (elapsed time: 9 s), and fully facing the front for the non-yielding AV (elapsed time: 8 s) so that the vehicle could be observed driving over the crossing area.

In addition to videos, side-by-side images were created per AR interface, for insertion in the questionnaire (see Figure 4.5 for an example). For the yielding AV, the frame where the vehicle came to a complete stop was selected, while for the non-yielding state, the frame at an elapsed time of 6 s was used so that each screenshot had a similar perspective on the road. The only exception was the side-by-side comparison of the phantom car, where the screenshots were taken with respect to the location of the phantom car interface, rather than the actual vehicle, so that both the interface and the vehicle could be seen in the screenshots. The 19 videos produced for the experiment are included in the Supplementary Material.

4.2.2 QUESTIONNAIRE PROCEDURE

The online questionnaire was administered to 1500 respondents from Germany, the Netherlands, Norway, Sweden, and the United Kingdom. These countries were selected based on the geographical locations of the participating partners of the Horizon 2020 SHAPE-IT

project, which funded this research. These five European countries also have a strong research base in automated vehicle development (Hagenzieker et al., 2020) and are likely candidates for the early deployment of eHMIs and AR interfaces. The questionnaire was developed in English using the Qualtrics XM (Qualtrics, 2022) survey platform and distributed to representative Internet panels through the German market research institute INNOFACT AG (Innofact, 2022), which has been used in previous research on the acceptance of AVs (Nordhoff et al., 2021).

A screening questionnaire, prepared in the national language of each of the target respondents' countries, was added by INNOFACT, to control for age, gender, and nationality and filter out respondents who were uncomfortable with completing the questionnaire in English. Our requested target sample was an equal distribution across countries, gender, and split between five (18–29, 30–39, 40–49, 50–59, 60–69) age groups. INNOFACT ensured that participants only participated using a desktop device, and safeguards against bots and duplicate respondents were also taken.

The survey ran from February to April 2022, and the respondents were financially compensated with approximately €3. The study was approved by the Human Research Ethics Committee of the TU Delft under application number 1984.

4.2.3 QUESTIONNAIRE DESIGN

INTRODUCTORY INFORMATION

First, a brief overview of AR and VR technologies was presented, together with examples of popular AR apps, so that the unfamiliar respondents would have a clearer picture of what would be discussed in the rest of the questionnaire. This was followed by an example of what the future could look like with the introduction of AR glasses, a brief introduction to the future urban environment, and the need for communication between AVs and pedestrians. The problem of having no clear signals from the car due to the lack of a driver was demonstrated through the baseline video (i.e., without AR interface) of a non-yielding AV. The respondents were provided with an explanation of the purpose of the study, where the potential of solving the communication issue using AR interfaces would be explored.

CONSENT

Respondents were provided with a consent section, which contained the experimenters' names, contacts, conditions to participate (being 18 years or older), the main purpose of the study, and the approximate length of the questionnaire (30 min). It was also highlighted that there were no risks associated with participation and that the questionnaire was anonymous and voluntary. Respondents were encouraged to close the page if they disagreed. Moreover, a question asking whether the instructions were read and understood was provided (Q1). If 'No' was selected, the questionnaire was terminated.

DEMOGRAPHICS

Next, respondents were asked about their identifying gender (Q2), age (Q3), country of residence (Q4), and their highest level of formal education completed (Q5). Respondents were presented with the Affinity for Technology Interaction (ATI) scale (Franke et al., 2019) to gauge their affinity with technological systems (Q6). Respondents were then asked if they had ever used VR headsets (Q7) and AR apps (Q8), and how willing they would be to use

AR wearables in general (Q9), specifically on the road as a pedestrian (Q10), and for the specific task of assisting pedestrians in crossing a road in front of an AV (Q11).

The respondents were then asked whether they had ever encountered AVs before (Q12), their daily walking time as pedestrians (Q13) (as used in Deb et al., 2017), and their primary mode of transportation (Q14). The last part in the demographic section treated any constraints in personal mobility (Q15) and included a colour blindness test (Q16) (Ishihara, 1917; as used in Bazilinskyy et al., 2020).

VIDEO PRESENTATION OF AR INTERFACES AND RATING QUESTIONS

Following a brief introduction to the experiment, participants proceeded to the main part of the study, where the yielding and non-yielding state of the nine interfaces were presented, together with various rating questions.

The videos from each interface were presented on a separate page, having the title of the respective interface (see Figure 4.2). The order in which the nine interfaces were presented was randomised for each respondent. Each interface page first presented the non-yielding-state video, followed by the yielding-state video. The videos auto-played and looped. All 18 videos were presented to each participant.

Below each video depicting a non-yielding AV, the respondents used a 7-point Likert scale (Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, agree, Strongly agree) to rate whether:

- "The interface in the video above is intuitive for signalling 'Please do NOT cross the road'" (Intuitiveness: Q17)
- "The interface in the video above convinced me NOT to cross the road" (Convincingness: Q18).

and below each video depicting a yielding AV, the following two questions were asked:

- "The interface in the video above is intuitive for signalling 'Please cross the road'" (Intuitiveness: Q19)
- "The interface in the video above convinced me to cross the road" (Convincingness: Q20).

Intuitiveness and convincingness were regarded as two key elements of interface quality, where the former refers to whether the message is readily understandable, and the latter refers to whether the interface would empower people to cross or not cross the road.

The video subsection containing the yielding and non-yielding videos along with the respective intuitiveness and convincingness items was followed by a side-by-side screenshot of the interface's states. A matrix table was presented with a 5-point descriptor scale (Q21) for interpretability, where the respondents had to rate the following:

- "Do you think that the interface was triggered too early or too late?" (too early – too late) (Q21.1)
- "Do you think that the interface is too small or too large? (too small – too large) (Q21.2)

- "How clear (understandable) was the interface to you?" (very unclear – very clear) (Q21.3)
- "How visually attractive is this interface to you?" (very unattractive – very attractive) (Q21.4)

Q17–Q21 were inspired from previous work which looked at perceived quality/clarity of information (Adell, 2010; Bazilinskyy et al., 2020; Lau et al., 2021; Rahman et al., 2017), and attractiveness, aestheticism, ease of understanding, and the adequacy of information, amongst others (Métayer & Coeugnet, 2021).

Each interface page ended with a 9-item acceptance scale (Van Der Laan et al., 1997) to collect further ratings on facets of usefulness and satisfaction (Q22.1–Q22.9). A free text area (Q23) was added to let respondents elaborate on their ratings, "Please add a few words to justify your choices above (eg. comment on the shape, colour, functionality, and the clarity of the interface)."

4.2.4 FINAL QUESTIONS

The final section of the questionnaire opened with a question on whether such AR interfaces would be useful for crossing the road in future traffic (Q24). This query was followed by three side-by-side screenshots contrasting various interface elements, and the following three statements:

- "I prefer interfaces mapped to the street rather than on the vehicle" (Q25)
- "I prefer interfaces with text rather than interfaces with just graphical elements" (Q26)
- "I prefer interfaces that move around with my head rather than interfaces that stay fixed" (Q27), to which the respondent was answered with a 5-point Likert agreement scale from Strongly disagree to Strongly agree.

The penultimate question related to whether the respondent would like to have the ability to customise the interfaces (Q28). The final question once again asked whether the respondent would be willing to use such interfaces as an aid for crossing after having seen all examples, assuming that they own AR glasses (Q29).

4.2.5 ANALYSIS

Mean item scores for the AR interfaces in their yielding and non-yielding states were computed and visualized in scatter plots, together with 95% confidence intervals. The confidence intervals were computed by applying a correction for within-subjects effects of the nine AR interfaces, according to a method presented by Morey (2008).

Differences between ratings of AR interfaces were examined using a repeated-measures ANOVA with an alpha level of 0.05. This was followed by paired-samples t-tests. Here, an alpha value of 0.005 was used to reduce the occurrence of false positives, compared to the more commonly used alpha value of 0.05 (Benjamin et al., 2018). It should be noted that because our sample size was large, even small within-subject differences between the AR interfaces were strongly significant.

For the assessment of the effects of the moderator variables (gender, age group, educational level), a repeated-measures ANOVA was used with the AR interface as a within-subject variable and the moderator variable subgroup (e.g., male, female) as a between-subjects variable (alpha = 0.05). Additionally, statistical comparisons between ratings for AR interfaces between participant groups (e.g., males vs. females) were performed using independent-samples t-tests (alpha = 0.005).

Apart from testing differences between AR interfaces and the effects of moderator variables, Pearson product-moment correlation coefficients among item scores were computed to evaluate redundancy among items. Highly correlated items were aggregated to form a composite score.

The textual responses were evaluated through thematic analysis (Kiger & Varpio, 2020). All text responses were read, with responses copied into a separate document if a common theme emerged. For example, if multiple participants commented that a particular interface was 'slow', then all comments with such a statement were extracted and placed in a text document under the section pertaining to the AR interface. Following the collation of all comments, four comments per interface (two per positive and two per negative valence were selected), depending on which theme was featured the most in that interface's comment section.

4.3 RESULTS

In total, 1500 participants answered the questionnaire. An initial quality filtering process was carried out to remove respondents who did not complete the entire questionnaire (n = 357) or answered 'no' to the consent item (Q1) (n = 39). Next, the recorded duration in seconds was used to omit the top 10% of fastest respondents (i.e., those who completed the questionnaire in 593 s or less, n = 110), since the fastest respondents are likely to yield relatively low-quality data (De Winter & Hancock, 2015). The resulting sample size was 992 (492 males, 491 females, 8 non-binary, 1 not specified). Within the resulting sample, the median time to complete the questionnaire was 23.3 min (25th percentile = 16.4 min, 75th percentile = 33.6 min).

General characteristics of the 992 retained respondents were as follows:

- Country: 202 were from Germany, 197 were from the Netherlands, 184 were from Norway, 197 were from Sweden, and 212 were from the United Kingdom (Q4).
- Age: The age (Q3) ranged from 18 to 69 (M = 45.10, SD = 14.17).
- Education: 54% (n = 536) indicated that they went to university, 25% (n = 246) attended trade or vocational school, whereas 21% (n = 210) indicated 'none of these' (Q5).
- Constraints: 17% (n = 170) reported some form of mobility constraint (Q15).
- Constraints: 3% (n = 32) were considered colour blind as they submitted three or more incorrect answers (Bazilinskyy et al., 2020) for the six-item Ishihara colour blindness test (Q16).

The results regarding AR and VR use indicated the following:

- 42% of respondents had used a VR headset before (Q7).
- 45% had used AR apps before (Q8).
- On a scale of 1 (Strongly unwilling) to 5 (Strongly willing), the mean response to "How willing would you be to use AR glasses?" (Q9) was 3.59 (*SD* = 1.04).
- For "How willing would you be to use AR glasses on the road as a pedestrian" (Q10), the mean was 3.10 (*SD* = 1.13).
- For "How willing would you be to use AR glasses on the road if these warn you about how safe it is to cross in front of a self-driving car?" (Q11), the mean was 3.30 (*SD* = 1.12).

Since the goal of this research was to perform a population-level evaluation of the AR interfaces, colour blind participants or participants with a mobility constraint were not excluded from the analysis.

4.3.1 RATINGS OF VIDEOS DEPICTING AR INTERFACES

Table S1 in the Supplementary material shows the means across the 992 respondents for the 17 items for each of the nine AR interfaces. From this table, it can be seen that there are clear redundancies among the items, with some AR interfaces producing considerably higher ratings than others on almost all of the 17 items.

In an attempt to better understand item redundancy, several correlational analyses were performed. In particular, Figure 4.6 shows the mean intuitiveness ratings (Q17, Q19) and convincingness ratings (Q18, Q20) for the nine AR interfaces. The ratings were very highly correlated ($r = 0.998$), indicating that the intuitiveness and convincingness questions yielded nearly the same information. Figure 4.6 also shows that the *Nudge HUD* scored highest, followed by the *Augmented zebra crossing*, *Fixed pedestrian lights*, *Pedestrian lights HUD*, and *Virtual fence*. The *Phantom car* yielded the lowest ratings.

In the same vein, Figure 4.7 shows the averaged intuitiveness and convincingness rating for the nine AR interfaces for yielding AVs versus non-yielding AVs. Again, a strong association ($r = 0.93$) is seen, indicating that the AR interfaces were rated similarly regardless of whether the vehicle was stopping or not. We performed a two-way repeated-measures ANOVA of the averaged intuitiveness and convincingness rating with AR interface and yielding state as within-subject factors. Results showed a significant effect of the AR interface, $F(8,7928) = 197.4$, $p < 0.001$, partial $\eta^2 = 0.17$, but not of yielding state $F(1, 991) = 0.12$, $p = 0.728$, partial $\eta^2 = 0.00$. There was, however, a significant AR interface × yielding state interaction, $F(8, 7928) = 41.5$, $p < 0.001$, partial $\eta^2 = 0.04$. Follow-up paired-samples t-tests showed that several AR interfaces (i.e., *Augmented zebra crossing, Field of safe travel, Fixed pedestrian lights, Nudge HUD, Pedestrian lights HUD*) yielded somewhat higher ratings for the non-yielding state than for the yielding state ($p < 0.005$ according to paired-samples t-tests). The *Virtual fence* and the *Planes on vehicle*, on the other hand, were rated statistically significantly higher for yielding AVs than for non-yielding AV.

A correlation matrix (Figure 4.8) of the mean ratings for each interface revealed strong associations between all 17 measured items, except for the small/large item (Q21, Item 1) and

early/late item (Q21, Item 2). The correlation coefficients between the means of the 15 other items ranged from $r = 0.862$ (for irritating/likeable [Q22, Item 6] vs. sleep-inducing/raising alertness [Q22, Item 9]) to $r = 0.999$ (unpleasant/pleasant [Q22, Item 2] vs. irritating/likeable [Q22, Item 6]).

4.3.2 DESCRIPTOR SCALE (Q21), ACCEPTANCE SCALE (Q22), AND COMPOSITE SCORE

Because correlation coefficients between items were very high, it was decided to compute a composite score of the 15 strongly-correlated items (unit-weight method, see DiStefano et al., 2009)[1]. More specifically, for each AR interface, a 992 participant x 15 matrix was available. The matrices were concatenated, yielding an 8928 x 15 matrix, and subsequently standardised, so that the item mean was 0 and the standard deviation was 1. The scores of the 15 items were summed, thus producing an 8928-long vector, which was then standardised. Finally, the 8928-long vector was partitioned back to the nine interfaces, so that a composite score was available for each participant and AR interface. Figure 4.8 shows that the mean composite score correlated very strongly with each of its defining items, which confirms that the composite score captures a large amount of the variance (96.5%) in the mean ratings of the nine AR interfaces. The strongest correlations between the composite score and the individual items ($r = 0.997$, 0.998) occurred for the items useful/useless (Q22.1), bad/good (Q22.3), and worthless/assisting (Q22.7) (see Figure 4.1). This suggests that the meaning of the composite score is well described by the colloquial phrase 'whether the AR interface is good or not'.

The mean and standard deviation of the composite score per AR interface are shown in Table 4.1 . The findings align with the above results (Figures 4.6 and 4.7) that the *Nudge HUD* was most favoured while the *Phantom car* was least favoured. A one-way repeated-measures ANOVA of the composite score showed a significant effect of the AR interface, $F(8,7928) = 195.0$, $p < 0.001$, partial $\eta^2 = 0.16$. A total of 32 of 36 pairs of AR interfaces were statistically significantly different from each other ($p < 0.005$), see Table 4.1.

4.3.3 ASSESSMENT OF MODERATOR VARIABLES

Gender: Figure 4.11 in the supplementary material shows a strong correlation ($r = 0.980$) between the mean composite scores for male and female respondents. A repeated-measures ANOVA of the composite score, with the AR interface as a within-subject factor and gender (male or female) as a between-subjects factor showed a significant effect of AR interface, $F(8, 7848) = 192.6$, $p < 0.001$, partial $\eta^2 = 0.16$, and no significant effect of gender, $F(1, 981) = 0.36$, $p = 0.547$, partial $\eta^2 = 0.00$, but a significant AR interface × gender interaction, $F(8, 7848) = 2.00$, $p = 0.043$, partial $\eta^2 = 0.00$. The interaction effect was extremely small, however, and scores for the nine AR interfaces did not differ significantly between males and females. More specifically, independent-samples t-tests for the nine AR interfaces yielded p-values between 0.087 and 0.952 (*Conspicuous looming planes*: Mean (*SD*) males

[1] An inspection of the eigenvalues of the correlation matrix of the (9 AR interfaces × 15 items) matrix showed strong uni-dimensionality. More specifically, the first component explained 96.5% of the variance in the participant means, and the corresponding Cronbach's alpha value was 0.990. Additionally, the correlation matrix at the participant level (992 participants x 15 items) showed strong uni-dimensionality as well, with the first component explaining 67.6% of the variance in the means of the 9 AR interfaces, and Cronbach's alpha being 0.962.

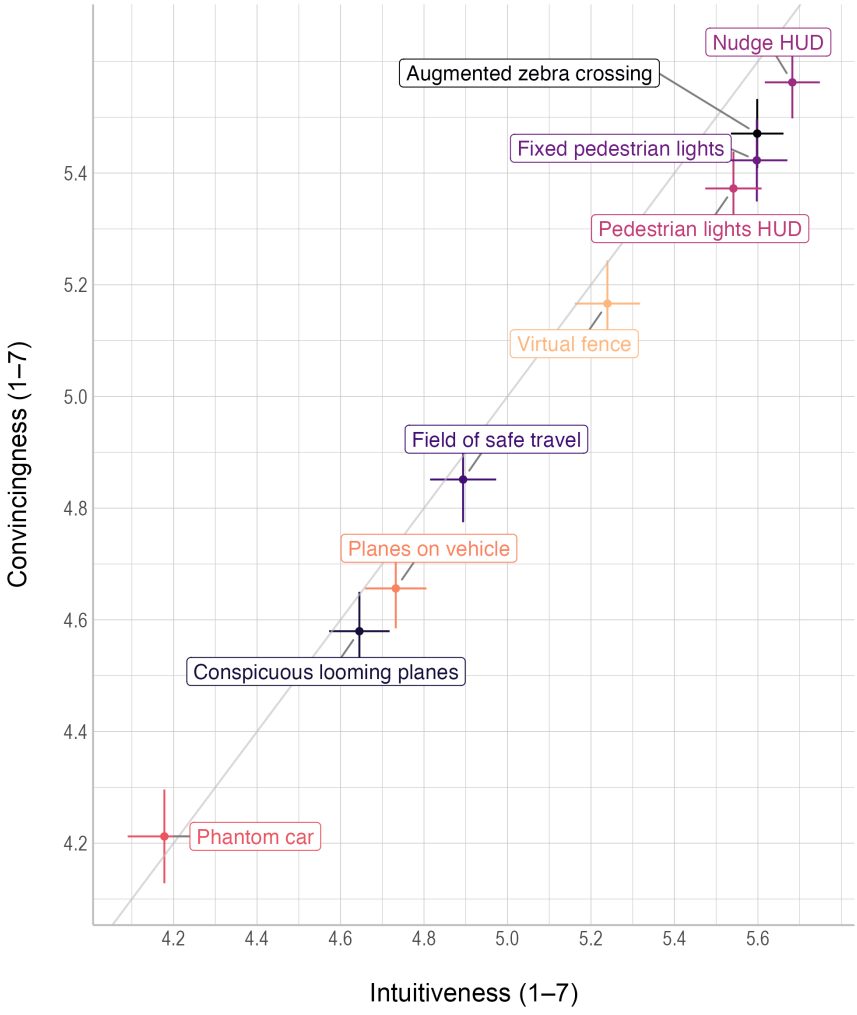


Figure 4.6: Scatter plot of intuitiveness ratings (mean of Q17 and Q19) and convincingness ratings (mean of Q18 and Q20) per AR interface. In this figure, ratings for the yielding and non-yielding states were averaged. The error bars represent 95% confidence intervals.

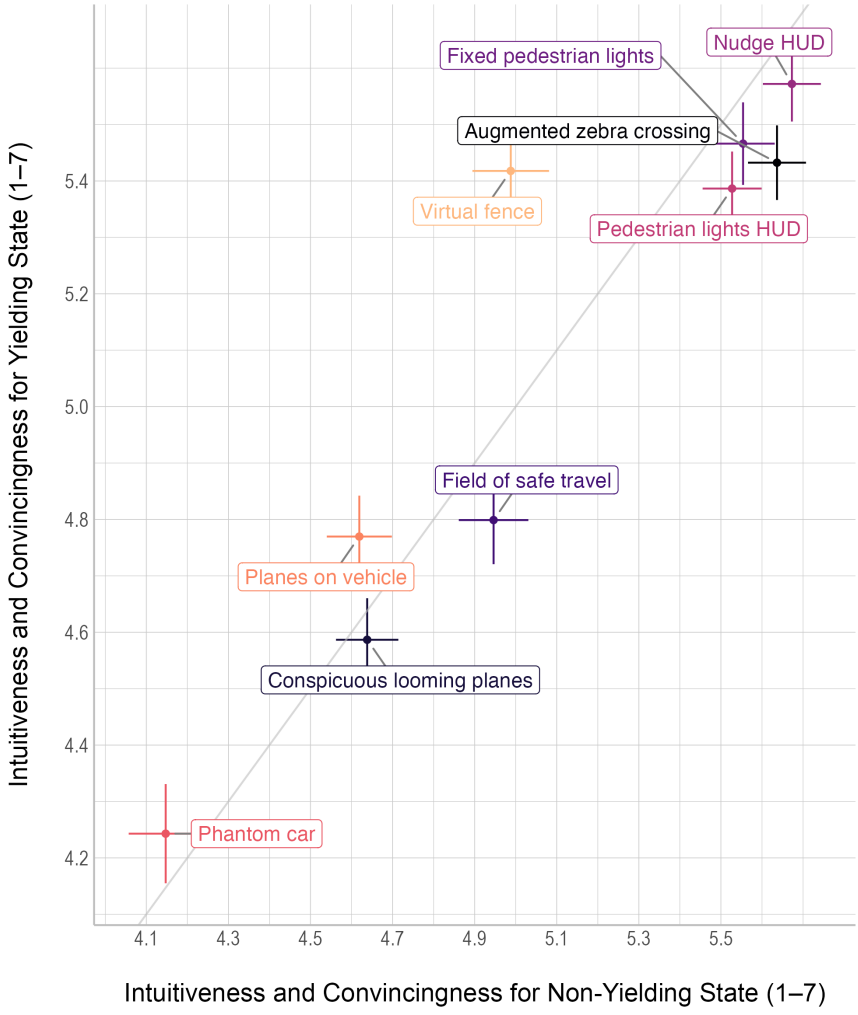


Figure 4.7: Scatter plot of averaged intuitiveness and convincingness ratings of the yielding state (mean of Q19 & Q20) versus the non-yielding state (mean of Q17 & Q18) of each AR interface. The error bars represent 95% confidence intervals.



Figure 4.8: Correlation matrix for the means of the scores of the AR interfaces (n = 9). Responses to Q22 Items 1, 2, 4, 5, 7, 9 were reversed with respect to the questionnaire. The variables are sorted based on hierarchical clustering, i.e., similarity with the other variables.

Table 4.1: Correlation matrix for measures relating to the online questionnaire study (Tabone et al., 2023), and the simulator study conducted in the HIKER

№	AR Interface	Composite Score	1	2	3	4	5	6	7	8	9
1	Augmented zebra crossing	0.32 (0.89)									
2	Planes on vehicle	- 0.26 (1.01)	x								
3	Conspicuous looming planes	- 0.35 (1.00)	x	x							
4	Field of safe travel	- 0.12 (1.00)	x	x							
5	Fixed pedestrian lights	0.28 (0.88)		x	x						
6	Virtual fence	0.04 (1.00)	x	x	x	x	x				
7	Phantom car	- 0.52 (1.05)	x	x	x	x	x	x			
8	Nudge HUD	0.37 (0.85)		x	x	x	x	x	x		
9	Pedestrian lights HUD	0.25 (0.86)		x	x	x		x	x	x	

Note. 'x' marks pairs of conditions that are statistically significantly different from each other, computed using paired-samples t-tests (df = 991).

/ females: -0.40 (1.03) / -0.29 (0.97), $t(981) = -1.71$, $p = 0.087$; *Nudge HUD*: Mean (*SD*) males / females: 0.37 (0.84) / 0.37 (0.85), $t(981) = -0.06$, $p = 0.952$).

Country: The composite score of each interface was examined across the respondents' countries of residence (Figure 4.9). The mean composite scores of the nine AR interfaces correlated again strongly. More specifically, for the 10 pairs of countries, correlations ranged between $r = 0.972$ (between Germany and Sweden) and $r = 0.992$ (between Norway and Sweden). A repeated-measures ANOVA of the composite score, with the AR interface as within-subject factor and country as between-subjects factor showed a significant effect of AR interface, $F(8, 7896) = 194.1$, $p < 0.001$, partial $\eta^2 = 0.16$, and no significant effect of country, $F(4, 987) = 0.82$, $p = 0.515$, partial $\eta^2 = 0.00$, and no significant AR interface × country interaction, $F(32, 7896) = 0.69$, $p = 0.902$, partial $\eta^2 = 0.00$.

Age: A repeated-measures ANOVA of the composite score, with the AR interface as a within-subject factor and age (45 or younger vs. 46 or older) as a between-subjects factor showed a significant effect of AR interface, $F(8, 7920) = 195.2$, $p < 0.001$, partial $\eta^2 = 0.16$, and no significant effect of age group, $F(1, 990) = 0.44$, $p = 0.506$, partial $\eta^2 = 0.00$, and no significant AR interface × age group interaction, $F(8, 7920) = 1.52$, $p = 0.143$, partial $\eta^2 = 0.00$. The corresponding scatter plot is found in the supplementary material (Figure 4.12).

Education: A repeated-measures ANOVA of the composite score, with the AR interface as a within-subject factor and educational attainment (university degree, trade/technical/vocational training, none of these) as a between-subjects factor showed a significant effect of AR interface, $F(8, 7912) = 167.8$, $p < 0.001$, partial $\eta^2 = 0.15$, and no significant effect of education, $F(2, 989) = 0.72$, $p = 0.489$, partial $\eta^2 = 0.00$, and no significant AR interface × education interaction, $F(16, 7912) = 0.98$, $p = 0.476$, partial $\eta^2 = 0.00$. The corresponding

scatter plots are found in the supplementary material (Figures 4.13 and 4.14).

It is noteworthy that although the overall composite score (i.e., averaged across the nine AR interfaces) did not correlate significantly with gender (r = 0.01 [1 = male, 2 = female]), age (r = 0.02), the highest level of education completed (r = 0.04, [1 = university degree, 2 = trade/technical/vocational training, 3 = none of these]), having ever used a VR headset (Q7; r = -0.01 [1 = no, 2 = yes]), or having ever used AR apps or games (Q8; r = 0.02 [1 = no, 2 = yes]), it did correlate moderately with willingness to use AR glasses (r = 0.33, 0.32, and 0.35 for Q9, Q10, and Q11, respectively) and with the ATI scale of technology affinity (Q6; r = 0.22). It is also noteworthy that older participants were less likely to have ever used VR (Q7; r = -0.30) or AR (Q8; r = -0.44, respectively).

4.3.4 TEXTUAL RESPONSES (Q23)

An average of 46 comments were extracted per interface. The subset of comments was further filtered down to retain four informative comments per concept, split equally between positive and negative valence (Table 4.2). These final selected comments were deemed representative of some of the major themes that arose per concept.

Table 4.2: Sample of four comments per interface, split based on positive or negative sentiment. Spelling and grammar mistakes were not corrected.

Interface	Positive Comments	Negative Comments
Augmented zebra crossing	"A good idea. The zebra crossing is familiar to everyone. The big red cross over the crossing should make it clear not to cross."	"It's clear what the images mean, but it doesn't fill me with confidence regarding when it would be safe to cross the road. I think if you are not looking at the approaching vehicle, you will always be in danger because you are not aware as to what it is doing, moving or stopping."
	"Very clear and presumably understandable by most people, including children, once the different colours are explained to them."	"The video signaling 'do not cross the road,' I think is very clear. However, the video signaling that it is safe to cross is not so clear. The green lines on either side of the pedestrian crossing did not immediately make me think it was safe; a green tick symbol may have been better."

Interface	Positive Comments	Negative Comments
Planes on vehicle	"[C]orrect colours for alert and safeness."	"[T]he walking man on the green background made sense, but the hand on the red background was unclear. I didn't like it moving with the car; I would prefer it to be more prominent..."
	"[B]etter variant because the size stays the same and symbols are clearer."	"The problem with this signal is that it just signals something about the car, not about the pedestrians."
Conspicuous looming planes	"Very effective; the colour and hand signal stand out well."	"[T]he colors are still very clear to understand: red for warning and green for no danger. However, as the vehicle approaches from the right side (around the corner), it was difficult to identify the signs written on the colored boxes. It was kind of a weird perspective and therefore irritating. [A]s the stop/go signs were moving with the car and were not fixed at the top of my AR glasses, I had to think twice if these instructions were meant for me as a pedestrian or if there was another issue not concerning me."
	"[T]he warning one was much better than the yielding one as the logo became larger as potential danger increased. [T]he change in size of the yielding one was hardly noticeable."	"I wondered when something would actually appear on the screen. It took forever before I realized the notification was actually on the car itself. I find this visualization absolutely useless."

Interface	Positive Comments	Negative Comments
Field of safe travel	"I think it is somewhat useful as it shows the path of the vehicle."	"[T]he green corridor has me confused. You see the car coming, with a corridor ahead, that makes me think it will drive on instead of stop."
	"There was good warning time to let me know whether I was to cross or not. I also liked how the red and green showed up a good distance off too. Very clear."	"The beam in the 'stop video' looks more like a red carpet, which I guess is something everyone would like to walk on."
Fixed pedestrian lights	"This interface has been familiar and useful to me for as long as I remember. Using it is highly intuitive, and I see no need to alter it."	"I think the sign for triggered too late for the non-yielding state, which would be more of a problem as I might already have started my journey across the street. This can be a risk if the vehicle expects pedestrians to stand still. Otherwise, the sign with a pole is very familiar to me in my cultural context and therefore easily understood."
	"[T]he interface is very clear and understandable, as traffic lights are common in everyday life. It includes people who are not able to read. It seems like a 'no energy' interaction for me as I already know everything I need to know and do not have to think about it."	"The signals are good, but optically too small and might well be overlooked depending on the device holder's age, sight, or the background's distractions."

Interface	Positive Comments	Negative Comments
Virtual fence	"It creates a safe feeling by creating a virtual wall."	"I like the crossing part of this as previously stated, but pairing it with walls is really confusing. When you just see the red one, you immediately think they are walls to stop the car from going through, and it looks like you are being given access through the crossing. The green one is better, but together, it's confusing."
	"Very clear in terms of the obvious color difference but also in the size of the warnings. Very functional!"	"I realized in all examples so far, I enjoy the green signs more. I found this red one being wayyyyy too big, and it literally made me jump when it appeared. It was also not clear to me that it signaled 'do not cross,' except for the red color. When I could compare it with the green sign, which was more intuitive, it was clear that red meant stop. Before that, I saw the red more as a frame/hallway around the zebra crossing."
Phantom car	"[T]he phantom was very fast and clear and really did signal the options I had. It's sustainable as well."	"[I] don't like the look. It reminds me of a video game. So, I guess it can be dangerous because you feel like you're in a game."

Interface	Positive Comments	Negative Comments
	"Really good-looking and easily understandable."	"[T]he trouble is it's just a bit too attractive, and your brain does what it always does when you see something really attractive (particularly cars) – it goes 'WOW!' When it does that, it sort of sucks up all of your attention, and you actually pay less attention to the other car. You almost forget about it."
Nudge HUD	"I liked this one. People are pretty used to something similar to a notification like this, and the color + text makes it even easier to understand it."	"[...] I feel the non-yielding state should specify 'do not cross' as opposed to just stating a vehicle is approaching. The yielding state clearly states 'safe to cross,' so the message is much clearer with no room for misinterpretation."
	"This again empowers the user to make a choice based on their actions, not based on what the car is doing. It is much bigger than some, but in some ways, less distracting. More functional."	"This example is clear enough, but a busy road is not like this. Except for cars, there can be running pets, pedestrians, bicycles coming from behind... It is dangerous to rely on this system, I think."

Interface	Positive Comments	Negative Comments
Pedestrian lights HUD	"The best so far because you get the information in the same direction, so you are looking for incoming traffic. Very nice."	"This is a lot clearer since it already relies on traffic rules that are now established in our society. I still have the feeling though that even if it is green, you would hold back a little bit with crossing the road since the car drives pretty fast towards you, and I would only cross the street if the car is completely still."
	"Because the interface uses an image that I am already acquainted with (as are most members of the general public, including children and senior citizens), I found it to be very effective in indicating to me whether I could or could not cross the road safely."	"The image is clearly recognizable as one which indicates whether or not to cross. My only concern is that it is too small. It actually took me a few seconds to work out where it was. It could, of course, be that in time users would automatically focus on that part of their vision and see the signal, but for this test, I found it worrying."

4.3.5 PREFERRED AR INTERFACES AND USE OF AUGMENTED REALITY IN TRAFFIC

The results of the final questionnaire section (Table 4.3) showed that 66% of respondents felt that communication using AR interfaces in future traffic would be useful (Q24). Furthermore, 72% preferred interfaces mapped to the street over those on the vehicle (Q25), 52% preferred interfaces that included text rather than just graphical elements (Q26), and 51% preferred head-locked over world-locked interfaces (Q27). Moreover, 62% would like to have the ability to customise the AR interfaces (Q28), and 47% indicated they would likely use AR interfaces as an aid for crossing in front of vehicles if they owned AR glasses (Q29).

Table 4.3: Descriptive statistics (i.e., means (M), standard deviations (SD), and relative frequencies) for the final questions.

Question	**M**	**SD**	**Relative Frequencies**				
			Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly Agree (5)
In future traffic, the communication from AR interfaces would be useful for crossing the road (Q24)	3.70	1.02	4.7 %	6.7 %	23.1 %	45.4 %	20.2 %
I prefer interfaces mapped to the street rather than on the vehicle (Q25)	3.98	0.98	1.9 %	6.1 %	19.5 %	37.3 %	35.2 %
I prefer interfaces with text rather than interfaces with just graphical elements (Q26)	3.44	1.09	5.3 %	14.0 %	28.5 %	35.4 %	16.7 %
I prefer interfaces that move with my head rather than interfaces that stay fixed (Q27)	3.38	1.13	7.3 %	14.3 %	27.5 %	35.0 %	15.9 %
I would like to have the ability to customize these AR interfaces (Q28)	3.71	0.95	3.0 %	5.4 %	29.5 %	41.4 %	20.6 %
Now that I have seen these interfaces, if I own AR glasses, I am likely to use such interfaces as an aid for crossing in front of vehicles (Q29)	3.30	1.10	8.9 %	11.6 %	32.9 %	34.3 %	12.3 %

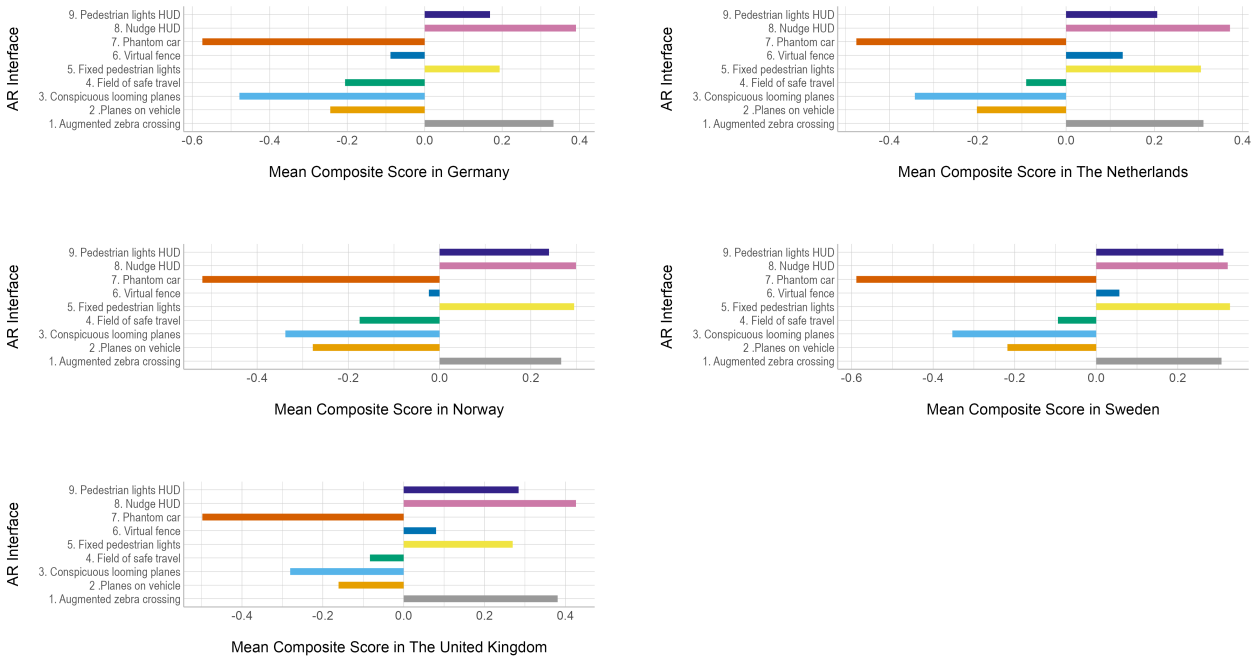


Figure 4.9: Bar plots of the composite score of each interface per respondents' country. The standard deviation across respondents for the 45 depicted AR interface × country combinations ranges between 0.78 and 1.12.

4.4 DICUSSION

An online questionnaire study, aiming to evaluate nine AR interfaces for pedestrian-vehicle interaction, resulted in 992 valid respondents. Respondents were asked to rate the interfaces, presented in videos, on several qualities such as intuitiveness, convincingness, aesthetics, usefulness, and satisfaction.

4.4.1 INTERFACE PREFERENCE BY RESPONDENTS

When considering the intuitiveness and convincingness ratings (Figures 4.6 and 4.7), and the composite score in Table 4.1, it can be asserted that AR interfaces that incorporated traditional traffic elements (*Augmented zebra crossing, Fixed pedestrian lights*, and *Pedestrian lights HUD*) and those that were head-locked performed better than the others. In addition, respondents indicated their preference for head-locked interfaces in the final responses of the questionnaire (Table 4.3).

The 'genius' design approach yielded a number of AR interfaces that were theoretically interesting but flawed from a user's point of view. The findings can retrospectively be explained by legacy design principles, which some AR interfaces adhered to and others did not (see Wickens et al., 2004, for thirteen established principles of display design). For example, although the *Phantom car* was designed to adhere to the principle of predictive aiding (since it showed the future position of the car), and the *Field of safe travel* adhered to the principle of ecological interface design (Kadar & Shaw, 2000; Tabone et al., 2021b; Waldenström, 2011), these two interfaces may have failed to comply with other design principles, such as redundancy gain (these interfaces displayed a coloured element, but no redundant icon or text), the proximity compatibility principle (it may be hard to perceptually separate the *Phantom car* from the real car), and the principle of top-down processing (participants are likely unfamiliar with these concepts). The most successful AR concepts, such as the *Augmented zebra* crossing and *Pedestrian lights* did adhere to the latter three principles, as described by Tabone et al. (2021b). The current observations also highlight the importance of involving the target user earlier on in the process through the use of a user-centred design methodology (Gulliksen et al., 2003) and to not rely on genius design only. The involvement of the target user early in the process could be achieved through focus groups, interviews, and card sorting, among other methods (Norman, 2013).

On the technical side, it is to be noted that the different AR interfaces involve different sensor and computational requirements (for an overview, see Tabone et al., 2021b). For instance, AR interfaces presented on the AV itself would have to rely on computer-vision techniques on the pedestrian's side, or vehicle-to-pedestrian communication of the AV's position and speed. The nudge interfaces, however, are considerably simpler and would only require the wireless communication of the AV's stopping intent to the pedestrian. These different sensor requirements were not presented to the respondents, nor were they considered in the evaluation of the AR interfaces.

Additionally, our study found that the means of questionnaire items were very strongly correlated and that the 15 acceptance-related items, in the aggregate, were well-represented by a single composite score. A recommendation that follows is that future research into the population-level mean acceptance of HMI concepts could just as well use a single acceptance item (such as a five-point scale ranging from bad to good) instead of multiple acceptance-related items. This finding aligns with previous research on the acceptance of

automated driving systems, which indicated that different acceptance dimensions are hardly distinguishable and that a single factor of acceptance provides a better representation of the data (De Winter & Nordhoff, 2022; Nees & Zhang, 2020).

There were, however, two items that did not correlate strongly with the composite score, namely items related to the physical parameters of interface size and timing. As shown in Table S1 in the Supplementary Material, all nine AR interfaces yielded equivalent ratings (between 2.91 and 3.12) on the scale from 1 (too early) to 5 (too late) (Q21, Item 1), which can be explained by the fact that all interfaces were triggered at the same moment in the video. The small differences may have been caused by proximity (e.g., *Field of safe travel* extends in front of the car, "a sort of tongue protruding forward along the road"; Gibson & Crooks, 1938, p. 454), which might give participants the illusion that the interface was triggered early. The size ratings (Q21, Item 2) were also close to the midpoint for the nine interfaces, i.e., between 2.56 for the *Pedestrian lights HUD* and 3.37 for the *Virtual fence*. The differences in perceived size can also be explained by the actual size of the interfaces (see Figure 4.2).

In the aggregate, different groups of participants reached similar conclusions on what they deemed to be a 'good' interface, i.e., results were similar regardless of gender, age, or country. Anecdotally, it is often believed that there are major cultural differences among pedestrians in that an eHMI that is found to work well in one country may not be received well in another country (see quotes of Bärgman, Hagenzieker, Krems and Ackerman, and Stanton in Tabone et al., 2021a). The results of the present study suggest that these cultural differences are less strong as may be believed, at least for the five European countries under investigation. Our findings mirror those of others (Bazilinskyy et al., 2019; Singer et al., 2022) who found cross-cultural robustness of eHMIs in a larger number of countries on different continents.

While the online questionnaire was generally well distributed across the set quotas, it should be noted that the represented countries of residence were exclusively Western and Northern European. Therefore, cultural differences may have been relatively small. Several studies reported differences between the perceived clarity of eHMIs among participants from China versus Western Europe (Joisten et al., 2021; Lanzer et al., 2020; Weber et al., 2019). Whether or not cultural differences become apparent may depend on the clarity of the task instructions in the experiment and participants' prior expectations rather than the eHMI content itself, as noted by Singer et al. (2022).

4.4.2 FREE-TEXT COMMENTS

The textual inputs and opinions of the respondents were varied. Some respondents reported that the interfaces on the road surface could distract pedestrians from approaching vehicles, while others considered interfaces on the vehicle a hazard, since these blocked the visibility of the oncoming vehicle. In a number of instances, respondents indicated that they preferred the non-yielding state over the yielding state, with the former being regarded as more clear and intuitive. In fact, the intuitiveness and convincingness, as shown in Figure 4.7, tended to favor the non-yielding state, except for a number of interfaces (*Planes on vehicle, Virtual fence*). Respondents described the yielding state for *Virtual fence* as clearer, but some labelled the non-yielding state as dangerous because the presence of a zebra crossing might tempt pedestrians to cross irrespective of the red gate. Similarly, the non-yielding state for

the *Field of safe travel* was labelled as potentially dangerous by some because it looked like a red carpet that invited them to walk on it.

Another prevalent theme was that some respondents felt that at times it was not clear to whom the communication referred, i.e., the pedestrian or the vehicle itself. For example, the hand symbol on the *Planes on vehicle* was sometimes misinterpreted as indicating a problem with the vehicle. The *Planes on vehicle* and *Conspicuous looming planes* interfaces, which project planes on the vehicle, drew concerns about a blocked view of the vehicle, yet at the same time, the looming planes concept was commended for its clarity in communicating danger. These observations reveal the issue of unintended effects resulting from 'genius designs', where the intention is not fully grasped by the user. Our findings resonate with broader issues in human factors, namely that *"the actual, rather than presumed, impact of new technology is usually quite surprising, unintended, and even counterproductive"* (Woods & Dekker, 2000, p. 276).

Similar to the observations derived from the statistical analysis, the interfaces based on more traditional traffic elements were labelled as more understandable and intuitive due to familiar symbology (e.g., zebra crossing, traffic light). The 'worst' performing interface (*Phantom car*), while commended for its aesthetic qualities, received various critical descriptions, such as 'confusing', 'frightening', 'scary', 'startling', 'spooky', and 'unclear'. In fact, some described the interface as a video game, which in a sense confirms the original design direction of the *Phantom car* concept from Tabone et al. (2021b), where the idea of ghost cars from racing video games was drawn upon.

The HUD interfaces were praised for being 'logical', 'visible', 'clear', and 'perfect' to capture the attention of distracted pedestrians. However, it was also stated that HUDs could be a distraction from other hazards, especially when text is used (for further discussion on text-based eHMIs, see Bazilinskyy et al., 2019). Moreover, a number of respondents complained that the text was in English, and that this would be a danger for pedestrians unfamiliar with the language. The latter feedback resonates with an advantage of AR communication, where personalization of the interface could solve the language issue. In fact, 62% of the respondents were in favour of such a possibility. Finally, a number of times, respondents suggested that they would still rely on the vehicle coming to a full stop before making any decision, confirming that implicit communication plays an important role in shaping pedestrian decisions (Lee et al., 2021).

4.4.3 LIMITATIONS AND FUTURE WORK

Although the online questionnaire was distributed to a wide respondent pool, the analysis revealed that more than half of the respondents (54%) reported having attained a university degree. Research suggests that university graduates are more inclined towards the adoption and usage of technology (Burton-Jones & Hubona, 2005; Nielsen & Haustein, 2018). At the same time, we found strong convergence in ratings for participants with and without a university degree, suggesting that educational level was not an important moderator of the current findings (see Figures 4.13 and 4.14). A possible reason is that participants were not asked to understand or use complex technology; instead, the present task was largely one of perceptual nature.

A further limitation is that the high correlation of acceptance-related items may have arisen from the uniform questionnaire format, giving rise to acquiescence bias. However,

this limitation may not be significant as the acceptance scale (Q22) contained reversed items (from high to low, and from low to high), yet these items still correlated very strongly with the responses to the intuitiveness and convincingness items.

A number of free-text comments mentioned drivers being blinded by the interfaces that appeared on the car, indicating that those respondents did not fully grasp what AR technology is. Additionally, there were instances where the terms 'AR' and 'VR' were used interchangeably in the comments, with a number of respondents expressing total opposition towards wearing 'VR headsets' when they walk around outside. This confusion may have been caused by the fact that participants only saw VR videos of AR concepts, rather than experiencing AR themselves. That said, such confusion would only have affected the overall understanding of AR, and probably not the relative differences in the participants' assessments of the nine AR concepts.

Many respondent comments were unusable in the thematic analysis. While gibberish text entries were uncommon, many of the textual comments were too brief to provide useful information (e.g., "This one was clear"). This highlights a limitation of online studies, where there is the risk that some respondents do not thoroughly read the information provided at the beginning or aim to complete the questionnaire items quickly. A further limitation of online studies with videos is that, while offering high repeatability, they do not offer high ecological validity and present only low perceived risk to participants (for a similar discussion, see Fuest et al., 2020; Petzoldt et al., 2018; Tabone et al., 2021a).

A further limitation was that the environment consisted of a one-way road, with only one vehicle. The addition of more traffic, with varying trajectories, would add more natural cues to the testing environment. It can be hypothesized that the *Nudge HUD* will be particularly effective when multiple vehicles approach from different directions since the *Nudge HUD* does not require the pedestrian to distribute attention across those vehicles. In contrast, the *Planes* on vehicle require the pedestrian to first locate the planes in the environment before crossing, which may be time-consuming and inefficient. A potential advantage of *Planes on vehicle*, on the other hand, is that it may prevent overreliance in situations of e.g., vehicle-to-pedestrian communication failure. Another limitation was the lack of environmental sound, and the fact that participants were not asked to interact with the scene (e.g., to indicate when it is safe to cross). To better understand the behaviour of users of such interfaces, ecological validity must be increased. Therefore, in the future, the stimuli could be presented to the participants in a virtual simulation environment and ultimately, in the real world.

4.5 CONCLUSION

NINE augmented reality interfaces for pedestrian-vehicle interaction were presented in a video-based online study that yielded 992 respondents from Germany, the Netherlands, Norway, Sweden, and the United Kingdom. Each interface was shown in its non-yielding and yielding states at a pedestrian crossing area represented in a VR environment. Respondents were asked to rate each interface based on its intuitiveness and convincingness in communicating whether or not a vehicle would yield. Other ratings related to functional and aesthetic qualities, usefulness, and satisfaction.

Statistical and qualitative thematic analysis indicated that respondents preferred head-locked interfaces over their world-locked counterparts, with interfaces employing traditional traffic interface elements receiving higher ratings than others. These results indicated that

legacy design principles performed better than designs generated through an expert-based approach ('genius' design), further highlighting the importance of involving the user early in the process. A further qualitative analysis provided more context to the ratings, such as the preference of the non-yielding state over the yielding state for a number of interfaces, preference towards traditional traffic symbols, and reliance on implicit cues.

Responses related to the general use of interfaces indicated a preference for interfaces that are mapped to the street instead of the vehicle. Moreover, respondents preferred interfaces that make use of text compared to interfaces that use just graphical elements, and interfaces that are head-locked rather than world-locked. Most respondents also indicated that they would like to personalise the AR interfaces, and that communication using AR interfaces in future traffic would be useful.

Although the current online study offered an indication of what kinds of AR interfaces, placement in the world, and design elements are more suitable for pedestrian-vehicle interactions, there are limitations related to the ecological validity dimension of the study. In order to better understand the behaviour of potential users of the system, in the future, the ecological validity of such a user evaluation should be increased.

The practical implications of the present study depend on the progression in vehicle automation and communication, and in AR. It seems plausible that computers will become increasingly compact, and that the use of AR, either via handheld or head-mounted devices will become increasingly feasible in the real world. At the same time, questions about inclusivity, affordability, and user acceptance remain to be addressed, as discussed by Tabone et al. (2021a). A likely way forward is that the use of AR for pedestrians will see its introduction first in professional transportation contexts (e.g., warehouses, airport personnel) before becoming available to the general public.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860410.

REFERENCES

Ackermans, S., Dey, D., Ruijten, P., Cuijpers, R. H., & Pfleging, B. (2020). The effects of explicit intention communication, conspicuous sensors, and pedestrian attitude in interactions with automated vehicles. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI. https://doi.org/10.1145/3313831.3376197

Adell, E. (2010). Acceptance of driver support systems. Proceedings of the European Conference on Human Centred Design for Intelligent Transport Systems, Berlin, Germany, 475–486.

Bai, S., Legge, D. D., Young, A., Bao, S., & Zhou, F. (2021). Investigating external interaction modality and design between automated vehicles and pedestrians at crossings. Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, 1691–1696. https://doi.org/10.1109/ITSC48978.2021.9564867

Bazilinskyy, P., Dodou, D., & De Winter, J. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. Transportation Research Part F: Traffic Psychology and Behaviour, 67, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2020). External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'? IEEE International Conference on Systems, Man and Cybernetics (SMC), 3721–3728, Toronto, Canada. https://doi.org/10.1109/SMC42975.2020.9282998

Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2021). How should external Human-Machine Interfaces behave? Examining the effects of colour, position, message, activation distance, vehicle yielding, and visual distraction among 1,434 participants. Applied Ergonomics, 95, 103450. https://doi.org/10.1016/j.apergo.2021.103450

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., ... Johnson, V. E. (2018). Redefine statistical significance. Nature Human Behaviour, 2, 6–10. https://doi.org/10.1038/s41562-017-0189-z

Bhorkar, G. (2017). A survey of augmented reality navigation. arXiv. https://arxiv.org/abs/1708.05006

Bindschädel, J., Krems, I., & Kiesel, A. (2022). Active vehicle pitch motion for communication in automated driving. Transportation Research Part F: Traffic Psychology and Behaviour, 87, 279–294. https://doi.org/10.1016/j.trf.2022.04.011

Buker, T. J., Vincenzi, D. A., & Deaton, J. E. (2012). The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: Reducing apparent latency with predictive compensation. Human Factors, 54, 235–249.

https://doi.org/10.1177/0018720811428734

Burton-Jones, A., & Hubona, G. S. (2005). Individual differences and usage behavior: revisiting a technology acceptance model assumption. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 36, 58–77. https://doi.org/10.1145/1066149.1066155

Calvi, A., D'Amico, F., Ferrante, C., & Ciampoli, L. B. (2020). Effectiveness of augmented reality warnings on driving behaviour whilst approaching pedestrian crossings: A driving simulator study. Accident Analysis & Prevention, 147, 105760. https://doi.org/10.1016/j.aap.2020.105760

Cavallo, V., Dommes, A., Dang, N. T., & Vienne, F. (2019). A street-crossing simulator for studying and training pedestrians. Transportation research part F: traffic psychology and behaviour, 61, 217-228. https://doi.org/10.1016/j.trf.2017.04.012

Carmona, J., Guindel, C., Garcia, F., & De la Escalera, A. (2021). eHMI: Review and guidelines for deployment on autonomous vehicles. Sensors, 21, 2912. https://doi.org/10.3390/s21092912

Colley, M., Eder, B., Rixen, J. O., & Rukzio, E. (2021). Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445351

Currano, R., Park, S. Y., Moore, D. J., Lyons, K., & Sirkin, D. (2021). Little road driving hud: Heads-up display complexity influences drivers' perceptions of automated vehicles. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan. https://doi.org/10.1145/3411764.3445575

Dancu, A., Vechev, V., Ünlüer, A. A., Nilson, S., Nygren, O., Eliasson, S., Barjonet, J.-E., Marshall, J., & Fjeld, M. (2015). Gesture bike: Examining projection surfaces and turn signal systems for urban cycling. Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, Madeira, Portugal, 151–159. https://doi.org/10.1145/2817721.2817748

Deb, S., Strawderman, L., Carruth, D. W., DuBien, J., Smith, B., & Garrison, T. M. (2017). Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. Transportation Research Part C: Emerging Technologies, 84, 178–195. https://doi.org/10.1016/j.trc.2017.08.029

De Winter, J., & Dodou, D. (2022). External human-machine interfaces: Gimmick or necessity. Transportation Research Interdisciplinary Perspectives, 15, 100643. https://doi.org/10.1016/j.trip.2022.100643

De Winter, J., & Nordhoff, S. (2022). Acceptance of conditionally automated cars: Just one factor? Transportation Research Interdisciplinary Perspectives, 15, 100645. https://doi.org/10.1016/j.trip.2022.100645

De Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE), Las Vegas, NV, 5334–5341. https://doi.org/10.1016/j.promfg.2015.07.641

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., & Terken, J. (2020a). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives, 7, 100174. https://doi.org/10.1016/j.trip.2020.100174

Dey, D., Habibovic, A., Pfleging, B., Martens, M., & Terken, J. (2020b). Color and animation preferences for a light band eHMI in interactions between automated vehicles and pedestrians. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI. https://doi.org/10.1145/3313831.3376325

Dey, D., & Terken, J. (2017). Pedestrian interaction with vehicles: roles of explicit and implicit communication. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 109–113. https://doi.org/10.1145/3122986.3123009

Dietrich, A., Tondera, M., & Bengler, K. (2020). Automated vehicles in urban traffic: The effect of kinematics and eHMIs on pedestrian crossing behavior. Advances in Transportation Studies, 2020, 73–84.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. Practical Assessment, Research, and Evaluation, 14, 20. https://doi.org/10.7275/da8t-4g52

Dong, W., Wu, Y., Qin, T., Bian, X., Zhao, Y., He, Y., Xu, Y., & Yu, C. (2021). What is the difference between augmented reality and 2D navigation electronic maps in pedestrian wayfinding? Cartography and Geographic Information Science, 48, 225–240. https://doi.org/10.1080/15230406.2021.1871646

Endsley, T. C., Sprehn, K. A., Brill, R. M., Ryan, K. J., Vincent, E. C., & Martin, J. M. (2017). Augmented reality design heuristics: Designing for dynamic interactions. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61, 2100–2104. https://doi.org/10.1177/1541931213602007

Faas, S. M., Mathis, L.-A., & Baumann, M. (2020). External HMI for self-driving vehicles: which information shall be displayed? Transportation Research Part F: Traffic Psychol-

ogy and Behaviour, 68, 171–186. https://doi.org/10.1016/j.trf.2019.12.009

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. International Journal of Human–Computer Interaction, 35, 456–467. https://doi.org/10.1080/10447318.2018.1456150

Fuest, T., Michalowski, L., Träris, L., Bellem, H., & Bengler, K. (2018). Using the driving behavior of an automated vehicle to communicate intentions – a Wizard of Oz study. Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, 3596–3601. https://doi.org/10.1109/ITSC.2018.8569486

Fuest, T., Schmidt, E., & Bengler, K. (2020). Comparison of methods to evaluate the influence of an automated vehicle's driving behavior on pedestrians: Wizard of Oz, virtual reality, and video. Information, 11, 291. https://doi.org/10.3390/info11060291

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. The American Journal of Psychology, 51, 453–471. https://doi.org/10.2307/1416145

Ginters, E. (2019). Augmented reality use for cycling quality improvement. Procedia Computer Science, 149, 167–176. https://doi.org/10.1016/j.procs.2019.01.120

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. Behaviour & Information Technology, 22, 397–409. https://doi.org/10.1080/01449290310001624329

Hagenzieker, M., Boersma, R., Velasco, P. N., Ozturker, M., Zubin, I., & Heikoop, D. (2020). Automated buses in Europe: An inventory of pilots (Technical Report). Delft University of Technology.

Hensch, A.-C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2019). Effects of a light-based communication approach as an external HMI for Automated Vehicles — a Wizard-of-Oz study. Transactions on Transport Sciences, 10, 18–32. https://doi.org/10.5507/tots.2019.012

Hesenius, M., Börsting, I., Meyer, O., & Gruhn, V. (2018). Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Barcelona, Spain, 261–268. https://doi.org/10.1145/3236112.3236148

Innofact. (2022). INNOFACT AG Das MarktVORSPRUNGSinstitut. https://www.innofact.com

Ishihara, S. (1917). Tests for colour-blindness. Tokyo: Kanehara Shuppan Co., Ltd.

Joisten, P., Liu, Z., Theobald, N., Webler, A., & Abendroth, B. (2021). Communication of automated vehicles and pedestrian groups: An intercultural study on pedestrians' street crossing decisions. Proceedings of MuC '21: Mensch und Computer 2021, Ingolstadt, Germany, 49–53. https://doi.org/10.1145/3473856.3474004

Kadar, E. E., & Shaw, R. E. (2000). Toward an ecological field theory of perceptual control of locomotion. Ecological Psychology, 12, 141–180. https://doi.org/10.1207/S15326969ECO1202_02

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & De Winter, J. C. F. (2020). External human–machine interfaces can be misleading: An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human Factors. https://doi.org/10.1177/0018720820970751

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. Medical Teacher, 42, 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Kim, H., Gabbard, J. L., Anon, A. M., & Misu, T. (2018). Driver behavior and performance with augmented reality pedestrian collision warning: An outdoor user study. IEEE Transactions on Visualization and Computer Graphics, 24, 1515–1524. https://doi.org/10.1109/TVCG.2018.2793680

Lanzer, M., Babel, F., Yan, F., Zhang, B., You, F., Wang, J., & Baumann, M. (2020). Designing communication strategies of autonomous vehicles with pedestrians: an intercultural study. Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Virtual Event, 122–131. https://doi.org/10.1145/3409120.3410653

Lau, M., Le, D. H., & Oehl, M. (2021). Design of external human-machine interfaces for different automated vehicle types for the interaction with pedestrians on a shared space. In: N. L. Black, W. P. Neumann, & I. Noy (Eds.), Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021) (pp. 710–717). Cham: Springer. https://doi.org/10.1007/978-3-030-74608-7_87

Lee, Y. M., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C., Camara, F., Rothmueller, M., Vendelbo-Larsen, S. A., Rasmussen, P. H., Dietrich, A., Nathanael, D., Portouli, V., Schieben, A., & Merat, N. (2021). Road users rarely use explicit communication when interacting in today's traffic: Implications for automated vehicles. Cognition, Technology & Work, 23, 367–380. https://doi.org/10.1007/s10111-020-00635-y

Li, Y., Kamalasanan, V., Batista, M., & Sester, M. (2022). Improving pedestrian priority via grouping and virtual lanes. arXiv. https://arxiv.org/abs/2205.08783

Löcken, A., Golling, C., & Riener, A. (2019). How should automated vehicles interact with pedestrians? A comparative analysis of interaction concepts in virtual reality. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 262–274. https://doi.org/10.1145/3342197.3344544

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing pedestrian behavior in augmented reality—proof of concept. Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, 313–321. https://doi.org/10.1109/VR46266.2020.00051

Matviienko, A., Müller, F., Schön, D., Seesemann, P., Günther, S., & Mühlhäuser, M. (2022). BikeAR: Understanding cyclists' crossing decision-making at uncontrolled intersections using Augmented Reality. Proceedings of the CHI Conference on Human Factors in Computing Systems, New Orleans, LA. https://doi.org/10.1145/3491102.3517560

Métayer, N., & Coeugnet, S. (2021). Improving the experience in the pedestrian's interaction with an autonomous vehicle: An ergonomic comparison of external HMI. Applied Ergonomics, 96, 103478. https://doi.org/10.1016/j.apergo.2021.103478

Mok, C. S., Bazilinskyy, P., & De Winter, J. C. F. (2022). Stopping by looking: A driver-pedestrian interaction study in a coupled simulator using head-mounted displays with eye-tracking. Applied Ergonomics, 105, 103825. https://doi.org/10.1016/j.apergo.2022.103825

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Nees, M., & Zhang, J. (2020). Acceptance of highly automated vehicles: A factor analysis approach. PsyArXiv. https://doi.org/10.31234/osf.io/9qcjp

Nielsen, J. (2007). The myth of the genius designer. https://www.nngroup.com/articles/the-myth-of-the-genius-designer

Nielsen, T. A. S., & Haustein, S. (2018). On sceptics and enthusiasts: What are the expectations towards self-driving cars? Transport Policy, 66, 49–55. https://doi.org/10.1016/j.tranpol.2018.03.004

Nordhoff, S., Beuster, A., Kessel, T., Bjorvatn, A., Innamaa, S., Lehtonen, E., Malin, F., Madigan, R., Lee, Y.-M., Merat, N., & Louw, T. (2021). Annual quantitative survey about user acceptance towards ADAS and vehicle automation (Deliverable D7.1 of L3Pilot project). European Commission. https://l3pilot.eu/downloads#c81

Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

Onkhar, V., Bazilinskyy, P., Stapel, J. C. J., Dodou, D., Gavrila, D., & De Winter, J. C. F. (2021). Towards the detection of driver–pedestrian eye contact. Pervasive and Mobile Computing, 76, 101455. https://doi.org/10.1016/j.pmcj.2021.101455

Oudshoorn, M., De Winter, J. C. F., Bazilinskyy, P., & Dodou, D. (2021). Bio-inspired intent communication for automated vehicles. Transportation Research Part F: Traffic Psychology and Behaviour, 80, 127–140. https://doi.org/10.1016/j.trf.2021.03.021

Petzoldt, T., Schleinitz, K., & Banse, R. (2018). Potential safety effects of a frontal brake light for motor vehicles. IET Intelligent Transport Systems, 12, 449–453. https://doi.org/10.1049/iet-its.2017.0321

Pichen, J., Yan, F., & Baumann, M. (2020). Towards a cooperative driver-vehicle interface: Enhancing drivers' perception of cyclists through Augmented Reality. Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, 1827–1832. https://doi.org/10.1109/IV47402.2020.9304621

Pompigna, A., & Mauro, R. (2022). Smart roads: A state of the art of highways innovations in the Smart Age. Engineering Science and Technology, an International Journal, 25, 100986. https://doi.org/10.1016/j.jestch.2021.04.005

Pratticò, F. G., Lamberti, F., Cannavò, A., Morra, L., & Montuschi, P. (2021). Comparing state-of-the-art and emerging Augmented Reality interfaces for autonomous vehicle-to-pedestrian communication. IEEE Transactions on Vehicular Technology, 70, 1157–1168. https://doi.org/10.1109/TVT.2021.3054312 Qualtrics. (2022). Qualtrics XM // The leading experience management software. https://www.qualtrics.com

Rahman, M. M., Lesch, M. F., Horrey, W. J., & Strawderman, L. (2017). Assessing the utility of TAM, TPB, and UTAUT for advanced driver assistance systems. Accident Analysis & Prevention, 108, 361–373. https://doi.org/10.1016/j.aap.2017.09.011

Ranieri, M. (2020). A new sense of direction with Live View. https://blog.google/products/maps/new-sense-direction-live-view

Rasouli, A., & Tsotsos, J. K. (2020). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems, 21, 900–918. https://doi.org/10.1109/TITS.2019.2901817

Rolland, J. P., Gibson, W., & Ariely, D. (1995). Towards quantifying depth and size perception in virtual environments. Presence: Teleoperators and Virtual Environments, 4, 24–49. https://doi.org/10.1162/pres.1995.4.1.24

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. Frontiers in Psychology, 10, 2757. https://doi.org/10.3389/fpsyg.2019.02757 Saffer, D. (2010). Designing for interaction: Creating innovative applications and devices. Berkeley, CA: New Riders.

Schieben, A., Wilbrink, M., Kettwich, C., Dodiya, J., Weber, F., Sorokin, L., Lee, Y.-M., Madigan, R., Markkula, G., Merat, N., Dietrich, A., & Kaup, M. (2019). Testing external HMI designs for automated vehicles–An overview on user study results from the EU project interACT. 9. Tagung Automatisiertes Fahren.

Singer, T., Kobbert, J., Zandi, B., & Khanh, T. Q. (2022). Displaying the driving state of automated vehicles to other road users: An international, virtual reality-based study as a first step for the harmonized regulations of novel signaling devices. IEEE Transactions on Intelligent Transportation Systems, 23, 2904–2918. https://doi.org/10.1109/TITS.2020.3032777

Sripada, A., Bazilinskyy, P., & De Winter, J. (2021). Automated vehicles that communicate implicitly: Examining the use of lateral position within the lane. Ergonomics, 64, 1416–1428. https://doi.org/10.1080/00140139.2021.1925353

Tabone, W., De Winter, J. C. F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B., & Stanton, N. A. (2021a). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transportation Research Interdisciplinary Perspectives, 9, 100293. https://doi.org/10.1016/j.trip.2020.100293

Tabone, W., Lee, Y. M., Merat, N., Happee, R., & De Winter, J. C. F. (2021b). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Leeds, UK, 209–218. https://doi.org/10.1145/3409118.3475149

Toh, C. K., Sanguesa, J. A., Cano, J. C., & Martinez, F. J. (2020). Advances in smart roads for future smart cities. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 476, 20190439. https://doi.org/10.1098/rspa.2019.0439

Tong, Y., & Jia, B. (2019). An Augmented-Reality-based warning interface for pedestrians: User interface design and evaluation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63, 1834–1838. https://doi.org/10.1177/1071181319631413

Tran, T. T. M., Parker, C., Wang, Y., & Tomitsch, M. (2022). Designing wearable augmented reality concepts to support scalability in autonomous vehicle–pedestrian interaction. Frontiers in Computer Science, 4, 866516. https://doi.org/10.3389/fcomp.2022.866516

Unity. (2022). Unity real-time development platform | 3D, 2D VR & AR engine. https://www.unity.com

University of Leeds. (2022). Highly Immersive Kinematic Experimental Research (HIKER) pedestrian lab. https://uolds.leeds.ac.uk/facility/hiker-lab

Uttley, J., Lee, Y. M., Madigan, R., & Merat, N. (2020). Road user interactions in a shared space setting: Priority and communication in a UK car park. Transportation Research Part F: Traffic Psychology and Behaviour, 72, 32–46. https://doi.org/10.1016/j.trf.2020.05.004

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies, 5, 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Volkswagen. (2020). From the luxury class to the compact segment: the augmented reality head-up display. https://www.volkswagen-newsroom.com/en/press-releases/from-the-luxury-class-to-the-compact-segment-the-augmented-reality-head-up-display-6730

Von Sawitzky, T., Wintersberger, P., Löcken, A., Frison, A. K., & Riener, A. (2020). Augmentation concepts with HUDs for cyclists to improve road safety in shared spaces. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI. https://doi.org/10.1145/3334480.3383022

Waldenström, C. (2011). Visualizing the field of safe travel increases performance in a naval movement task. Proceedings of the 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Miami Beach, FL, 252–256. https://doi.org/10.1109/COGSIMA.2011.5753454

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. Vision Research, 35, 2731–2736. https://doi.org/10.1016/0042-6989(95)00018-U

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). Crossing the street across the globe: A study on the effects of eHMI on pedestrians in the US, Germany and China. In H. Krömker (Ed.), HCI in Mobility, Transport, and Automotive Systems. HCII 2019 (pp. 515–530). Cham: Springer. https://doi.org/10.1007/978-3-030-22666-4_37

Wickens, C. D., Gordon, S. E., & Liu, Y. (2004). An introduction to human factors engineering. Upper Saddle River, NJ: Pearson Prentice Hall.

Woods, D., & Dekker, S. (2000). Anticipating the effects of technological change: a new era of dynamics for human factors. Theoretical Issues in Ergonomics Science, 1, 272–282.

https://doi.org/10.1080/14639220110037452

APPENDIX

The 19 videos, raw data, and a PDF version of the questionnaire are available on a repository (https://doi.org/10.4121/21603678) to facilitate reproducibility and encourage further development.

		1. Augmented zebra crossing	2. Planes on vehicle	3. Conspicuous looming planes	4. Field of safe travel	5. Fixed pedestrian lights	6. Virtual fence	7. Phantom car	8. Nudge HUD	9. Pedestrian lights HUD
Q17	Intuitiveness, Non-yielding	5.67	4.61	4.64	4.95	5.62	5.00	4.07	5.68	5.59
Q18	Convincingness, Non-yielding	5.60	4.63	4.64	4.94	5.48	4.97	4.23	5.67	5.47
Q19	Intuitiveness, Yielding	5.53	4.85	4.65	4.84	5.57	5.48	4.29	5.69	5.50
Q20	Convincingness, Yielding	5.34	4.69	4.52	4.76	5.36	5.36	4.20	5.46	5.28
Q21, 1	Too early - Too late	3.03	3.02	3.06	2.91	3.12	2.99	3.10	2.94	2.99
Q21, 2	Too small - Too large	3.05	2.95	2.88	3.17	2.60	3.37	3.13	2.92	2.56
Q21, 3	Very unclear - Very clear	3.77	3.26	3.17	3.33	3.79	3.60	2.92	3.91	3.77
Q21, 4	Very unattractive - Very attractive	3.52	2.91	2.86	3.16	3.47	3.20	2.85	3.55	3.43
Q22, 1	Useless - Useful	3.84	3.25	3.17	3.38	3.81	3.61	2.96	3.89	3.79
Q22, 2	Unpleasant - Pleasant	3.70	3.15	3.04	3.32	3.69	3.28	2.97	3.69	3.66
Q22, 3	Bad - Good	3.79	3.23	3.10	3.32	3.71	3.44	2.97	3.79	3.69
Q22, 4	Annoying - Nice	3.62	3.11	3.01	3.23	3.59	3.24	2.99	3.63	3.53
Q22, 5	Superfluous - Effective	3.79	3.28	3.17	3.35	3.67	3.57	3.04	3.79	3.66
Q22, 6	Irritating - Likeable	3.62	3.11	3.01	3.26	3.64	3.24	2.94	3.64	3.59
Q22, 7	Worthless - Assisting	3.78	3.27	3.17	3.39	3.72	3.56	3.01	3.80	3.68
Q22, 8	Undesirable - Desirable	3.58	3.10	3.04	3.22	3.55	3.29	2.92	3.59	3.50
Q22, 9	Sleep-inducing - Raising Alertness	3.72	3.38	3.31	3.49	3.59	3.66	3.27	3.76	3.58

Figure 4.10: Table S1. Means for the 17 questions asked for each AR interface. Note: Q17–Q20 were measured on a scale of 1 = Strongly disagree to 7 = Strongly agree. Q21 and Q22 were measured on five-point scales. Colour coding is applied for Q17–Q20 and Q21 & Q22 separately, where the lowest value is red, the median is white, and the highest value is blue.

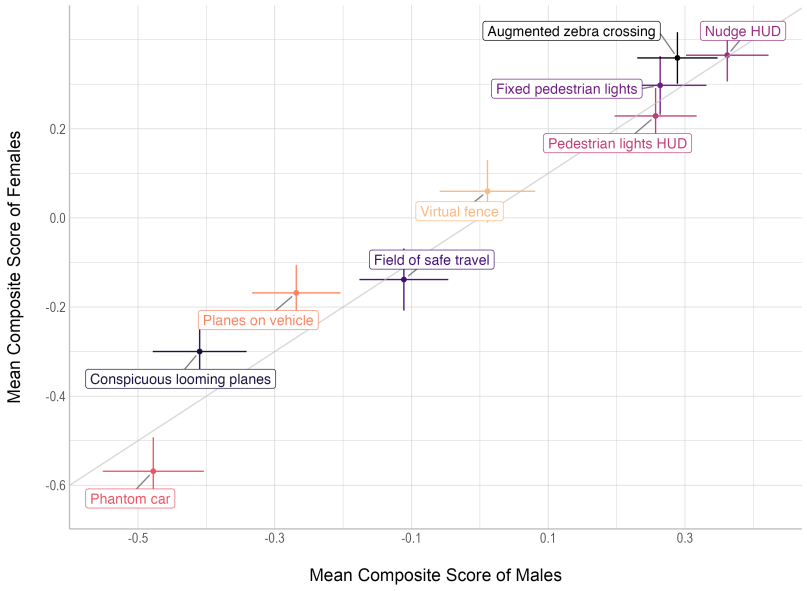


Figure 4.11: Scatter plot of the composite score for the nine AR interfaces, for females (n = 491) versus males (n = 492). The error bars represent 95% confidence intervals.

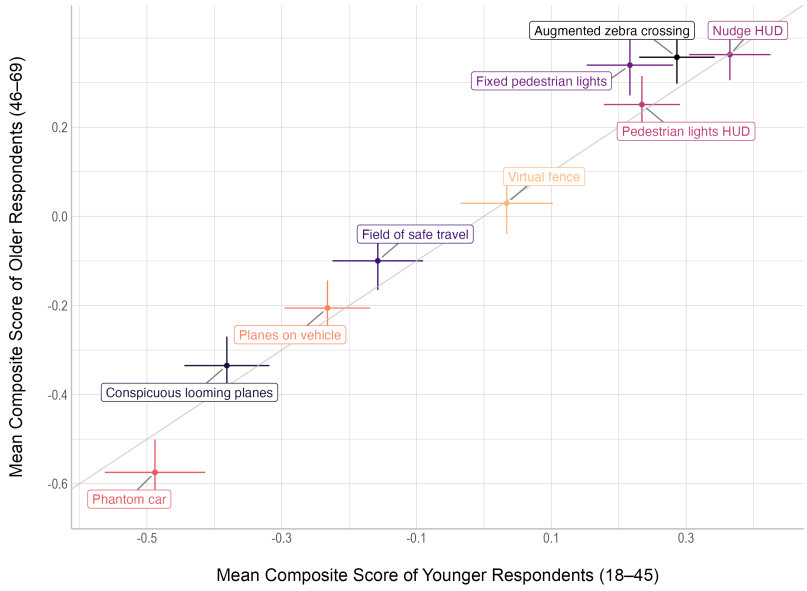


Figure 4.12: Scatter plot of the composite score for the nine AR interfaces, for older respondents (n = 493) versus younger respondents (n = 499) (r = 0.988). The error bars represent 95% confidence intervals.

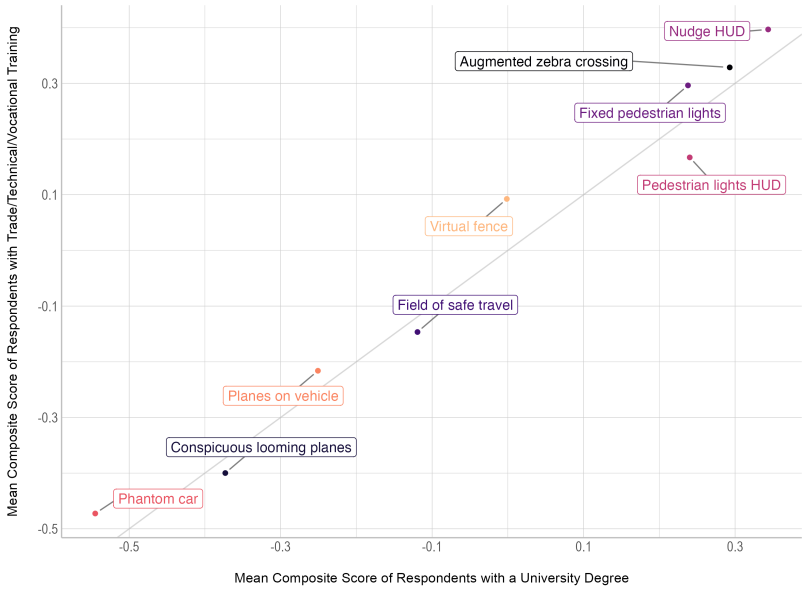


Figure 4.13: Scatter plot of the composite score for the nine AR interfaces, for respondents with trade/technical/vocational training (n = 246) versus respondents with a university degree (n = 536) (r = 0.986). The error bars represent 95% confidence intervals.

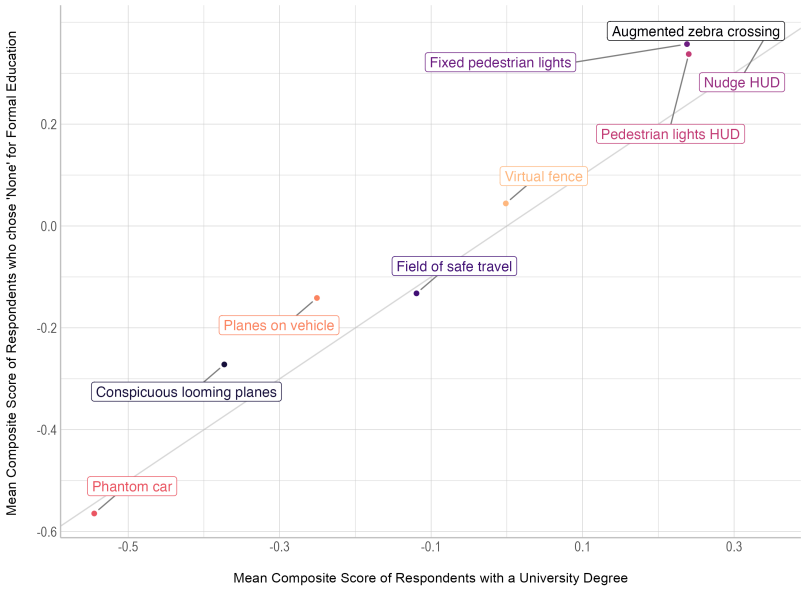


Figure 4.14: Scatter plot of the composite score for the nine AR interfaces, for respondents who indicated ‘none of these’ for the choice of trade/technical/vocational training or university degree (n = 210) versus respondents with a university degree (n = 536) (r = 0.990). The error bars represent 95% confidence intervals.

5

Using ChatGPT for Human–Computer Interaction Research: A Primer

This chapter was published as: *Tabone, W.*, & De Winter, J. C. F. (2023). Using ChatGPT for Human–Computer Interaction Research: A Primer. Royal Society Open Science, 10, 231053.

ABSTRACT

ChatGPT could serve as a tool for text analysis within the field of Human-Computer Interaction, though its validity requires investigation. This study applied ChatGPT to: (1) textbox questionnaire responses on nine augmented-reality interfaces, (2) interview data from participants who experienced these interfaces in a virtual simulator, and (3) transcribed think-aloud data of participants who viewed a real painting and its replica. Using a hierarchical approach, ChatGPT produced scores or summaries of text batches, which were then aggregated. Results showed that (1) ChatGPT generated sentiment scores of the interfaces that correlated extremely strongly ($r > 0.99$) with human rating scale outcomes and with a rule-based sentiment analysis method (criterion validity). Additionally, (2) by inputting automatically-transcribed interviews to ChatGPT, it provided meaningful meta-summaries of the qualities of the interfaces (face validity). One meta-summary analysed in depth was found to have substantial but imperfect overlap with a content analysis conducted by an independent researcher (criterion validity). Finally, (3) ChatGPT's summary of the think-aloud data highlighted subtle differences between the real painting and the replica (face validity), a distinction corresponding with a keyword analysis (criterion validity). In conclusion, our research indicates that, with appropriate precautions, ChatGPT can be used as a valid tool for analysing text data.

5.1 INTRODUCTION

OPENAI'S ChatGPT has taken the research world by storm, garnering media attention and sparking debates. Released in November 2022, ChatGPT has shown great promise in various areas, such as interpreting computer code (Sobania et al., 2023; Tate et al., 2023), generating prompts for art-generating tools (Pavlik, 2023), document writing and translation (Tate et al., 2023), text interpretation and review (Zhang & Simeone, 2022; Zhong et al., 2023), generating creatively written text such as poetry (Kirmani, 2023), writing hospital discharge summaries (Patel & Lam, 2023), and as an assistive tool in education (Baidoo-Anu & Owusu Ansah, 2023; Bommarito & Katz, 2022; Gao et al., 2022), amongst many other applications. Although ChatGPT offers many opportunities (Bubeck et al., 2023), others have highlighted several risks, including the potential for plagiarism in academic writing (González-Padilla, 2023; Krukar & Dalton, 2020; Thorp, 2023) and education (De Winter, 2023; Gilson et al., 2022; Rudolph et al., 2023; Stoker-Walker, 2022), generation of incorrect computer code (Vincent, 2022) or incorrect mathematical solutions (Frieder et al., 2023), and the possibility of inaccurate or biassed output (Alba, 2022; Borji, 2023; Council of the European Union, 2023; Van Dis et al., 2023).

In order to better appreciate its functioning, it is useful to elucidate how ChatGPT was created (Karpathy, 2023). ChatGPT is a machine learning model that has been designed and fine-tuned in several stages. The first is a pre-training stage where multiple terabytes of text data, which includes a general crawl of the internet, including sources like Github, Wikipedia, and others, have been tokenised into numerical sequences. These sequences are processed through a transformer neural network, a model architecture proposed by Vaswani et al. (2017) that uses self-attention mechanisms to understand the contextual relationship between words in a sentence. The transformer analyses the entire sequence of tokens at once rather than sequentially, making it effective in modelling complex patterns in the data.

During the pre-training phase, the decrease in loss function corresponds to an improvement in the prediction accuracy of the subsequent word, a procedure demanding significant computational resources involving 1000s of GPUs over a period of months. The eventual base model effectively operates as a 'next-word predictor' and has a refined understanding of diverse concepts and textual structures. However, to function as a useful AI assistant, further refinements are implemented. In the supervised fine-tuning stage, human-created 'ideal' (i.e., helpful, truthful, harmless) responses to prompts are used to refine the model. Then, reinforcement learning from human feedback (RLHF) is applied. In the reward modelling stage, human workers rank multiple generated responses to prompts, to train the model to rank response quality. Finally, reinforcement learning is applied to the previous stage's reward model to refine the assistant model. The final product, ChatGPT, is a chat interface designed to function as a user-friendly assistant. An API is available as well, allowing researchers and practitioners to connect it to programming software and to set a number of parameters. One such parameter is the temperature, which controls the level of randomness in the output.

Despite its advanced training, GPT fundamentally differs from human information processing. ChatGPTm an autoregressive model, sequentially generates each new token predicated on the previously generated tokens within the same sequence (Zhou et al., 2023). It does not possess introspective capabilities, meaning it cannot self-correct or perform sanity checks on its outputs (Long, 2003). Furthermore, ChatGPT lacks deep reasoning abilities, and it cannot use tools like humans do, such as accessing web information or calculators (Amer-Yahia et al., 2023; Bubeck et al., 2023; Frieder et al., 2023). Several strategies can enhance ChatGPT's performance, such as prompt engineering to encourage more thoughtful responses by spreading reasoning across multiple prompts ("chain of thought prompting", Kim et al., 2023; Wei et al., 2022). Methods such as self-consistency (Wang et al., 2022), try-again (Yang et al., 2023), and backtracking (Long, 2023) enable the model to review its output, and its capabilities can be enhanced by connecting it to tools like online apps or calculators (Eulerich et al., 2023; Karpathy, 2023; OpenAI, 2023b; Peng et al., 2023).

Considering the immense popularity of the ChatGPT web interface and API, there is a need for further research to ascertain how it can be used for scholarly purposes and to evaluate its validity in such contexts. In this work, we explored whether ChatGPT can be applied as a valid tool in Human–Computer Interaction (HCI) research. HCI is the field that focuses on the design and use of computer technology and examines interfaces between people and computers. HCI research may involve various types of research methods, including questionnaires, interviews, psychophysics methods, virtual reality setups, and field tests. In this process, the HCI researcher often collects text data. For example, in questionnaire research, a number of free-response items may be included to allow the respondents to reflect on a specific type of HCI design, while in interviews, verbal data is collected that are later transcribed and analysed (Gerlach & Kuo, 1991; Gubrium & Holstein, 2001; Maraj et al., 2016; Schelble et al., 2022), and in field or VR studies, a think-aloud protocol is often used to enrich the data collection (Clemmensen & Roese, 2010; Kjeldskov & Skov, 2003; Zhang & Simeone, 2022; Zhao & McDonald, 2010). The analysis of text-based data has faced criticism due to its perceived lack of reproducibility (Humphreys & Wang, 2018; Indulska et al., 2012; Kitto et al., 2023). For example, despite the frequent use and acclaim for the thematic analysis method (Braun & Clarke, 2006; De Carvalho & Fabiano, 2021; Kiger &

Varpio, 2020) in revealing underlying themes in texts, the consistency of themes identified by independent researchers analysing the same text, such as interview transcripts, remains a contentious issue (DeCuir-Gunby et al., 2011; Roberts et al., 2019).

This study posed the research question: To what extent does ChatGPT produce valid sentiment scores and summaries when applied to different forms of text data in HCI research? Our analysis was based on questionnaire responses, interviews, and think-aloud data available from three previous studies (De Winter et al., 2022; Tabone et al., 2023a, 2023b). Our objective was to determine if ChatGPT's outputs could uphold both face validity and criterion validity. Here, face validity refers to plausibility—the extent to which ChatGPT appears, on face value, to accurately measure what it is supposed to. Conversely, criterion validity in our context refers to the extent to which the outputs of ChatGPT correlate with other established measures, namely automated or human-generated scores or counts. We conducted three distinct studies, using either the ChatGPT API (introduced on March 1, 2023) or the online website. Each study applied ChatGPT to different types of data (questionnaire data, interview data, and transcribed think-aloud data) to test its validity.

5

5.2 STUDY 1: QUESTIONNAIRE TEXTBOX DATA

In an online study by Tabone et al. (2023a), 992 respondents rated nine new augmented reality (AR) interfaces for pedestrian-vehicle interaction. Each interface was presented in two animated video clips depicting a crossing situation in a virtual environment, where a single vehicle approached from the right. The interface was presented in a yielding or non-yielding state, with the former communicating that the approaching vehicle would stop for the pedestrian and the latter communicating the opposite. Figure 5.1 provides still frames from the videos as an example. As part of the questionnaire, respondents had to complete several rating scales related to the interface's intuitiveness and convincingness in communicating the message to cross or not to cross the road. Rating scales regarding acceptance, attractiveness, aesthetics, ease of understanding, and the adequacy of information were also presented. A free-response item was added at the end, allowing respondents to elaborate on their ratings: *"Please add a few words to justify your choices above (e.g., comment on the shape, colour, functionality, and the clarity of the interface)"*.

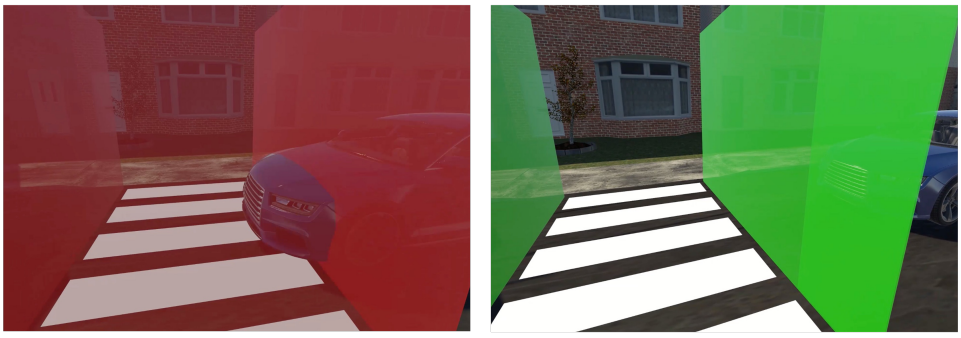


Figure 5.1: Frames from the videos presented in the online questionnaire. In this instance, the 'Virtual fence' (interface number 6) interface is seen in its non-yielding (left) and yielding states (right) (Tabone et al., 2023a).

In the original study, the 15 quantitative ratings for intuitiveness, convincingness etc. were statistically analysed and averaged into a composite score, which was calibrated to have a grand mean of 0 and a standard deviation of 1. The meaning of the composite score can be characterised as "whether the AR interface is good or not" (Tabone et al., 2023a).

With regard to the statistical reliability of the composite scores, Tabone et al. (2023a) reported that the Cronbach alpha was 0.990, as calculated across the matrix of 15 mean scores of the rating scales x 9 AR interfaces. We also determined the statistical reliability of the composite score using the split-half method. To this end, we repeatedly divided the 15 rating scale items into two groups of 8 and 7 random items, and calculated the composite score over these two sets separately. The two vectors of composite scores (9 × 1) thus obtained correlated strongly, on average, $r = 0.995$. Finally, we determined the composite score reliability by repeatedly splitting the group of 992 participants into two groups of 496 randomly selected participants. The two vectors of composite scores (9 × 1) that are thus obtained also correlated strongly on average, $r = 0.992$.

In the original study, the text-based responses were analysed through a manual review process where representative answers were selected. From the average 46 text comments selected per interface, 4 representative comments (2 with positive sentiments and 2 with negative sentiments) were chosen for inclusion in the publication by Tabone et al. (2023a). A limitation of this thematic analysis was the potential influence of the researcher's subjectivity on the analysis. Furthermore, the manual text analysis that we conducted had limitations in the sense that we could not easily compare or interpret the results in relation to other findings. More specifically, it was difficult to assess whether the results from our thematic analysis were somehow linked to the calculated composite score.

5.2.1 METHODS

In the new analysis, all text responses were extracted from the raw questionnaire data, available from the online supplementary material of Tabone et al. (2023a). This yielded 9 columns (1 for each interface) and 992 rows, for a total of 8,928 entries. The data was unaltered, and the responses were submitted to the ChatGPT API through a custom script (date: 4 March 2023, model: gpt-3.5-turbo). The parameter governing randomness, referred to as 'temperature', was set to a value of 0. This configuration may be advantageous in tasks such as sentiment analysis, where the absence of random variation is preferred. This deterministic attribute of the API contrasts with the web interface of ChatGPT, which features some randomness in its output.

The raw data submitted to ChatGPT included not only meaningful statements, but also contained meaningless ones, staccato text, and text in different languages (see Table 5.1 for a random selection of 20 comments for one of the nine interfaces). We have chosen to retain these occasionally meaningless quotes to evaluate ChatGPT's sentiment analysis in a standalone way, without human intervention. By not filtering the text, the sentiment analysis for ChatGPT becomes more challenging compared to a situation where a human researcher might preprocess the data; it falls to ChatGPT's model itself to handle gibberish text in an appropriate manner.

Since there are limits to the amount of text (tokens) that could be submitted at any one time, it was decided to adopt a 'hierarchical approach' to data processing, where batches of text were submitted, and the numeric outputs were subsequently averaged. Specifically, each

Table 5.1: Selection of comments about the 'Virtual fence' interface (20 of 992 comments were selected using a random number generator).

Interesting and very good.
The would be fun to test in the future
I have no idea
Really clear, but a bit too much
Full colour very clear
Easy to see with the bright colours
Best one so far...
Not so great graphic quality
To big sign
I have no comment
On det kommer någon gående då stannar man eftersom det är det det betyder.
Extremely clear way to show when it is safe and unsafe to cross
Big, bold and very clear
Bruce Springsteen is a communist
Want to see by myself if it's safe
Felt more intuitive than the second option but I still prefer the first option
The interface is too big
It is hard to trust.
I think the fencing is a good way to signal whether or not it's safe to cross the street. However, the non-yielding fence isn't as effective at getting the message across, apart from the colour, as the pedestrian crossing is still open. I think it could be confusing, especially to someone who's colourblind. Maybe a big X to cross out the pedestrian crossing/obscure the path would be more effective here.
Niet van toepassing
It was a bit confusing to show a crossing if you were not supposed to cross. Even though it was red I still saw the crossing and thought it was maybe ok.

column was automatically divided into nine batches of 100 rows each, and a tenth batch consisting of 92 rows. Each batch was then separately submitted to the API, with a specific prompt:

"The comments below were obtained from respondents in an online survey about an AR interface in a road-crossing scenario. If you read the comments below, what is the quality of the interface on a scale from 1 = bad to 10 = good? Only report a number between 1 and 10. Round to two decimals", followed by the comments from each batch.

After all the batches were submitted, the mean score for each interface was calculated by averaging the sentiment scores of its 10 batches. The correlation coefficient was computed between the mean sentiment score of the interfaces and the composite score of the interfaces as previously reported by Tabone et al. (2023a).

We opted to evaluate the correlation with the rating scales because these scales, which had been gathered from the respondents, provide a clear numerical criterion, enabling us to quantitatively assess the validity of ChatGPT's sentiment analysis. In the free-response item, participants were asked to elaborate on their previous answers to the rating scales. We therefore anticipate that if ChatGPT can produce a more meaningful sentiment score from the participants' text responses, there will be a stronger correlation (i.e., closer to 1.00) with the composite score. The entire process used to calculate the mean sentiment score is detailed in the flowchart presented in Figure 5.2.

Next, we examined whether the prompt could be improved by trial and error. More specifically, we repeated the above process using different prompts, in an attempt to maximise the correlation coefficient between the mean sentiment score obtained through ChatGPT and the mean composite score (Tabone et al., 2023a).

A potential limitation of the above analysis is the calculation of correlations between the ChatGPT sentiment score based on textual responses and a numerical composite score grounded in rating scale responses. As previously indicated, we assume that a higher correlation coefficient is indicative of higher criterion validity. However, this supposition relies on the assumption that both the rating scales and textual responses have accurately captured the sentiment of the AR interface. Since this is not certain, we have substituted the composite score based on the rating scales with an alternative criterion, where we applied a traditional sentiment analysis method to the texts. To accomplish this, we first tokenised all the statements of the respondents for each AR interface. Subsequently, we subjected the tokens, for the 992 participants and 9 AR interfaces separately, to the VADER (Valence Aware Dictionary and sEntiment Reasoner) algorithm (Hutto & Gilbert, 2014). VADER is a lexicon and rule-based sentiment analysis tool that quantifies the sentiment of text by assigning polarity scores to words. It generates a compound sentiment score ranging from -1 (most negative) to +1 (most positive), taking into account factors like negations, punctuation, capitalisation, and emoticons. We then computed an average VADER sentiment score per AR interface by averaging over the 992 values, and correlated this VADER sentiment score with the ChatGPT sentiment score, which was based on the same text.

One limitation of ChatGPT we discovered is its peculiar preference for certain discrete values. These discrete outputs inherently have a negative effect on the sensitivity of the sentiment scores, especially when the number of batches is small. We also found that the order of the texts within a given batch had an impact on the output. These findings led us to realise that a refined method of sentiment analysis could be found in bootstrapping. To this end, for each of the nine AR interfaces, we drew a large number (200) of batches of 100 texts, randomly selected from the 992[1]. We submitted these batches in the same manner as

[1]The number of outputs over which the sentiment score was calculated varied between 189 and 200, due to

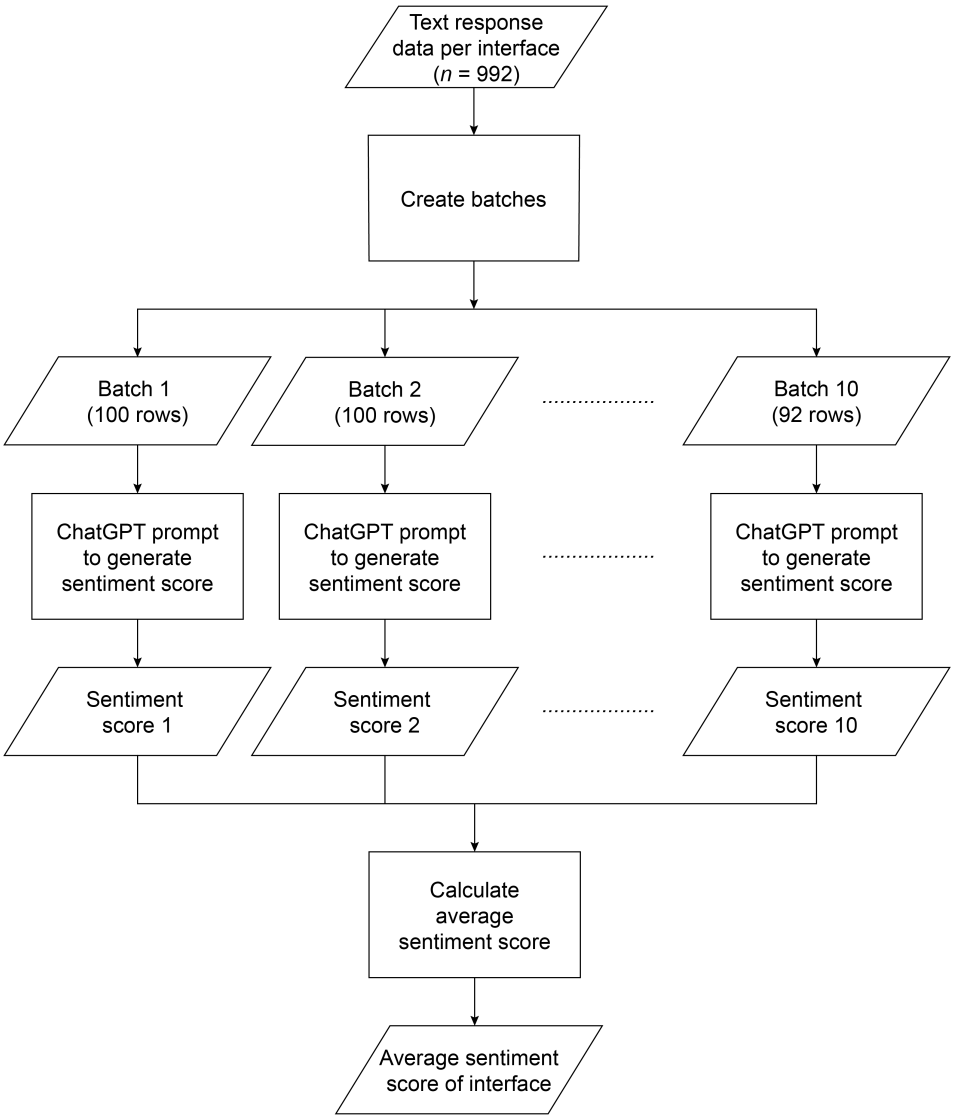


Figure 5.2: Flowchart detailing the process of the hierarchical approach used to calculate the average ChatGPT sentiment score per interface. The entire process was conducted 9 times (once for each interface).

above, and then took an average sentiment score over the outputs. We repeated this process for different prompts ranging from extensive prompts providing context about the survey in which videos were shown to respondents, to very short prompts, where the shortest prompt was *"Rate texts, from 1 to 100. Only report a single number between 1 and 100, rounded to two decimals. No text!:"*.

5.2.2 RESULTS

The ChatGPT API provided outputs of the 90 batches (i.e., 9 interfaces x 10 batches per interface) in a total of 72 seconds on a standard laptop and internet connection. The first prompt we tried (see Methods) yielded a very strong correlation between the ChatGPT outcomes and the composite from the numeric rating scales ($r = 0.975$). In other words, ChatGPT produced sentiment scores that correlated extremely highly with aggregated human ratings.

Through trial and error, we modified the prompt as follows: *"Looking at the comments, score the interface, from 1 to 100. Only report a number between 1 and 100, rounded to two decimals"*. The correlation coefficient between the two measures was found to be close to 1 ($r = 0.993$). The scatter plot in Figure 5.3 shows the mean sentiment score from the ChatGPT outputs and the composite score based on the rating scales.

The ChatGPT sentiment score not only strongly correlated with the composite score as calculated from the rating scales completed by the same participants; it also correlated strongly with the VADER sentiment score, a more classical lexicon rule-based method, specifically $r = 0.996$. This correlation coefficient is illustrated in Figure 5.4. The VADER sentiment score also strongly correlated with the composite score ($r = 0.992$).

It is noteworthy that the prompt used is short, and devoid of context. During the trial-and-error process, we discovered that longer prompts, which provided more context (such as that the experiment involved AR and a car and pedestrian in a road crossing scenario) were not needed for achieving a high correlation. It was also not needed to mention the dimension across which the sentiment score should be provided (e.g., quality, clarity).

We also explored the impact of batch size and the randomness parameter on sentiment analysis results. We found that using a smaller batch size (of 25 instead of 100 comments) did not substantially affect the correlation coefficient; it changed from 0.993 to 0.987. We also tried the alternative, namely submitting all 992 comments per AR interface at once, making use of a recent model of ChatGPT that can process 16,384 tokens (gpt-3.5-turbo-16k-0613). This approach turned out to be less effective, resulting in a correlation coefficient between the output and the composite score of only 0.724.

Furthermore, we found that adjusting the randomness parameter (by setting it to a higher value than 0 and averaging the results of multiple repetitions did not increase the correlation coefficient either. Figure 5.5 provides an illustration of this finding. We generated 100 sentiment scores for the first batch of comments, specifying the number of scores by setting the 'n' parameter to 100 in the API request. We repeated this process for different 'temperature' settings, from 0 up to the maximum possible value of 2. We then calculated the mean and standard deviation across these sentiment scores. As shown in Figure 5.5, the mean remained relatively constant across the different temperature settings. Conversely, the standard deviation grew with higher temperature values. This indicates that the sentiment

exceeding the token rate limit in certain instances, which resulted in no data being output in our loop.

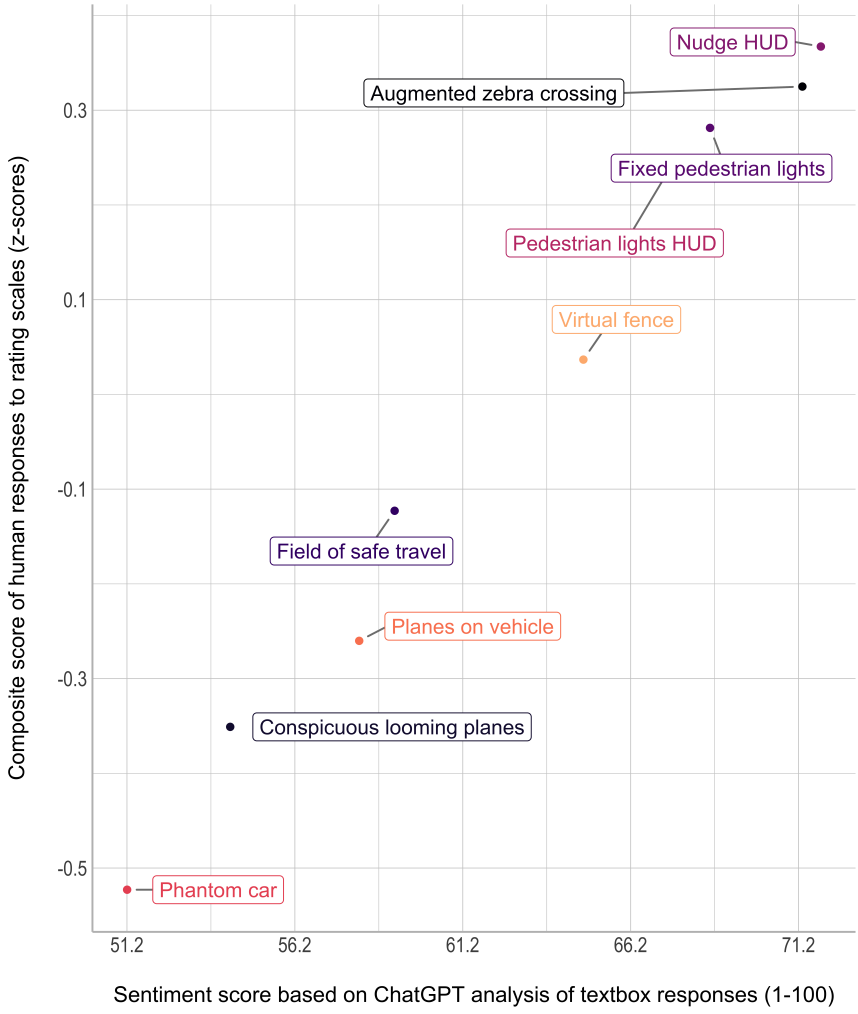


Figure 5.3: Scatter plot showing the composite score based on rating scales in the online questionnaire study (extracted from Tabone et al., 2023a) against the mean sentiment score calculated from the ChatGPT outputs, for nine augmented reality interfaces. The composite score has a mean of 0 and a standard deviation of 1 in the population of respondents.

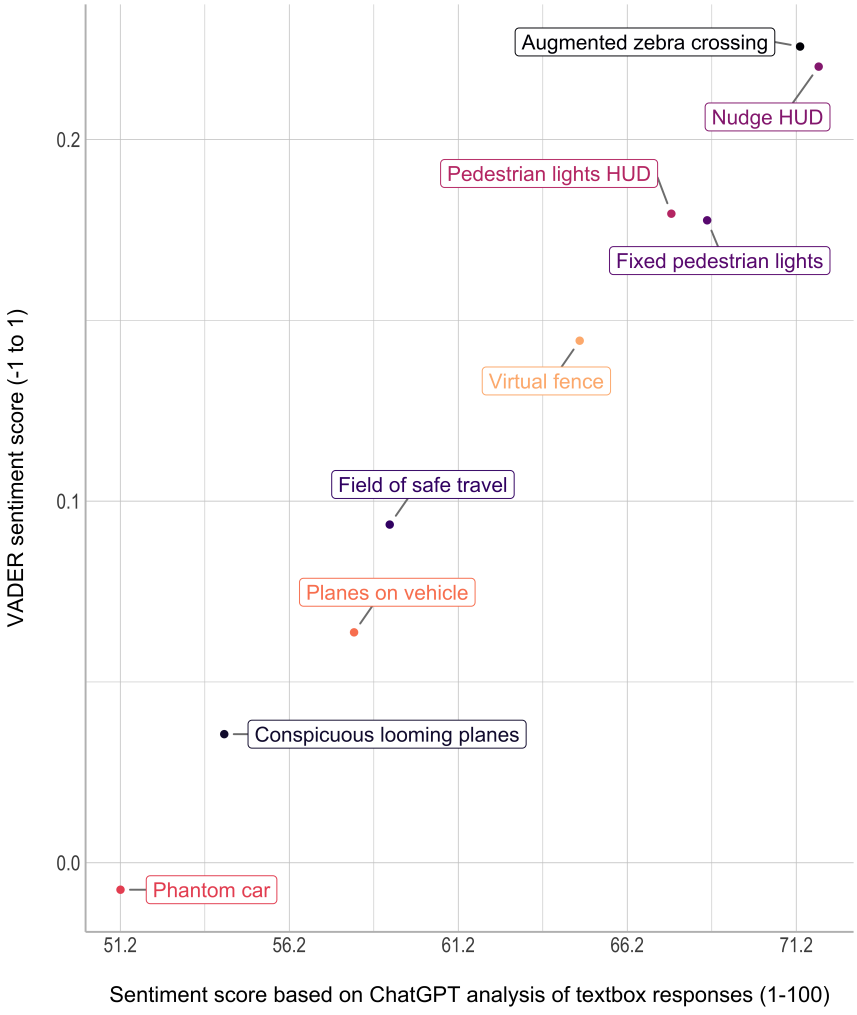


Figure 5.4: Scatter plot showing the sentiment score calculated using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment method, a lexicon and rule-based algorithm (Hutto & Gilbert, 2014) against the mean sentiment score calculated from the ChatGPT outputs, for nine augmented reality interfaces.

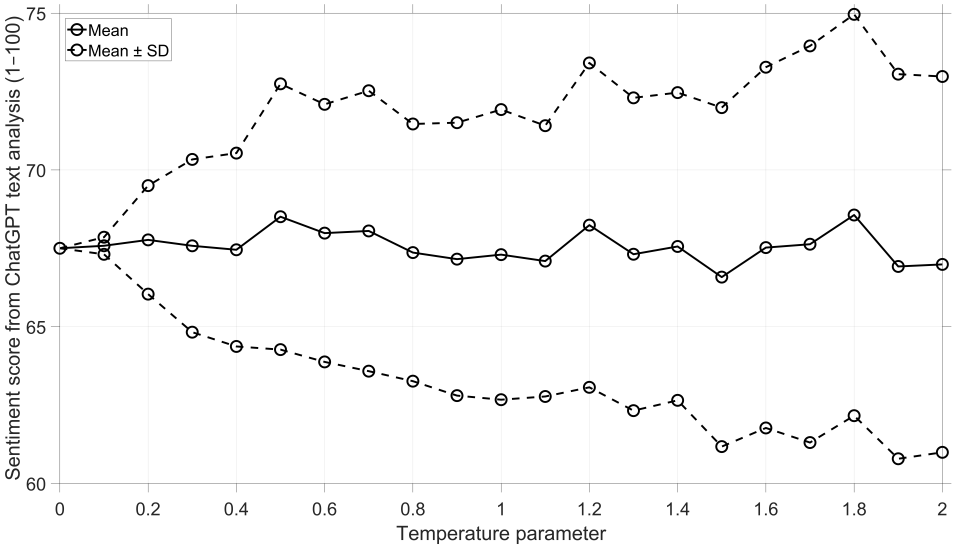


Figure 5.5: Mean and mean ± standard deviation of the sentiment score based on ChatGPT analysis of textbox responses, for one batch of texts (Augmented zebra crossing, texts 1–100 of 992), as a function of the temperature parameter (in increments of 0.1).

analysis output remained unbiased irrespective of the temperature setting used. In other words, in the context of this sentiment analysis task, there seems to be no specific need for a nonzero temperature setting.

As mentioned, using the initial prompt yielded a strong correlation of 0.975 between ChatGPT's sentiment analysis and the average human scores. By modifying the prompt, we were able to increase this correlation to 0.993. However, one limitation was that the output was of limited resolution. To illustrate, of the 90 outputs (9 AR interfaces × 10 batches per AR interface), the score '67.50' occurred 18 times, the score '62.50' 13 times, and the score '54.29' 6 times. A subsequent question we addressed is whether the correlation coefficient could be further improved through bootstrapping. Using the bootstrapping method, after trying 24 different prompts, we were able to increase the correlation coefficient to 0.997, with the following prompt: *"These text messages are obtained from a textbox in a questionnaire about nine interfaces for pedestrians who interact with an automated vehicle. I need you to provide a single sentiment rating about the interface being discussed in the comments, from 1 to 100. Only report a single number between 1 and 100, rounded to two decimals. No text!:"* It was also found that overly truncating the prompt did not help; the shortest prompt mentioned above resulted in a correlation of 0.984.

In summary, it was found to be possible to obtain a near-perfect association between text sentiment scores obtained with ChatGPT, on the one hand, and numerical sentiment scores obtained using human responses to rating scales and lexicon and rule-based sentiment scores, on the other.

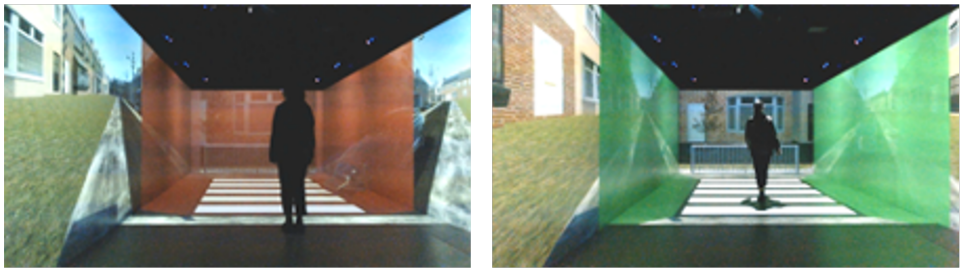


Figure 5.6: Non-yielding (left) and yielding (right) state of the *Virtual fence* interface.

5.3 STUDY 2: INTERVIEW DATA

IN Tabone et al. (2023b), the effect of nine AR interfaces, as well as a control condition without AR interface, on pedestrian crossing behaviour was investigated in a virtual simulator environment. This was a follow-up of the online study described in the previous section and was conducted at the University of Leeds using the Highly Immersive Kinematic Experimental Research (HIKER) simulator (Figure 5.6). The experiment was approved by the University of Leeds Research Ethics Committee under ethics reference number LLTRAN-150. All participants provided written informed consent.

The experiment consisted of 120 trials per participant, each featuring one of the interfaces, a different location of a visual attention attractor, and a vehicle which could either be yielding or non-yielding. The 120 trials followed a counterbalanced blocked design, divided into 10 blocks (with each block corresponding to one interface). After each block, the participant was interviewed for about 3 minutes, starting by asking whether they were comfortable, followed by, *"what did you think of this particular interface/situation?*". Depending on the participant's response, follow-up questions about the clarity of the colour coding (e.g., *"is the colour code making any sense to you?"*) and a comparison between different interface states (e.g., *"between the two states, were there any preferences"*) were asked.

Processing the interview data would be a challenging task, involving considerable transcription work and subjective interpretation. Therefore, it was decided to transcribe the audio-recorded interviews automatically and submit the transcripts to ChatGPT. In this paper, we examine whether ChatGPT provided valid insights into the qualities of each interface.

5.3.1 METHODS

First, the 300 voice recordings were transcribed by submitting the audio files to the paid online service Otter.ai (2023). An example of a transcript for the *Virtual fence* is shown in Figure 5.7. As can be seen, the transcript did not distinguish between the interviewer and interviewee, and contained some transcription errors.

Next, each anonymous transcript was submitted to the ChatGPT API (date: 4 March 2023, model: gpt-3.5-turbo) using a custom script. The following prompt was used: *"Looking at the participant's responses, provide a summary of the interface"*, followed by the participant's interview on the subsequent lines.

After collecting all responses from all 300 interviews, the responses per interface were again submitted to ChatGPT. The same prompt was used: *"Looking at the participant's*

responses, provide a summary of the interface", including all 30 summaries from the participants. The entire process used is detailed in the flowchart presented in Figure 5.8.

A key challenge in assessing the validity of interview summaries stems from ChatGPT's tendency to always generate a plausible output. Deciding if the output is truth-based (i.e., valid) or a hallucination (a known issue with large language models; Ji et al., 2023), requires further consideration. We have chosen to evaluate the validity of one interview summary by involving an independent human rater. A fellow researcher (not an author) was provided with the 30 interview transcripts for one of the interfaces (*Virtual fence*), with the task of performing a content analysis. This fellow researcher created categories (e.g., 'Interface intuitive', 'Walls gave a safe feeling'), split the transcript into individual sentences or sub-sentences expressing a single notion, and classified each of these into one of the categories. The researcher adopted an iterative procedure: as more transcripts were examined, categories were added and refined to capture emerging categories.

Upon completion of this independent analysis, the authors quantified the frequency of each category's occurrence by counting how many of the 30 participants made a remark per category. If the same participant mentioned a topic for a category more than once, this was counted as one instance for the category. Subsequently, we evaluated whether the summary generated by ChatGPT corresponded with the counts obtained by the fellow researcher.

Lastly, we evaluated whether ChatGPT4, OpenAI's paid version that outperforms ChatGPT3.5 on various benchmark tests (Bubeck et al., 2023; Katz et al., 2023; Nori et al., 2023), would yield a richer and more accurate summary. To do this, we repeated the aforementioned process, again via the API, but this time using the gpt-4-0613 model.

5.3.2 RESULTS

The result was a summary of the qualities of each interface (see Tabone et al., 2023b, for all summaries). Our script, which connected to the ChatGPT3.5 API, took a total of 25 min to process all 300 interviews, including creating the overall summaries. The overall summary for the Virtual fence is provided below:

Overall, the participants had mixed feelings about the interface. Some found it clear and easy to use, while others found it confusing and overwhelming. The green and red states were generally well-received, with participants feeling safer and more confident when the green state was active. However, some participants had concerns about the interface obscuring their view of the road and potentially causing accidents. The zebra crossing was generally appreciated, but some participants found it confusing and potentially misleading. Accessibility features were also noted as a positive aspect of the interface. Overall, there were varying opinions on the effectiveness and usability of the interface.

The results of the content analysis, shown in Table 5.2, are generally in line with the above summary of the interviews. For example, the content analysis indicated that 8 participants stated that the tunnel walls inhibited visibility, and 3 participants suggested that the walls provided a false sense of security, which relates to the summary statement: *"... some participants had concerns about the interface obscuring their view of the road and potentially causing accidents."* The remark about the zebra being potentially misleading was noted by 2 participants in the content analysis (*"Red is counterintuitive because it shows a zebra"*), which aligns with the summary (*"...some participants found it confusing and potentially misleading."*). The overall result of 'mixed feelings' is also in alignment with

00:00
So, say 2.8 and of block five. Do you feel comfortable?

00:06
Still feel okay. Yeah.

00:08
Okay, so how to evaluate this interface?

00:12
I'm certainly uncomfortable with the designing. I mean, because of you have like a blocking both sides and I can't really see clearly what's happening behind the color color. And then with the crossing lanes is kind of disturbing. Yeah, in terms of the designing is kind of disturbing to the eye because I prefer to look natural than having like too much digitalization on ice. Where the

00:43
starting point, influence your decision. Like, when you look into different directions. The interface with this turbine all the time, oh, it's somehow less disturbing when you are facing a specific direction.

01:01
It doesn't matter. It doesn't matter which direction I'm facing. But it I feel disturbing with the designs. I'm not preferring to have like, full color coded around my eyes.

01:14
Okay, so is the color code making any sense to you? Very clearly indicates its meaning.

01:25
I'd say the color choosing is indicating correct meaning, but I don't prefer like to have like a full block. Designing throughout my viewing. Yeah.

01:39
Okay, so do you comfortable with the choice, the confidence in the choice you're making with this interface?

01:48
Because if I'm slightly uncomfortable with designing, it doesn't mean that I didn't trust the decision by interfaces. I'd say, Oh, I see. I still have have a trust. It's just that I don't feel really comfortable. Okay, that's good. Thank you.

Figure 5.7: Example of an automated transcription (Participant 8, Virtual fence), created by Otter.ai (2023).

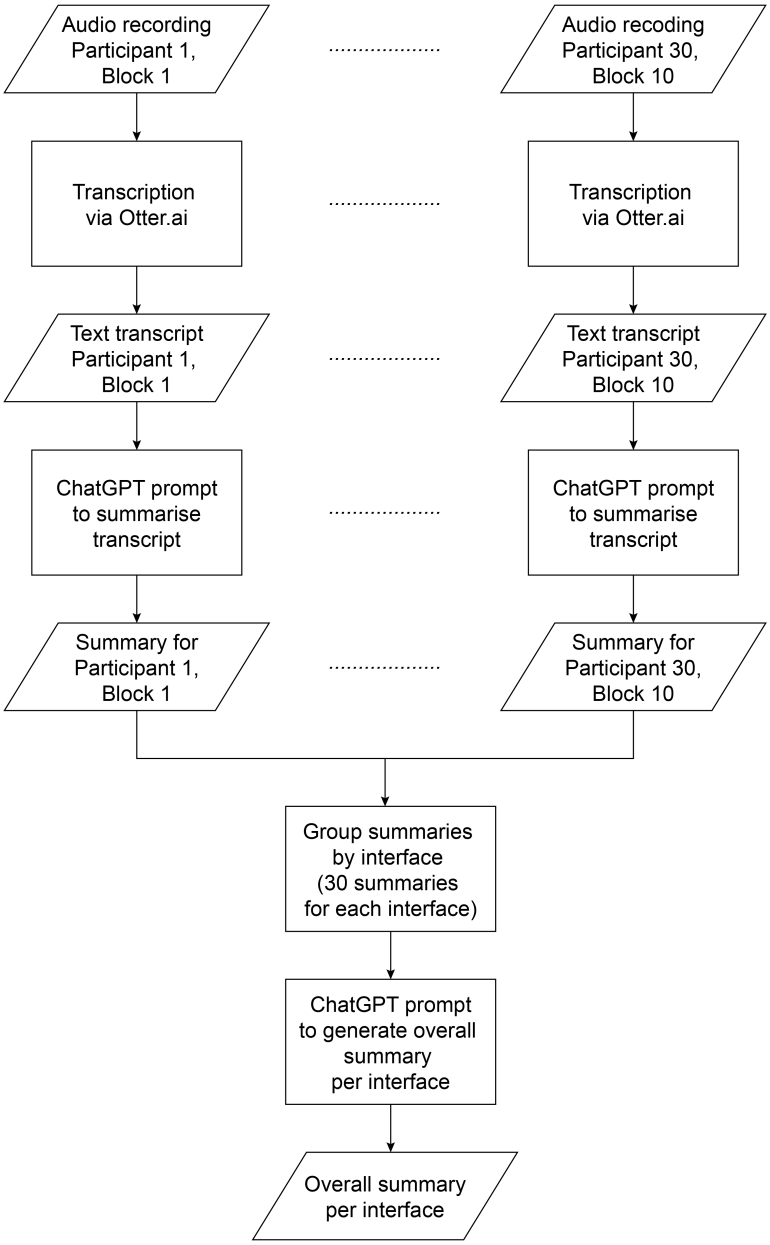


Figure 5.8: Flowchart detailing the process of transcribing and summarising each transcript. Transcripts were then grouped per interface, and an overall summary was generated using a further ChatGPT prompt.

the content analysis, where some participants found the tunnel intuitive, but others found it confusing.

The final summary using ChatGPT4 is provided below:

The participants' responses to the interface were mixed. Some found it intuitive and easy to understand, appreciating the clear signals of when to cross and when not to cross. They felt safer and more confident when using the interface, particularly appreciating the green signal that indicated it was safe to cross. However, some participants found the interface confusing and overwhelming, particularly in relation to the green and red signals. They felt that the interface was too busy and cluttered, and that the colors could be confusing, especially for colorblind individuals. Some participants also expressed concern about the interface potentially impacting their safety due to its high walls and lack of transparency, which could obscure their view of other vehicles or cyclists. Despite these concerns, most participants were able to adapt to the interface by the end of the session and found it generally helpful in making crossing decisions. Some participants suggested improvements, such as adding audio support or making the interface more natural-looking rather than fully digitalized.

It can be seen that the ChatGPT4 summary is richer and touches on more facets than the ChatGPT3.5 summary. The ChatGPT4 summary makes the point that the interface was too overwhelming or cluttered, which coincides with the content analysis, in which 4 participants mentioned that the interface was large, 3 found it too large, and 3 said there 'too much going on'. The summary also offers recommendations on how the interface can be improved. One point of criticism is that the ChatGPT4 summary mentions issues with audio support and colour blindness, while these topics did not emerge in the content analysis. Our inspection of the transcripts mentioned the word 'colorblind' only once, and the word 'audio' (or 'sound') also only once.

Table 5.2: Experimenter Questions and Participant Responses

Experimenter Questions	**Number of Participants**
Wellbeing monitoring	26
Thoughts on interface	26
Preference between interfaces	20
Clarification/follow-up question	19
Participant Responses	**Number of Participants**
No preference between green and red	10
Walls inhibited visibility	8
Relied on the interface more than the car/interface influenced decision	8
Walls gave a safe feeling	8
Interface clear	7
Interface intuitive	7
Green preferred	7
Interface not intuitive/counterintuitive/confusing	5
Red preferred	5
Interface salient/large	4
Interface safe	4
Relied on the car more than the interface/interface did not affect decision	4
One trial needed for understanding	3
Interface easy	3
Interface too large	3
Red not intuitive/counterintuitive/confusing	3
Interface not safe	3
Walls provided false safety	3
Interface had too many signals/too much going on	3
Bad that interface did not explicitly relate to the approaching car	3
Felt confident	3
Realistic	3
Unclear/gibberish	3
Green intuitive	2
Red is counterintuitive because it shows a zebra	2
Good that the interface was not attached to the approaching car	2
Liked the interface	2
Did not feel comfortable with the interface	2

5.4 STUDY 3: THINK ALOUD DATA

IN this study, data from a think-aloud protocol was recorded while participants, wearing a mobile eye tracker, freely gazed at Rembrandt's *The Night Watch* (1642) in Amsterdam's Rijksmuseum (De Winter et al., 2022). Unlike Studies 1 and 2, Study 3 does not involve HCI; participants expressed their thoughts through verbalisation while examining a painting, rather than interacting with a computerised system. Nevertheless, the use of the think-aloud method is firmly rooted (Lewis, 1982; Rasmussen & Jensen, 1974) and often employed (Fan et al., 2019; Hoppmann, 2009; Trukenbrod et al., 2020) in the HCI domain, and can therefore be viewed as a prototypical tool within this field.

The aim of the original study was primarily to assess the participants' attentional distribution for this work of art. The analysis of the gaze data was supported by the recorded statements from the participants as they freely spoke out their thoughts while gazing at the painting (Figure 5.9, left).

The recorded voice data (in Dutch) was previously transcribed and analysed using a hybrid approach, combining thematic analysis with a quantitative method of tabulating word frequencies. The main themes were identified, and the number of times a target word belonging to each theme was spoken was also counted.

A follow-up experiment was conducted to assess the robustness of the first experiment, using a replica of the painting in a laboratory setting and with new participants (Figure 5.9, right). The participants in the Rijksmuseum study were recruited by contacting acquaintances of the authors and television crew and by inviting (ex-)students, whereas participants in the replica study were recruited from the student population and the teaching and administration staff of the Delft University of Technology. The task instructions and experiment duration (5 min of viewing per participant) were identical between the two settings.

The authors discovered that admiration-related words (i.e., beautiful, cool, enjoy, fantastic, fascinating, great, impressive, nice, special, splendid, unique, wonderful, wow) were uttered more frequently for the real painting compared to the replica painting; other comparisons of think-aloud data between the two versions of the painting did not yield statistically significant differences (De Winter et al., 2022).

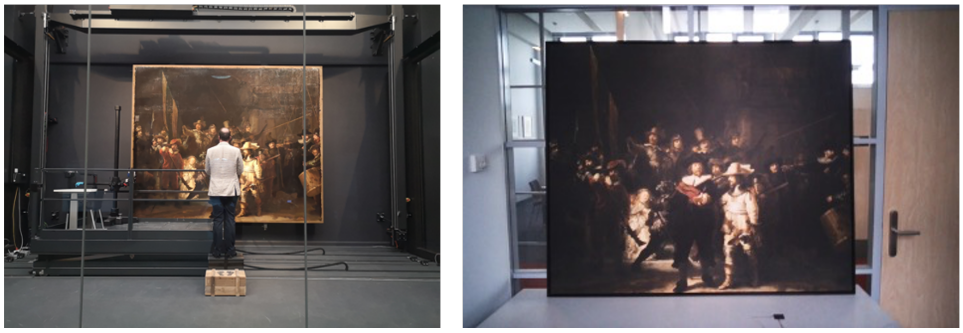


Figure 5.9: Experimental setup for *The Night Watch* study in a glass chamber in the Rijksmuseum (left) and a replica in an office environment (right) (De Winter et al., 2022).

5.4.1 METHODS

The supplementary material for *The Night Watch* experiment (De Winter et al., 2022) contains two transcript files in Dutch, one for participants observing the real painting in the Rijksmuseum, and another for participants looking at the replica painting in the laboratory.

A pairwise comparison was conducted between transcripts by order of participant numbers, i.e., the first transcript in one file was compared to the first transcript in the other, and so on. Since there were a total of 21 participant transcripts in the recordings taken at the Rijksmuseum, and 27 for the replica, the new analysis included only the first 21 participant pairs.

We used the online chat window, instead of the API used in Studies 1 and 2. On January 16, 2023, each participant pair was compared in a separate chat window, using three prompts. The first prompt was: *"Consider the following transcripts:"*, followed by the transcript of the original painting. Based on this prompt, ChatGPT automatically provided a short summary in English. This was followed by a transcript related to the original painting: *"How does it differ from the following transcripts, in essence?"*, followed by a transcript from the replica painting. Based on this prompt, ChatGPT produced a comparison of the two transcripts. Next, the following prompt was given: *"What would you regard as the key difference between both transcripts?"*.

To arrive at the final output, a hierarchical approach was again adopted. Specifically, once all 21 comparisons were performed, the key difference outputs were grouped into batches of 5, 5, 5, and 6. Each batch was summarised into a 200-word paragraph using the prompt *"Please provide a 200 word summary for the following text:"*. The four resulting summaries were then combined and condensed into a general 400-word summary using the prompt: *"Please provide a 400 word summary for the following text:"*. The entire process used is detailed in the flowchart presented in Figure 5.10.

In Study 1, the sequence in which texts were submitted to ChatGPT was found to have some influence on the obtained sentiment score, even when the content of the texts remained the same. In addition to the fact that ChatGPT has a token limit, this sensitivity to the prompt was a reason for us to submit the text in batches and subsequently average the sentiment scores. A similar limitation might affect the aforementioned analysis, where we entered the transcripts corresponding to the real *The Night Watch* painting first, followed by those pertaining to the replica. To investigate whether this sequence introduces a potential bias, we performed an additional analysis through the API (model: gpt-3.5-turbo-0613). Our employed method is more comprehensive than the approach executed with ChatGPT via the web interface and involved prompting ChatGPT as follows: *"TRANSCRIPT 1: <transcript of a participant for the real painting> TRANSCRIPT 2: <transcript of a participant for the replica painting> What would you regard as the key difference between both transcripts?*"

This procedure was conducted for all 567 possible pairs, combining participants from the real painting (21 participants) with those from the replica painting (27 participants). The 567 pairs were sorted in random order. We then repeated this process but for the real painting and the replica submitted in reverse order. Accordingly, we prompted ChatGPT as follows: *"TRANSCRIPT 1: <transcript of a participant for the replica painting> TRANSCRIPT 2: <transcript of a participant for the real painting> What would you regard as the key difference between both transcripts?"*. Consequently, we obtained two sets of 567 descriptions of how the transcripts for replica and real *The Night Watch* differed from one another, once for the

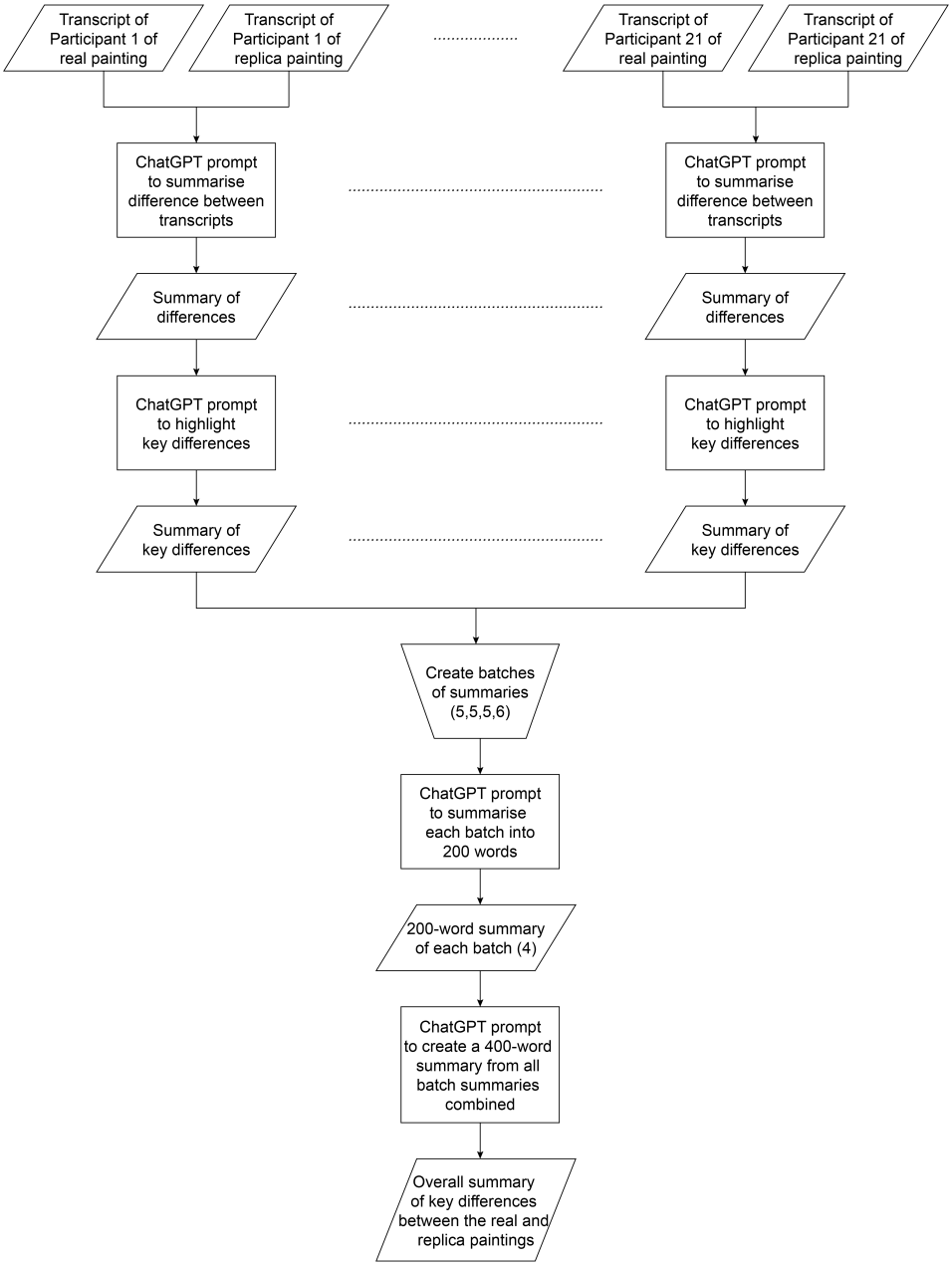


Figure 5.10: Flowchart detailing the process used to summarise the key differences between the real and replica The Night Watch, as stated by participants in a think-aloud protocol.

situation where the real painting was first mentioned in the prompt, and another time where the replica was first mentioned in the prompt.

Subsequently, we implemented the following procedure for each sequence to obtain a final summary. First, we separated the 567 pairwise comparisons into 7 batches of 81 each, maintaining their original random order. We then submitted each batch to the ChatGPT API (model: gpt-3.5-turbo-16k) using the following prompt: *"These are transcripts of different pairs of speakers. Provide one single 200-word summary of all Transcripts 1 combined; as well as one single 200-word summary of all Transcripts 2 combined. Report as follows: "Transcript 1: ...; Transcript 2: . . . ""*. Lastly, for both sequences, we executed a prompt to create a final summary, using the following prompt: *"Provide one single 200-word summary of the differences between Transcripts 1 and 2, of all batches combined:"*, followed by the summaries of the 7 batches. The objective of this analysis was to determine whether the final summaries were consistent, irrespective of whether the real painting or the replica painting was presented first in the prompt.

5.4.2 RESULTS

The text below shows the final 400-word summary that was generated by ChatGPT, highlighting the differences between what participants said while observing *The Night Watch* in the museum and its replica in the laboratory:

"The text describes the differences between two transcripts of two people's observations of a painting. The first transcript is more emotive, focusing on the personal experience of looking at the painting, showing awe and amazement, trying to understand the meaning, and expressing it as a once in a lifetime experience. The person in this transcript is deeply moved by the painting and finds it fascinating, providing a detailed description of the painting and its intricate details, dynamic lighting and highlights on the faces, and expressing its beauty and realism.

The second transcript is more analytical, providing a more general and surface-level description of the painting, focusing primarily on the activity of the figures and their presence of weapons and musical instruments. It also notes the lighting, but makes less interpretive statements and has less focus on the individual figures and their meaning in the scene. Additionally, the second speaker seems to have a better knowledge of the painting and its historical context, including information on its restoration and the identification of specific elements such as a dog, chicken or dove, and the presence of Rembrandt himself in the painting.

The main difference between the two transcripts is the tone, focus, and level of detail provided by the speakers. The first transcript is more emotional and expresses awe and wonder towards the artwork, while the second transcript is more analytical, providing objective observations and less emotive language. Additionally, the second speaker seems to be more knowledgeable about the painting and its historical context, but less engaged with the painting, providing a more general and surface-level description of the painting."

The generated summary above indicates that ChatGPT captured how the participants looking at the real painting were *"deeply moved by the painting"*, and it commented on the *"dynamic lighting and highlights on the faces"*. Moreover, the participants who viewed the real painting were deemed to be 'more emotional' and expressed *"awe and wonder towards the artwork"*, in contrast to the participants who viewed the replica, which was described as

being *"more analytical, providing objective observations and less emotive language"*.

The aforementioned analysis was conducted using the ChatGPT web interface for 21 paired comparisons. Our more refined analysis, in which all 567 possible pairs between participants were reduced to final summaries through the API, resulted in the following final summary for the situation where the transcripts for the real painting (Transcript 1) were presented before the transcripts of the replica painting (Transcript 2):

"The key difference between all the transcripts in the combined batches is the level of detail, analysis, and engagement provided by the speakers. Transcript 1 consistently offers more detailed descriptions and observations of the painting, focusing on elements such as lighting, composition, colors, clothing, and specific details. The speakers in Transcript 1 express fascination, admiration, and curiosity about the painting, providing personal insights, interpretations, and emotional responses. They also discuss the historical context and significance of the artwork. In contrast, Transcript 2 consistently provides more general and superficial descriptions, focusing on the overall impression, recognition of figures, and some specific details. The speakers in Transcript 2 express uncertainty, confusion, and a lack of knowledge or familiarity with the painting, often speculating on possible meanings and interpretations. They do not provide as much detail, analysis, or personal engagement as the speakers in Transcript 1. Overall, Transcript 1 demonstrates a deeper level of engagement, understanding, and appreciation for the artwork, while Transcript 2 offers a more casual and less informed perspective."

And for the reversed situation, where the transcripts for the replica painting were submitted first (Transcript 1) followed by the transcripts for the real painting (Transcript 2), the final summary was as follows:

"The key difference between the transcripts in all batches combined is the level of detail and analysis provided. Transcript 1 tends to provide more general and surface-level descriptions of the painting, focusing on the overall scene, main figures, and notable details. The speakers in Transcript 1 also express uncertainty and lack of knowledge about the historical context. In contrast, Transcript 2 demonstrates a deeper appreciation and engagement with the painting. The speakers in Transcript 2 mention specific details they find fascinating, such as the eyes, hidden elements, and clothing. They also reflect on their personal experiences and emotions while viewing the painting. Transcript 2 provides a more in-depth observation and analysis of the artwork, discussing elements such as composition, lighting, colors, and interactions between figures. The speakers in Transcript 2 also question the authenticity and meaning of certain elements, providing a more analytical and subjective response to the painting. Overall, Transcript 2 offers a more detailed and analytical description of the painting, while Transcript 1 tends to provide more general and observational descriptions."

It can be seen that the two summaries above for the real painting have several similarities. Both refer to detailed descriptions and observations of the painting, with emphasis on composition, lighting, and colours. They also show an emotional and personal connection to the artwork, discussing their feelings and insights (*"fascination, admiration, and curiosity about the painting, providing personal insights, interpretations, and emotional responses"* vs. *"details they find fascinating, such as the eyes, hidden elements, and clothing. They also reflect on their personal experiences and emotions while viewing the painting"*). Additionally, the speakers touch upon the deeper meanings, significance, or questions related to the

artwork. (*"discuss the historical context and significance of the artwork"* vs. *"question the authenticity and meaning of certain elements, providing a more analytical and subjective response to the painting"*.)

With regard to the replica, both summaries mention providing more general descriptions (*"more general and superficial descriptions"* vs. *"more general and surface-level descriptions"*). The summaries also highlight a focus on the overall impression (*"overall impression"* vs. *"overall scene"*) and the key figures (*"recognition of figures"* vs. *"main figures"*), and both emphasise the speakers' *"uncertainty"* and *"lack of knowledge"* in their discussions about the painting.

In conclusion, the final summaries are robust across prompting sequences (real painting first vs. real painting second) and in line with the keyword analysis by De Winter et al. (2022) which found more admiration for the real painting compared to the replica. Simultaneously, it can be observed that the two summaries pertaining to the transcripts of the real painting and the two pertaining to the replica exhibit differences in their specific phrasings, even though they originate from the same transcripts and the only difference was the sequence of prompting. The full analysis can be found in the supplementary material, and the findings are further discussed in the Discussion section.

5.5 DISCUSSION

In this study, we examined whether ChatGPT can extract valid patterns from text data. Firstly (Study 1), we had ChatGPT perform a sentiment analysis on texts from a questionnaire study, and assessed criterion validity by calculating the correlation coefficient with average numeric rating scale scores of AR interfaces as well as with a lexicon and rule-based sentiment analysis method. Secondly (Study 2), we had ChatGPT summarise raw interview transcripts without preprocessing, and assessed face validity by evaluating the summaries, as well as criterion validity by comparing it with the results of a manual content analysis. Thirdly (Study 3), we examined whether ChatGPT is powerful enough to detect differences between think-aloud transcripts related to two different experimental conditions, namely a real painting versus a replica. Here, we assessed face validity by seeing if the summaries were sensible, as well as criterion validity by comparing the summaries with the results of prior keyword counts.

In the evaluation of questionnaire textbox responses (Study 1), ChatGPT exhibited remarkable proficiency in assessing sentiment. This was demonstrated by the virtually perfect correlation ($r > 0.99$) between the ChatGPT sentiment scores and the pre-existing scores derived from both rating scales. This correlation was on par with the statistical reliability of the criterion rating scales themselves, effectively reaching the upper ceiling of statistical similarity. The correlation between the ChatGPT sentiment score and the scores obtained through a traditional lexicon and rule-based sentiment analysis (VADER) was nearly perfect as well ($r = 0.996$). Collectively, these results indicate that ChatGPT possesses criterion validity in sentiment analysis with respect to numerical responses from human participants, as well as in relation to a previously published popular method of sentiment analysis. The strength of ChatGPT lies in its generalisability; whereas the VADER method has been specifically developed for sentiment analysis, ChatGPT was not, which highlights its potential for different types of numerical analyses of text data.

We also demonstrated several techniques that can be used to improve the criterion

validity coefficient, namely, adjusting the prompt so that it contains sufficient (but not more than needed) context and the use of a bootstrapping method. The essence of the bootstrapping method is that text comments are repeatedly submitted to ChatGPT in random order, averaging out the effect of ChatGPT's preference for certain discrete numerical outputs as well as its sensitivity to the order in which texts are submitted.

In Study 2, ChatGPT was found to be a valid tool for summarising interview data from human subjects. The interview summary was found to be logically interpretable and largely corresponded with the results of a content analysis conducted by a human researcher who manually tabulated the interview results. In our case, we found that ChatGPT3.5 provided a basic yet accurate summary, while ChatGPT4 offered a more refined summary of the interviews. However, the ChatGPT4 summary also included some points which were perhaps considerate and relevant, but not main themes in the raw transcripts. Another topic from the content analysis, namely whether the pedestrian based the crossing decision on the car or the AR interface, did not appear in either of the two ChatGPT summaries. A possible explanation is the hierarchical summarisation method we used, which caused less essential yet recurrent information to fall away in the intermediate steps.

Finally, in Study 3, ChatGPT was found capable of highlighting subtle differences in what participants said when observing *The Night Watch* in the Rijksmuseum compared to a replica in an office environment. These results agree with the manual analysis of the think-aloud method, which revealed that the participants uttered more words related to admiration for the real *The Night Watch* than for the replica. Our final summary also adds to the ongoing discourse surrounding the effects of a painting in the context of the museum (Krukar & Dalton, 2020; Pelowski et al., 2017; Specker et al., 2017) and in a laboratory setting (DiPaola et al., 2013). However, unlike the manual analysis performed by De Winter et al. (2022), which provided keywords grouped by theme, ChatGPT was able to generate a paragraph that highlighted differences between the major themes identified in the two settings by submitting transcripts without further context.

It is important to note that if a single human researcher were to have generated the text output for either Study 1, 2, or 3, it may not be deemed trustworthy from a scientific perspective due to the limitations of human cognition, such as confirmation bias. It is possible that a human researcher, especially after having invested substantial effort into conducting the experiments, may expect or hope that certain differences between the two settings arise based on preconceptions or conflicts of interest. ChatGPT, having been provided with no prior information regarding the authenticity of the two paintings, created a discerning meta-summary. This approach, which is in alignment with the principles of researcher blinding, instils a certain level of confidence. However, as shown in Study 2, outputs from ChatGPT should not be unconditionally trusted. The value of ChatGPT may lie in pairing its outputs with other forms of analysis, such as counts and manual content analysis. Moreover, ChatGPT could serve as an auxiliary verification of human analyses, especially when there is a limited number of human annotators where, for example, an inter-rater reliability score cannot be calculated.

Apart from the validity assessments, it is noteworthy that the use of ChatGPT was time-efficient. Using the method presented in this paper, for Study 1 and 2, our script produced outputs in about 1 minute and 25 minutes, respectively, while for Study 3, it only required sorting the data and copying it into the ChatGPT interface. An additional benefit of our

pipeline was that the interview data of Study 2 was transcribed automatically, significantly reducing the time required compared to manual transcription. In contrast, the manual analysis of interviews can be challenging, particularly when dealing with large amounts of data, and subjective biases of the researcher may potentially influence the interpretation of the responses (Johnson et al., 2020; Morse, 2015). At the same time, it must be acknowledged that the development of the pipelines (as shown in the flowcharts in this paper) and the writing of the scripts that interacted with the API required a substantial amount of the researchers' time, scripts that are available for reuse (see supplementary material).

We also found that, although it is commonly advised that ChatGPT and similar large language models require prompts that provide suitable context (e.g., Jalil et al., 2023; White et al., 2023), we found that brief prompts were sufficient for performing a sentiment analysis. We also found that the randomness ('temperature') parameter should be set to 0 for accurate results.

When researchers wish to use ChatGPT, several limitations must be taken into consideration. Firstly, while ChatGPT offers high reproducibility when the temperature parameter is set to 0, it is crucial to understand that a reproducible output does not guarantee robustness. As pointed out above, ChatGPT is an autoregressive model, which implies that each produced token relies on the preceding prompt and the tokens generated up to that point. This dependency implies that, while ChatGPT's output is highly reproducible, it can be sensitive to small changes to the prompt (see also Bubeck et al., 2023; LeCun, 2023) and differently fine-tuned versions of the ChatGPT model (Chen, 2023; De Winter, 2023). For example, in Study 3, the sequence in which we presented the transcripts in the prompts (actual painting followed by replica or vice versa) affected the phrasing of the final summary. Similarly, in Study 1, we discovered that the sentiment score output could change when we submitted the same texts in a different order. A possible explanation for this is that any minor variations in the prompt and output continue to propagate (or even reinforce) in subsequent summaries, and that ChatGPT lacks a self-correction mechanism allowing it to converge to a specific optimal response. This sensitivity, combined with the limited context length offered by ChatGPT, motivated our choice to process the texts in batches. The rationale behind using multiple batches or the bootstrapping method is the expectation that random noise will average out with enough repetitions of varied prompts, a principle resembling the self-consistency method (Wang et al., 2022). In an alternative approach to Study 1, we submitted all 992 comments per AR interface simultaneously, using a newer model with an extended context length (gpt-3.5-turbo-16k). However, this approach yielded outputs of mediocre criterion validity, likely because of the single sentiment score given for the entire set of texts.

Another point of consideration pertains to the pipeline created from Study 3, which reduced the pairwise comparisons between transcripts of the original and replica paintings to a meta-summary. This process was found to be somewhat volatile and prone to errors. One significant challenge was the potential loss of vital information. For instance, depending on the prompt, ChatGPT might only highlight the difference between the two transcripts without indicating whether it is from Transcript 1 or 2. It might also produce output in inconsistent formats, such as narrative text versus bulleted statements. Furthermore, ChatGPT could generate over-generalised statements. An example is stating that participants closely examined the painting. Such a statement is likely applicable to both the original and the replica,

offering little distinction between the two. As in Study 2, we also observed instances where specific observations from a few participants could become unduly prominent in subsequent summaries. The robustness check we applied—where the transcripts were submitted in two different sequences—was beneficial in this regard. We believe that an effective approach to creating meta-summaries involves retaining information that consistently emerges across various prompting approaches. This strategy aligns with the bootstrapping method and the self-consistency method previously discussed (Wang et al., 2022). We further recommend involving a human researcher in the summarisation pipeline to ensure ChatGPT's outputs are neither too generic nor overly detailed.

In summary, it can be posited that ChatGPT produces reproducible output, but this does not imply that reliance should be placed on a single output. Researchers are advised to report the precise version number of ChatGPT in their paper, so that other scholars may select the same model if necessary. Secondly, researchers are recommended to examine the robustness of the output, for example, through methods such as bootstrapping or by applying minor variations to the prompt.

Moreover, it must be recognised that, as indicated above, the autoregressive nature of ChatGPT means it is not especially proficient at tasks that require retaining multiple variables in memory, including counting (Bubeck et al., 2023). For instance, in regard to Study 2, when trying to determine whether ChatGPT could perform a content analysis (akin to Table 5.2, created by a human researcher), we discovered that ChatGPT did not provide particularly accurate output. Here, we tried various prompt strategies, such as simultaneously submitting the interviews from all participants (using gpt-3.5-turbo-16k) and directly prompting ChatGPT to create a table. We also attempted a more structured approach, where we presented all categories from Table 5.2 to ChatGPT for each interview separately, prompting it to indicate whether or not the topic was addressed in the interview. In all cases, ChatGPT produced tables that were credible and fairly consistent with the narrative of the interview summry. However, the exact numbers (and the participants to whom these counts could be traced) were found to be insufficiently inaccurate and sensitive to the specific prompt. We believe these issues can be traced back to the aforementioned functioning of ChatGPT: It understands textual structures and can generate summaries of texts, but unlike humans, is not capable of keeping counts in working memory required to make an overview table. It should be noted, however, that a human researcher cannot carry out this task alone either, but typically employs tools such as spreadsheet software or other forms of annotation or note-taking tools. In our case, the researcher utilised Microsoft Excel in performing the content analysis that ultimately resulted in Table 5.2 (see supplementary material).

It should be further noted that ChatGPT's outputs also have limitations due to potential biases. These biases may arise from the fact that ChatGPT models have been trained on human-generated text and reinforcement learning from human feedback to better align with human values (OpenAI, 2022, 2023a). In particular, ChatGPT outputs could potentially contain biases toward political leanings (Hartmann et al., 2023; McGee, 2023; Motoki et al., 2023; Rutinowski et al., 2023). These possible biases are unlikely to have affected the results of Study 1, because ChatGPT was tasked to provide numeric scores and not to generate new ideas. However, it may have affected the results of Study 2 to some extent, wherein it caught our attention that ChatGPT-4 devoted attention to individuals who are colourblind in the generated summary. The inclusion of colourblind or otherwise impaired road users in traffic

indeed represents an important human value, and perhaps also one that a human researcher would have wanted to consider; however, strictly speaking, it was not something that was prominently featured in the 30 summaries and therefore did not emerge in the human content analysis. The challenge of alignment, that is, to what extent the output that large language models generate aligns with human values, is an ongoing research topic (Fischer et al., 2023; Liu et al., 2023). It can also be noted here that these values may be context-dependent, whereby in some cases, a researcher might want to receive a mechanistic output, such for a meta-analysis, and in other cases, they would prefer an output that takes into account more general human values such as safety and inclusiveness.

5.6 CONCLUSION

THIS work addressed the research question: To what extent does ChatGPT produce valid sentiment scores and summaries when applied to different forms of text data in HCI studies? We employed a hierarchical approach because ChatGPT has a limit to the number of tokens it can process. Specifically, we provided batches of text for it to summarise before producing a final score or summary. The results showed that (1) ChatGPT produced sentiment scores that correlated extremely highly with quantitative metrics, (2) produced meaningful summaries of interviews that aligned with content analysis, and (3) highlighted subtleties from think-aloud data that corresponded with a prior keyword count.

In conclusion, our findings suggest that ChatGPT is a valid tool for qualitative data analysis in HCI research. However, it is recommended to use ChatGPT as a complementary tool in conjunction with more quantitative analyses and critical human reasoning, areas in which the model currently shows limitations.

A strength of our study is that it examined various forms of validity across different types of texts, ranging from succinct textbox responses to patterns discernible in think-aloud data. Nevertheless, this work should be viewed as an initial step, and it is recommended that additional research, encompassing a more diverse range of text types and prompts, be undertaken. The rapidly evolving landscape of large language models, characterised by the increasing availability of a diversity of models, necessitates continual evaluation.

SUPPLEMENTARY MATERIAL

The MATLAB scripts, as well as the ChatGPT inputs and outputs, are available in the following repository: https://doi.org/10.4121/21916017

ACKNOWLEDGEMENTS

We thank Dimitra Dodou for performing the content analysis of Study 2.

REFERENCES

Alba, D. (2022). OpenAI chatbot spits out biased musings, despite guardrail. https://www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results

Amer-Yahia, S., Bonifati, A., Chen, L., Li, G., Shim, K., Xu, J., & Yang, X. (2023). From Large Language Models to databases and back: A discussion on research and education. arXiv. https://doi.org/10.48550/arXiv.2306.01388

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. http://dx.doi.org/10.2139/ssrn.4337484

Bommarito, M. J., II, & Katz, D. M. (2022). GPT takes the bar exam. arXiv. https://arxiv.org/abs/2212.14402

Borji, A. (2023). A categorical archive of ChatGPT failures. arXiv. https://doi.org/10.48550/arXiv.2302.03494

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3, 77–101. https://doi.org/10.1191/1478088706qp063oa

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv. https://arxiv.org/abs/2303.12712

Chen, L., Zaharia, M., & Zou, J. (2023). How is ChatGPT's behavior changing over time? arXiv. https://doi.org/10.48550/arXiv.2307.09009

Clemmensen, T., & Roese, K. (2010). An overview of a decade of journal publications about culture and human-computer interaction (HCI). In D. Katre, R. Orngreen, P. Yammiyavar, & T. Clemmensen (Eds.), Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts. HWID 2009 (pp. 98–112). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-11762-6_9.

Council of the European Union. (2023). ChatGPT in the public sector – overhyped or overlooked? https://www.consilium.europa.eu/media/63818/art-paper-chatgpt-in-the-public-sector-overhyped-or-overlooked-24-april-2023_ext.pdf

De Carvalho, P., & Fabiano, A. (2021). Thematic analysis for interactive systems design: A practical exercise. Proceedings of 19th European Conference on Computer-Supported Cooperative Work. European Society for Socially Embedded Technologies, Zürich, Switzerland. https://doi.org/10.18420/ecscw2021_wsmc06

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. Field Methods, 23, 136–155. https://doi.org/10.1177/1525822X10388468

De Winter, J. C. F. (2023). Can ChatGPT pass high school exams on English language comprehension? https://www.researchgate.net/publication/366659237_Can_ChatGPT_pass_high_school_exams_on_English_Language_Comprehension

De Winter, J. C. F., Dodou, D., & Tabone, W. (2022). How do people distribute their attention while observing The Night Watch? Perception, 51, 763–788. https://doi.org/10.1177/03010066221122697

DiPaola, S., Riebe, C., & Enns, J. T. (2013). Following the masters: Portrait viewing and appreciation is guided by selective detail. Perception, 42, 608–630. https://doi.org/10.1068/p7463

Eulerich, M., Sanatizadeh, A., Vakilzadeh, H., & Wood, D. A. (2023). Can artificial intelligence pass accounting certification exams? ChatGPT: CPA, CMA, CIA, and EA? SSRN. https://doi.org/10.2139/ssrn.4452175

Fan, M., Lin, J., Chung, C., & Truong, K. N. (2019). Concurrent think-aloud verbalizations and usability problems. ACM Transactions on Computer-Human Interaction, 26, 1–35. https://doi.org/10.1145/3325281

Fischer, R., Luczak-Roesch, M., & Karl, J. A. (2023). What does ChatGPT return about human values? Exploring value bias in ChatGPT using a descriptive value theory. arXiv. https://doi.org/10.48550/arXiv.2304.03612

Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., Chevalier, A., & Berner, J. (2023). Mathematical capabilities of ChatGPT. arXiv. https://doi.org/10.48550/arXiv.2301.13867

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. bioRxiv. https://doi.org/10.1101/2022.12.23.521610

Gerlach, J. H., & Kuo, F.-Y. (1991). Understanding human-computer interaction for information systems design. MIS Quarterly, 15, 527–549. https://doi.org/10.2307/249456

Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). How does ChatGPT perform on the medical licensing exams? the implications of large language models for medical education and knowledge assessment. medRxiv.

https://doi.org/10.1101/2022.12.23.22283901

González-Padilla, D. A. (2023). Concerns about the potential risks of artificial intelligence in manuscript writing. Journal of Urology. https://doi.org/10.1097/JU.0000000000003131

Gubrium, J. F., & Holstein, J. A. (2001). Handbook of interview research: Context and method. Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781412973588

Hartmann, J., Schwenzow, J., & Witte, M. (2023). The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation. arXiv. https://doi.org/10.48550/arXiv.2301.01768

Hoppmann, T. K. (2009). Examining the 'point of frustration'. The think-aloud method applied to online search tasks. Quality & Quantity, 43, 211–224. https://doi.org/10.1007/s11135-007-9116-0

Humphreys, A., & Wang, R. J.-H. (2018). Automated text analysis for consumer research. Journal of Consumer Research, 44, 1274–1306. https://doi.org/10.1093/jcr/ucx104

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 8, 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Indulska, M., Hovorka, D. S., & Recker, J. (2012). Quantitative approaches to content analysis: Identifying conceptual drift across publication outlets. European Journal of Information Systems, 21, 49–69. https://doi.org/10.1057/ejis.2011.37

Jalil, S., Rafi, S., LaToza, T. D., Moran, K., & Lam, W. (2023). Chatgpt and software testing education: Promises & perils. Proceedings of the 2023 IEEE International Conference on Software Testing, Verification and Validation Workshops, Dublin, Ireland, 4130–4137. https://doi.org/10.1109/ICSTW58534.2023.00078

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55, 1–38. https://doi.org/10.1145/3571730

Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. American Journal of Pharmaceutical Education, 84, 7120. https://doi.org/10.5688/ajpe7120

Karpathy, A. (2023). State of GPT | BRK216HFS [video]. https://www.youtube.com/watch?v=bZQun8Y4L2A

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). GPT-4 passes the bar exam. SSRN. https://doi.org/10.2139/ssrn.4389233

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. Medical Teacher, 42, 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Kim, Z. M., Taylor, D. E., & Kang, D. (2023). "Is the pope Catholic?" Applying chain-of-thought reasoning to understanding conversational implicatures. arXiv. https://doi.org/10.48550/arXiv.2305.13826

Kirmani, A. R. (2023). Artificial Intelligence-Enabled Science Poetry. ACS Energy Letters, 8, 574–576. https://doi.org/10.1021/acsenergylett.2c02758

Kitto, K., Manly, C. A., Ferguson, R., & Poquet, O. (2023). Towards more replicable content analysis for learning analytics. Proceedings of Learning Analytics and Knowledge 2023, Arlington, TX. https://doi.org/10.1145/3576050.3576096

Kjeldskov, J., & Skov, M. B. (2003). Creating realistic laboratory settings: comparative studies of three think-aloud usability evaluations of a mobile system. In M. Rauterberg, M. Menozzi, & J. Wesson (Eds.), Human-Computer Interaction (INTERACT'03) (pp. 663-670). Amsterdam, The Netherlands: IOS Press.

Krukar, J., & Dalton, R. C. (2020). How the visitors' cognitive engagement is driven (but not dictated) by the visibility and co-visibility of art exhibits. Frontiers in Psychology, 11, 350. https://doi.org/10.3389/fpsyg.2020.00350

LeCun, Y. (2023). Do large language models need sensory grounding for meaning and understanding? Spoiler: YES! [presentation]. https://drive.google.com/file/d/1BU5bV3X5w65DwSMapKcsr0ZvrMRU_Nbi/view Lewis, C. (1982). Using the "thinking-aloud" method in cognitive interface design (Research Report No. RC 9265). Yorktown Heights, NY: IBM TJ Watson Research Center.

Liu, R., Jia, C., Zhang, G., Zhuang, Z., Liu, T. X., & Vosoughi, S. (2023). Second thoughts are best: Learning to re-align with human values from text edits. arXiv. https://doi.org/10.48550/arXiv.2301.00355

Long, J. (2023). Large Language Model guided Tree-of-Thought. arXiv. https://doi.org/10.48550/arXiv.2305.08291

Maraj, C. S., Martinez, S. G., Badillo-Urquiola, K. A., Stevens, J. A., & Maxwell, D. B. (2016). Preliminary review of a virtual world usability questionnaire. In S. Lackey & R. Shumaker (Eds.), Virtual, Augmented and Mixed Reality: 8th International Conference (pp. 35–46). Cham: Springer. https://doi.org/10.1007/978-3-319-39907-2_4

McGee, R. W. (2023). Is Chat GPT biased against conservatives? An empirical study. https://doi.org/10.2139/ssrn.4359405

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. Qualitative Health Research, 25, 1212–1222. https://doi.org/10.1177/1049732315588501

Motoki, F., Pinho Neto, V., & Rodrigues, V. (2023). More human than human: Measuring ChatGPT political bias. SSRN. http://dx.doi.org/10.2139/ssrn.4372349

Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv. https://doi.org/10.48550/arXiv.2303.13375

OpenAI. (2022). Introducting ChatGPT. https://openai.com/blog/chatgpt

OpenAI. (2023a). How should AI systems behave, and who should decide? https://openai.com/blog/how-should-ai-systems-behave

OpenAI. (2023b). ChatGPT plugins. https://openai.com/blog/chatgpt-plugins

Otter.ai. (2023). Otter.ai - Voice meeting notes & real-time transcription. https://otter.ai

Patel, S. B., & Lam, K. (2023). ChatGPT: the future of discharge summaries? The Lancet Digital Health, 5, E107–E108. https://doi.org/10.1016/S2589-7500(23)00021-3

Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. Journalism & Mass Communication Educator. https://doi.org/10.1177/10776958221149577

Pelowski, M., Forster, M., Tinio, P. P. L., Scholl, M., & Leder, H. (2017). Beyond the lab: An examination of key factors influencing interaction with 'real' and museum-based art. Psychology of Aesthetics, Creativity, and the Arts, 11, 245–264. https://doi.org/10.1037/aca0000141

Peng, B., Galley, M., He, P., Cheng, H., Xie, Y., Hu, Y., Huang, Q., Liden, L., Yu, Z., Chen, W., & Gao, J. (2023). Check your facts and try again: Improving large language models with external knowledge and automated feedback. arXiv. https://doi.org/10.48550/arXiv.2302.12813

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: a case study of electronic trouble shooting. Ergonomics, 17, 293–307. https://doi.org/10.1080/00140137408931355

Roberts, K., Dowell, A., & Nie, J.-B. (2019). Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. BMC Medical Research Methodology, 19, 66. https://doi.org/10.1186/s12874-019-0707-y

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6.

https://doi.org/10.37074/jalt.2023.6.1.9

Rutinowski, J., Franke, S., Endendyk, J., Dormuth, I., & Pauly, M. (2023). The self-perception and political biases of ChatGPT. arXiv. https://doi.org/10.48550/arXiv.2304.07333

Schelble, B. G., Flathmann, C., Musick, G., McNeese, N. J., & Freeman, G. (2022). I see you: Examining the role of spatial information in human-agent teams. Proceedings of the ACM on Human-Computer Interaction, 6, 374. https://doi.org/10.1145/3555099

Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of ChatGPT. arXiv. https://doi.org/10.48550/arXiv.2301.08653

Specker, E., Tinio, P. P. L., & Van Elk, M. (2017). Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. Psychology of Aesthetics, Creativity, and the Arts, 11, 265–275. https://doi.org/10.1037/aca0000107

Stoker-Walker, C. (2022). AI bot ChatGPT writes smart essays — should professors worry? Nature. https://doi.org/10.1038/d41586-022-04397-7

Tabone, W., Happee, R., García, J., Lee, Y. M., Lupetti, M. L., Merat, N., & de Winter, J. (2023a). Augmented reality interfaces for pedestrian-vehicle interactions: An online study. Transportation Research Part F: Traffic Psychology and Behaviour, 94, 170–189. https://doi.org/10.1016/j.trf.2023.02.005

Tabone, W., Happee, R., Yang, Y., Sadraei, E., García, J., Lee, Y. M., Merat, N., & De Winter, J. (2023b). Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment. Manuscript in review.

Tate, T. P., Doroudi, S., Ritchie, D., Xu, Y., & Warschauer, M. (2023). Educational research and AI-generated writing: Confronting the coming tsunami. EdArXiv. https://doi.org/10.35542/osf.io/4mec3

Thorp, H. H. (2023). ChatGPT is fun, but not an author. Science, 379, 313. https://doi.org/10.1126/science.adg7879

Trukenbrod, A. K., Backhaus, N., & Thomaschke, R. (2020). Measuring subjectively experienced time in usability and user experience testing scenarios. International Journal of Human-Computer Studies, 138, 102399. https://doi.org/10.1016/j.ijhcs.2020.102399

Van Dis, E. A. M., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. Nature, 614, 224–226. https://doi.org/10.1038/d41586-023-00288-7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. arXiv. https://doi.org/10.48550/arXiv.1706.03762

Vincent, J. (2022). AI-generated answers temporarily banned on coding Q&A site Stack Overflow. https://www.theverge.com/2022/12/5/23493932/chatgpt-ai-generated-answers-temporarily-banned-stack-overflow-llms-dangers

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. arXiv. https://doi.org/10.48550/arXiv.2203.11171

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), Advances in neural information processing systems 35 (pp. 24824–24837). Curran Associates.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv. https://doi.org/10.48550/arXiv.2302.11382

Yang, J., Prabhakar, A., Narasimhan, K., & Yao, S. (2023). InterCode: Standardizing and benchmarking interactive coding with execution feedback. arXiv. https://doi.org/10.48550/arXiv.2306.14898

Zhang, X., & Simeone, A. L. (2022). Using the think aloud protocol in an immersive virtual reality evaluation of a virtual twin. Proceedings of the 2022 ACM Symposium on Spatial User Interaction, Online CA. https://doi.org/10.1145/3565970.3567706

Zhao, T., & McDonald, S. (2010). Keep talking: An analysis of participant utterances gathered using two concurrent think-aloud methods. Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, 581–590, Reykjavik, Iceland. https://doi.org/10.1145/1868914.1868979

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT. arXiv. https://doi.org/10.48550/arXiv.2302.10198

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2023). A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT. arXiv. https://doi.org/10.48550/arXiv.2302.09419

6

Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment

This chapter was submitted as: *Tabone, W.*, Happee, R., Yang, Y., Sadraei, E., García, J., Lee, Y. M., Merat, N., & De Winter, J. C. F. (2023). Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment.

Abstract

Augmented reality (AR) has been increasingly studied in transportation, particularly for drivers and pedestrians interacting with automated vehicles (AVs). Previous research evaluated AR interfaces using online video-based questionnaires but lacked human-subject research in immersive environments. This study examined if prior online evaluations of nine AR interfaces could be replicated in an immersive virtual environment and if AR design effectiveness depends on pedestrian attention allocation. Thirty participants completed 120 trials in a CAVE-based simulator with yielding and non-yielding AVs, rating AR interface intuitiveness and crossing the road when safe. To emulate visual distraction, participants had to look into an attention-attractor circle that disappeared 1 second after the AR interface appeared. The results showed that intuitiveness ratings from the CAVE and previous online study correlated strongly ($r \approx 0.90$). Head-locked interfaces and familiar designs (augmented traffic lights, zebra crossing) yielded higher intuitiveness scores and quicker crossing initiations than vehicle-locked interfaces. Vehicle-locked interfaces were less effective when the attention-attractor was on the environment's opposite side, while head-locked interfaces were relatively unaffected by attention-attractor position. In conclusion, this 'AR in VR' study showed strong congruence between intuitiveness ratings in a CAVE-based study and online research, and demonstrates the importance of interface placement in relation to user gaze direction.

6.1 Introduction

Road accidents rank among the top ten causes of human fatalities in middle-income countries, as reported by the World Health Organisation (2020). Academia and industry have been researching a number of ways to curb this problem, such as through the use of smart infrastructure (Banks et al., 2018; Sewalkar and Seitz, 2019; Toh et al., 2020) and connectivity. The future urban environment will likely be more connected than today (Alam et al., 2017), and traffic partners may communicate with each other in order to facilitate traffic flow (Cao et al., 2022), solve right of way conflicts (Li et al., 2019), and improve road safety for vulnerable road users, including pedestrians.

A specific solution is the use of external human-machine interfaces on automated vehicles (eHMIs), which could take the form of projections onto the road, LED strips, LED screens, anthropomorphic elements, amongst other examples (see Bazilinskyy et al., 2019; Dey, Habibovic, Pfleging, et al., 2020; Rouchitsas & Alm, 2019, for reviews of such interfaces). Although eHMIs have been shown to effectively convince VRUs whether or not to cross in front of approaching vehicles (De Clerq et al., 2019), there exist a number of drawbacks related to current designs, including the use of text, especially if the language is unfamiliar, cases where the eHMI needs to signal to a single pedestrian in a group, cultural differences in interpretability, and lack of standardisation across designs (De Winter & Dodou, 2022; Rasouli & Tsotsos, 2020; Tabone et al., 2021a; Tabone et al., 2023; Weber et al., 2019).

Augmented Reality (AR) has been proposed as a new type of communication in traffic, and as a possible solution to the aforementioned problems with eHMIs. In particular, AR offers the possibility of sending a customised signal to an individual pedestrian in a group (Tabone et al., 2021b; Tabone et al., 2023; Tran et al., 2022). Most AR studies so far have been conducted around the driver as the user (Calvi et al., 2020; Colley et al., 2021; Currano

et al., 2021; Kim et al., 2018; Mukhopadhyay et al., 2023; Pichen et al., 2020), or as a navigation assistant to VRUs (Bhorkar, 2017; Dancu et al., 2015; Dong et al., 2021; Ginters, 2019; Tran et al., 2023). However, more recently, studies have explored the use of AR as an assistive technology for pedestrians, to guide them in making safe crossing decisions by including road projections of zebra crossings, arrows, and safe paths (Hesenius et al., 2018; Li et al., 2022; Pratticò et al., 2021; Tran et al., 2022; Tran et al., 2023), visualisation of obstructed vehicles, or the collision time and conflict points (Matviienko et al., 2022; Tong & Jia, 2019; Von Sawitzky et al., 2020), and car overlays (Tran et al., 2022).

Nine novel AR interfaces for pedestrian-AV interaction were previously developed in Tabone et al. (2021b) to support pedestrians in crossing an urban road. These interfaces were designed using an experience-based approach through a theoretically-informed brainstorming session and based on expert perspectives extracted from Tabone et al. (2021a). An expert ('genius')-based design method was used (Saffer, 2010), where the designers drew upon known set theories such as predictive aiding, ecological interface design (Kardar & Shaw 2000), redundancy gain, and the proximity compatibility principle, amongst others (Wickens et al., 2004). Each AR interface was designed with two states (non-yielding, and yielding) to represent whether a vehicle would stop or not to that pedestrian, since the goal of these interfaces are to assist pedestrians in the decision to cross the road in front of an approaching AV.

Tabone et al. (2023) assessed these AR interfaces through an online questionnaire completed by 992 respondents in Germany, the Netherlands, Norway, Sweden, and the United Kingdom. The nine interfaces were recorded in a virtual reality (VR) environment and presented as videos to the respondents, who rated the interfaces for their intuitiveness and convincingness, aesthetics, and usefulness. Moreover, respondents were asked to provide free-text comments to further support their choices. Results indicated preference towards interfaces that incorporated familiar elements from existing traffic, as well as head-up displays (HUDs). These insights were possible through the statistical and qualitative thematic analyses, which also revealed a number of unintended effects of certain designs.

Despite the rich body of information that was extracted from the online questionnaire study, one of the limitations of Tabone et al. (2023), was that it did not offer high ecological validity and presented only low perceived risks to the participants. A possible solution to this problem would be to use a VR simulation method that embodies the participant. The use of an immersive environment is important to test the distributed attention of participants, since in real traffic, accidents do occur because, similar to drivers, pedestrians may fail to look at the right object, at the right time (Lanzer et al., 2023; Lee, 2008; Ralph, & Girardeau, 2020).

Two possible VR simulation methods are to use a head-mounted display (HMD) or a CAVE (CAVE automatic virtual environment). An advantage of a CAVE setup is that it allows participants to see their bodies as they move around (Cordeil et al., 2017). Previous experiments with pedestrians in a CAVE investigated pedestrians' overreliance on AVs equipped with eHMIs (Kaleefathullah et al., 2022) and crossing behaviour of pedestrians on a road with continuous traffic (Kalantari et al., 2023; Mollaro et al., 2016). In this study, we extend the research of Tabone et al. (2023) by assessing the nine AR interfaces delineated in Tabone et al. (2021b) within a CAVE-based pedestrian simulator, with the objective to examine the generalizability of findings from online questionnaire studies.

There exist fundamental differences among various types of Augmented Reality (AR)

designs (Arena et al., 2022; Carmigniani et al., 2011). Some AR designs are head-locked, wherein the message displayed follows the gaze of the participant. This feature ensures that the message is always visible to the pedestrian, allowing the user to benefit regardless of where the pedestrian is looking (Tabone et al., 2021a). Alternatively, AR systems may be positioned on the road infrastructure (Hesenius et al., 2018). The supposed advantage of this approach is that the interface can be found at known and expected locations (e.g., on the other side of the road, on the road surface). However, a disadvantage of this approach is that the AR interface may be overlooked when the participant glances left or right before deciding to cross the road. Finally, AR interfaces may be locked to the vehicle, just like eHMIs are. This approach has the advantage that the AR interface is congruent with the vehicle's motion, such that the pedestrian can process the implicit communication of the vehicle concurrently with the explicit AR signal. Furthermore, this approach eliminates the need for participants to distribute their attention towards multiple elements. However, a possible downside of vehicle-locked interfaces is that the vehicle must be identified before the pedestrian can benefit from the AR interface. For example, if a vehicle approaches from the right while the participant happens to be looking to the left and does not immediately scan to the right, identifying the AR interface may be inefficient.

In this study, we study effects of initial visual attention, using a novel technique to guide the pedestrian's initial attention towards specific regions of the road before the arrival of the vehicle. Our approach involved simulating the behaviour of pedestrians who initially fail to observe the approaching vehicle, as attention resources are finite (e.g., Wickens et al., 2004; Ralph & Girardeau, 2020), and individuals cannot attend to the entire traffic scenario simultaneously. Specifically, we used an attention-attractor circle in the form of circles placed either on the left, front, or right side of the scene to investigate our hypothesis about whether the effectiveness of the AR interface type would interact with the participant's initial attentional position. Specifically, the attention attractor was to be looked at for 1 s before any interface would appear in the environment. The aforementioned method was used to emulate and enforce initial distraction towards a certain region of the environment.

The aim of the experiment is to examine the effects of the nine different AR interfaces (previously described in Tabone et al., 2021b; Tabone et al., 2023) on pedestrian crossing behaviour and perceived intuitiveness, relative to a no-AR baseline condition, as well as their relative effects on each other. It is expected that the quality ratings observed in Tabone et al. (2023a) will be replicated in the present immersive CAVE-based environment. Moreover, the nine AR interfaces are anticipated to be regarded as clearer than the no-AR baseline conditions.

Additionally, the current study investigated the effect of the above-mentioned position-based attention attractor on pedestrian crossing behaviour and perceived intuitiveness for different AR interfaces. It is hypothesised that there will be an AR/attention-attractor position interaction, where head-locked AR interfaces (e.g., *Nudge HUD*, *Pedestrian lights HUD*) will be perceived as more intuitive when the attention-attractor is presented on the left relative to the otherwise mapped interfaces, while infrastructure-locked (road-mapped) AR interfaces (e.g., *Augmented zebra crossing*, *Fixed pedestrian lights*, *Virtual fence*) will be perceived as more intuitive when the attention-attractor is presented in the centre location of the CAVE. Finally, vehicle-locked AR interfaces (e.g., *Planes on vehicle*, *Conspicuous and looming planes*, *Field of safe travel*, *Phantom car*) are expected to be perceived as more

intuitive when the attention-attractor circle is presented to the right.

6.2 METHODS

6.2.1 PARTICIPANTS AND RECRUITMENT

Thirty participants (20 identified as male, 9 identified as female, and 1 was unspecified), aged between 22 and 53 (M = 31.50, SD = 7.98) were recruited for the study. The 30 participants were of 12 different nationalities, namely British (10), Chinese (5), Greek (3), Indonesian (3), Malay (2), German (1), Maltese (1), Norwegian (1), Romanian (1), Saudi (1), Turkish (1), and Zimbabwean (1). From the participant pool, 56.7% (n = 17) indicated that they had never been in a CAVE, while 43.3% (n = 13) indicated that they had. Further general characteristics of the 30 participants, which were collected in the demographic survey, are available in the Appendix A2.

Participants were recruited using an opportunistic sampling approach. Internal emails were sent to a pool of people interested in participating in the University of Leeds Driving and Pedestrian Simulation studies, and to various schools at the University of Leeds. Moreover, adverts were posted on a Facebook group consisting of students studying at the same university, and a further group composed of residents of Leeds.

Criteria for participating in the study were highlighted in both the emails and adverts. Specifically, participants were only eligible to participate if they were over the age of 18, had a good command of English, did not suffer from severe mobility issues, and did not suffer from epilepsy, claustrophobia, or feelings of disorientation. Moreover, we asked participants to wear lenses if they had prescription glasses, and if they had long hair, to tie it back on the day of the experiment due to the eye tracking equipment.

Participants were able to select an available time slot using an online calendar system which was linked to the experimenter's and simulator's calendars. The experiment ran for two weeks between June and July 2022, and the participants were financially compensated with a £15 Amazon gift voucher for their time spent, which was roughly 60–90 minutes. The experiment was approved by the University of Leeds Research Ethics Committee under ethics reference number LLTRAN-150.

6.2.2 APPARATUS

The experiment was created using Unity v.2020.3.35f1 (Unity, 2022) and performed in the Highly Immersive Kinematic Experimental Research (HIKER) simulator located at the University of Leeds (University of Leeds, 2022). The HIKER is a 9 × 4 m Cave Automatic Virtual Environment (CAVE) simulator, composed of eight 4K high resolution (3840 × 2400 px) projectors, and 10 Vicon Vero 2.2 IR cameras, which were calibrated and controlled using Vicon Tracker 3.9. The simulator supports both stereo and mono modes. For this experiment, mono mode was used.

Gaze data were sampled at 50 Hz and collected using the Tobii Pro Glasses 2 (firmware 1.25.6-citronkola-0; head unit 0.062) mobile eye-tracker, which was operated and calibrated using the Tobii Controller Software v.1.114.20033. Participants' verbal statements were recorded using an Olympus VP-20 microphone. A Logitech web camera mounted on a tripod was used to record the entire experiment per participant in low resolution. High-resolution video clips for dissemination were recorded using a GoPro Hero 10 camera and an iPhone

13 Pro.

The entire experiment was run on an eight-computer rack, with seven Image Generator (IG) machines brandishing an Intel®Core™i9-7900X CPU @ 3.30 GHz, 32 GB RAM, and 8 GB Nvidia Quadro P6000, and the host machine equipped with Intel®Core™i9-7900X CPU @ 3.30 GHz, 128 GB RAM, and 8 GB Nvidia Quadro P4000.

6.2.3 AUGMENTED-REALITY INTERFACES

The interfaces that were evaluated in this simulator study were adopted from the designs in Tabone et al. (2021b), and identical to the VR implementation used in Tabone et al. (2023a). In total, nine AR interfaces were designed and developed as functioning AR prototypes (Tabone et al., 2021b).

In general, the interfaces were split into three categories: interfaces that were mapped to the road, to the vehicle, and the user's head position (ie. HUDs). There are three interfaces that were mapped to the road:

- *Augmented zebra crossing* (Labelled 1 in Figure 6.1), which is a conventional traditional zebra crossing
- *Fixed pedestrian traffic lights* (5), which depicts a familiar pedestrian traffic light design across the road, and a
- *Virtual fence* (6), which displays semi-translucent walls around a zebra-crossing and a gate that opens in the yielding state.

The interfaces that were mapped to the vehicle include:

- *Planes on the vehicle* (2), which displays a plane on the windshield area of the vehicle,
- the *Conspicuous looming planes* (3), which is a scaling version of (2), as it grows or shrinks as the vehicle approaches the pedestrian depending on the AV's yielding state,
- the *Field of safe travel* (4) which projects a field on the road in front of the vehicle to communicate safety, and
- the *Phantom car* (7), which projects the vehicle's predicted future motion.

The final category of interfaces are head-up displays that are head-locked to the user's head position, i.e., they follow and remain in the user's field of view. These are:

- the *Nudge HUD* (8), which displays text ('SAFE TO CROSS' vs. 'DANGER! VEHICLE IS APPROACHING' with a corresponding icon on a green or red background, and
- the *Pedestrian lights HUD* (9), which displays a head-locked version of the pedestrian traffic lights.

Interfaces mapped to the road or the vehicle were positioned within the environment or attached to the vehicle, while the HUDs moved with the participant's camera view. The *Nudge HUD* measured 65 cm in width, and 20 cm in height and was situated 2.5 m away

from the participant, while the *Pedestrian lights HUD* measured 20 cm in width and 40 cm in height. The bottom edge of the HUDs aligned with the participant's eye level. In comparison to the online questionnaire study, the dimensions of the HUDs were reduced due to the restricted field of view (FOV) in the CAVE. The red (RGB: 244, 0, 0) and green (RGB: 32, 244, 0) colours selected for the non-yielding and yielding states were chosen based on their high intuitiveness score ratings for signalling 'please (do not) cross' (Bazilinskyy et al., 2020).

6.2.4 SCENARIO DESIGN

The current study adopted the road environment and vehicle behaviours from Kaleefathullah et al. (2022) and Tabone et al. (2023). The AV spawned out of sight (Figure 6.2, point A) and moved at a constant speed of 48 km/h (30 mph). The AR interfaces were activated when the AV was 43 m away in virtual space (Figure 6.2, point B) from the participant, who was located at point E. For non-yielding states, the vehicle maintained its initial speed of 48 km/h throughout. In contrast, for yielding states, the vehicle started decelerating at Point C, which was 33 m from Point E (the participant), at a rate of 2.99 m/s2, and came to a complete stop at Point D, located 3 m from Point E.

In each trial, an attention-attractor circle was presented in the shape of a stationary open circle with cyan border (RGB: 0, 255, 255), which was presented in three different positions: left, front-centre (i.e., across the street), and to the right of the participant, at the start of each trial (Figure 6.3). Cyan was chosen as the colour since it has been suggested to be considered neutral, and without any meaning in current traffic (Bazilinskyy et al., 2020). The circle would disappear if the participant looked into the area enclosed by its circumference for 1 second. The detection of whether the participant was looking into the circle was conducted through the head-tracker. If this condition was met, the trial commenced.

Two seconds later, the vehicle arrived at point B (see Figure 6.2) and the AR interface was activated (Figure 6.4iii). If the condition was an interface in yielding state, the vehicle started decelerating at point C (see Figure 6.2). The circle disappeared 1 s after the AR interface was activated (see Table 6.1 for the general timings of the trials). If this constraint was not applied, participants may have looked at the car as soon as possible since they expect the car to appear from behind the corner of the road, or because they could see it in their peripheral vision.

6.2.5 EXPERIMENTAL DESIGN

The experiment was a within-subject design, consisting of 120 trials per participant: 90 trials with a yielding AV and 30 trials with a non-yielding AV. More specifically, the yielding AV trials consisted of 10 AR interface conditions (i.e., including the Baseline conditions) x 3 circle positions x 3 repetitions, while the non-yielding AV trials comprised 10 AR interface conditions x 3 circle positions x 1 repetition. Each interface condition was presented in a block. Therefore, ten blocks were presented, and a counterbalancing technique was applied. Each block consisted of 12 trials (4 yielding/non-yielding conditions x 3 circle positions) which were presented in a randomised order.

After each trial, a question was projected on the centre screen of the CAVE. Participants had to verbally indicate their agreement from 1 (Strongly disagree) to 7 (Strongly agree), with the statement: *"The interface was intuitive for signalling: 'Please (do NOT) cross the*

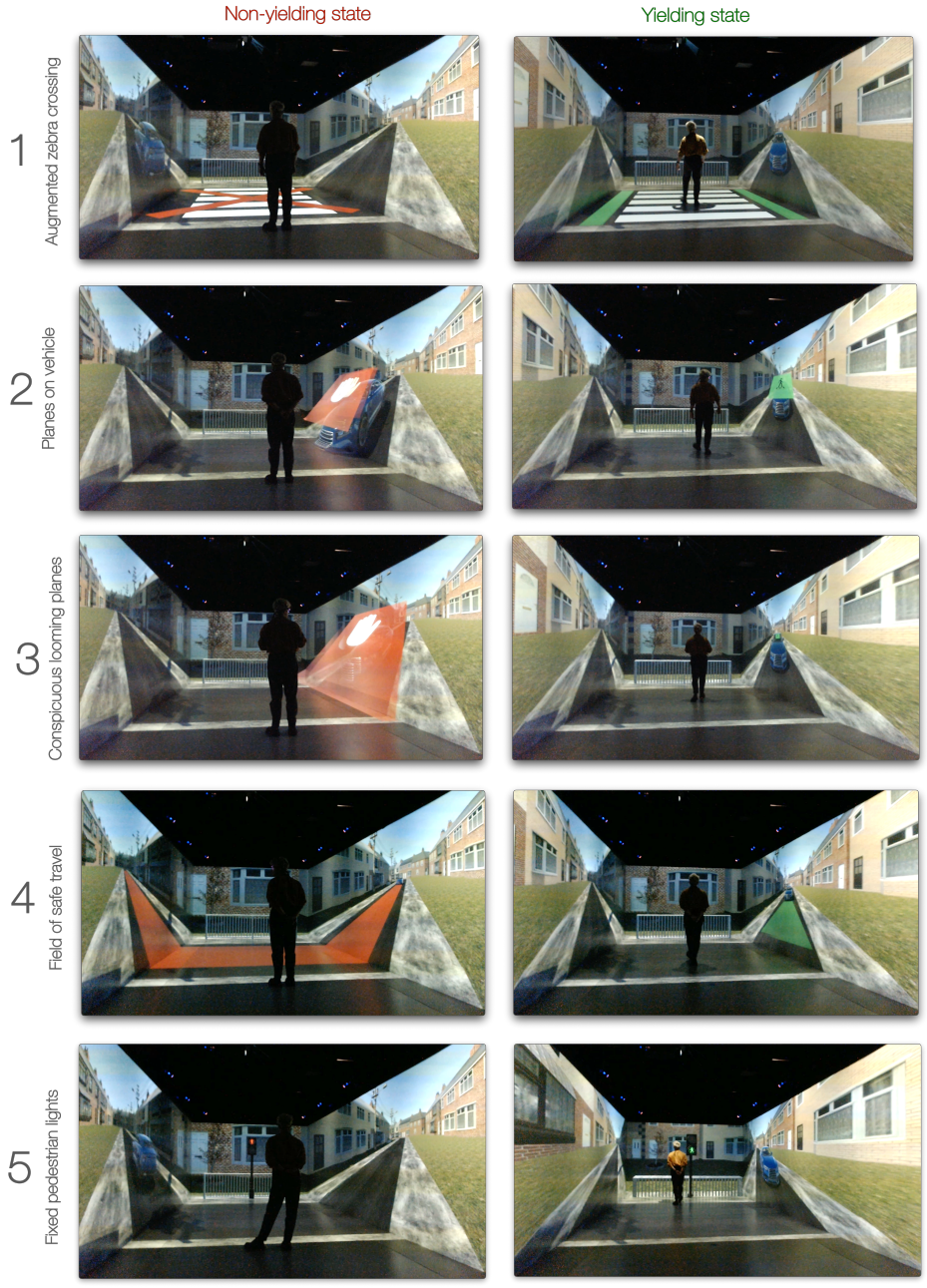


Figure 6.1: The nine AR concepts presented in the HIKER environment. Interfaces 1, 5, and 6 are projected on the road surface, 4 and 7 are also projected on the road surface, but mapped to the vehicle, while Interfaces 2 and 3 are projected on the vehicle. Interfaces 8 and 9 are head-locked.

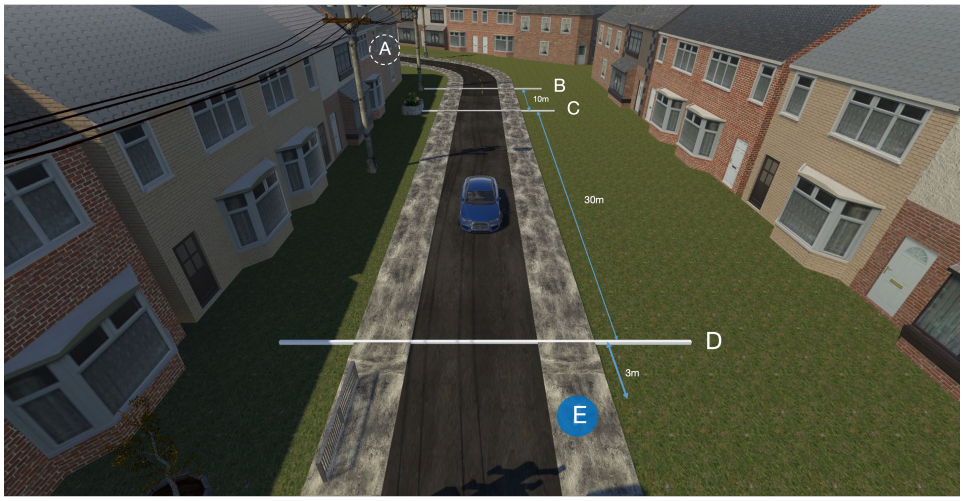


Figure 6.2: The simulation environment used for the HIKER experiment. Each salient point is demarcated by a label, together with the distance (in metres) between each point. A: spawn point, B: AR interface onset, C: AV deceleration onset, D: stopping point, E: participant location. Image adapted from Tabone et al. (2023).



Figure 6.3: A participant is seen standing on the blue marker and looking at the cyan attention-attractor circle, which is displayed in the centre position.

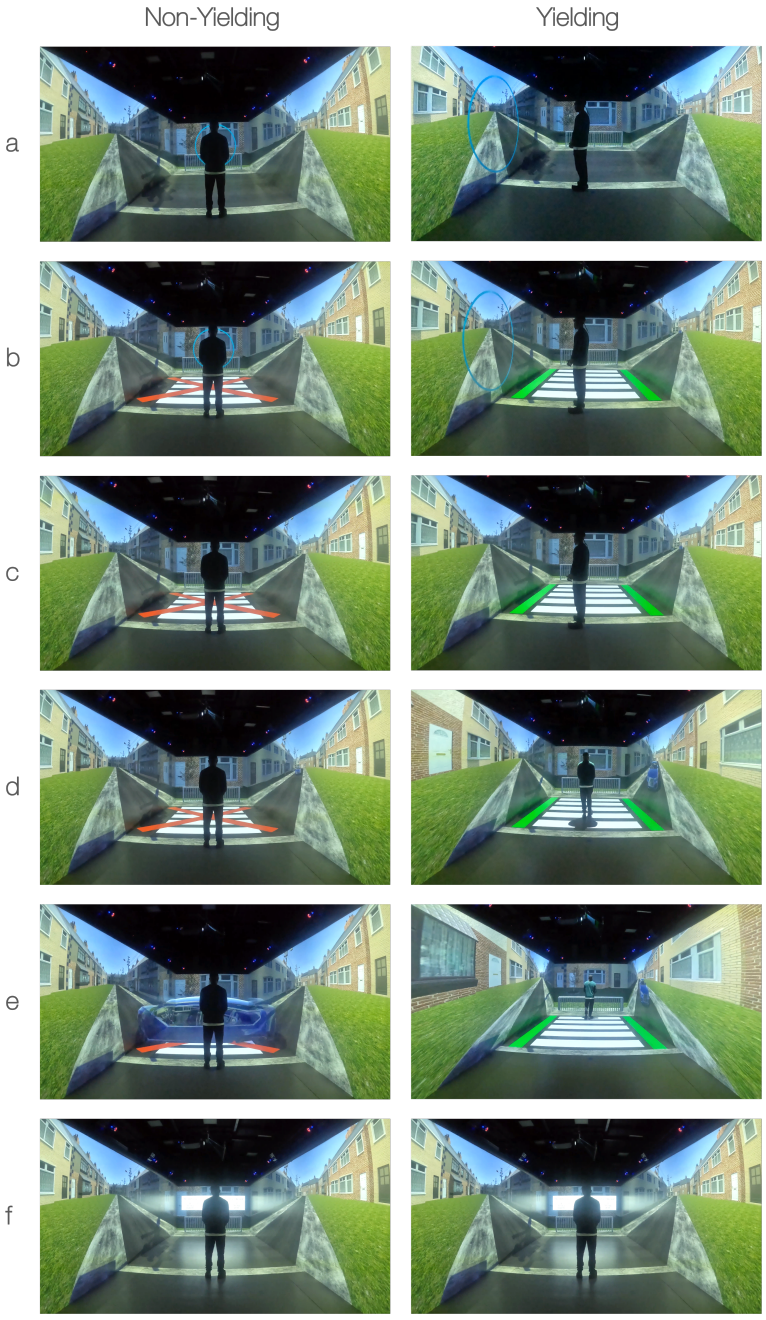


Figure 6.4: Timeline of the crossing experiment, with the left column representing the non-yielding condition, and the right column the yielding condition.

Table 6.1: Timings of events of trials, centred around the moment the circle disappeared. The letters in parentheses correspond to the positions in Figure 6.2.

Event	Elapsed time (s)	Elapsed time (s)
	Non-yielding condition	Yielding condition
Attention-attractor circle appears	-9.0	-9.0
Attention-attractor circle looked at for 1 s and vehicle starts driving (A)	-8.0	-8.0
Vehicle appears from behind curve	-3.0	-3.0
AR interface appears (B)	-1.0	-1.0
Vehicle starts to decelerate (C)	—	-0.2
Circle disappears	0.0	0.0
Vehicle comes to a halt (D)	—	3.8
Vehicle passes participant (E)	2.3	—
Intuitiveness question appears	7.0	8.8

road'". In the case of the Baseline condition, which featured no AR interface, the word, 'interface' was replaced by *'situation'*. Figure 6.5 demonstrates various moments in a trial.

6.2.6 EXPERIMENTAL PROCEDURE

Participants were welcomed into the room containing the CAVE simulator by an experimenter. An information sheet containing the experiment protocol (also previously sent to the participant by email) was presented. If the participant agreed with all the information, they were asked to sign the consent form. Participants were reminded that they can choose to stop the experiment at any time. Participants were then asked to complete a demographics questionnaire, which was developed using the Qualtrics XM (Qualtrics, 2022) survey platform. Details of the questionnaire flow are presented in Appendix 6.5.

Following the demographics questionnaire, the participants were asked to wear overshoe covers to protect the HIKER flooring, and the Tobii eye tracker system, which was also equipped with infrared markers to track the head position and correctly align the CAVE to the participant's FOV. The eye tracker was subsequently calibrated, and the participant walked into the simulator and stood on a blue marker demarking the starting position for each trial (Figure 6.4i).

The experimenter then reminded the participant of some key points highlighted in the information sheet. They were asked to look into the cyan circle for 1 consecutive second (Figure 6.4ii). If the participant violated this rule, a beeping sound was automatically produced to draw the attention of the participant back to the circle. If the participant adhered to the instructions, the vehicle would come out from behind the corner.

Furthermore, participants were reminded that they might encounter various suggestive interfaces, and that they were to cross the entire road (from one curb to the other) only if they felt that it was safe to do so (Figure 6.4iv, right). If the vehicle drove past them (Figure 6.4v left), then they no longer needed to cross. Once they had crossed the road, and the intuitiveness question was displayed (Figure 6.4vi), they were allowed to walk back to the starting position and read out their answer there. The participants were also reminded about

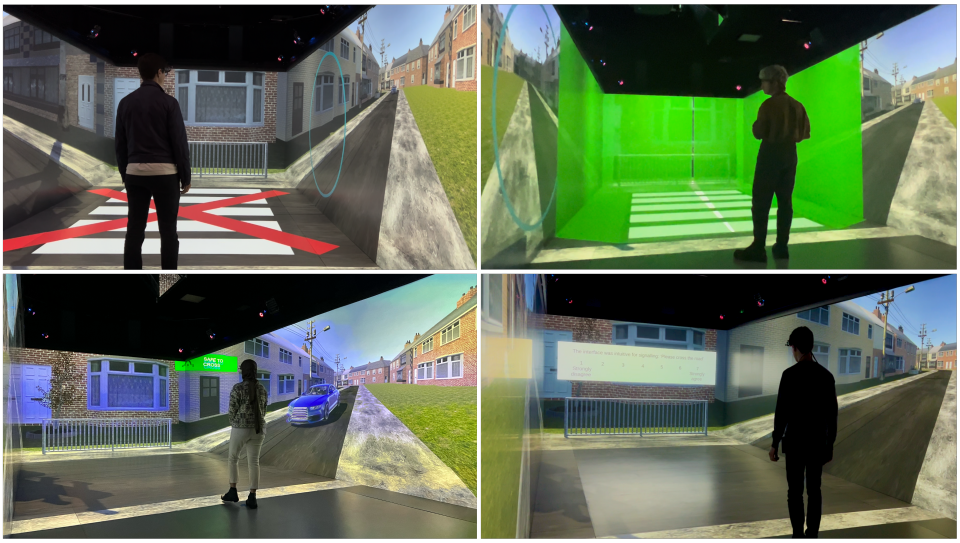


Figure 6.5: *Top-left:* The non-yielding state of *Augmented zebra crossing* is projected just before the attention-attractor circle in the right position disappears. *Top-right:* Yielding state of *Virtual fence*, projected before the circle in the left position disappears. *Bottom-left: Nudge HUD* in a yielding state is seen projected in the CAVE. *Bottom-right:* The intuitiveness rating scale is displayed at the end of a trial.

a digital mesh that appears across the nearest CAVE wall if they get too close, to warn participants to stop walking and avoid crashing into a physical barrier.

For improved understanding, two practice trials of the baseline conditions, one with a non-yielding vehicle and one with a yielding vehicle, were conducted prior to the initiation of the main experiment. When the participant indicated that they had understood the procedure, the experimenter announced that the actual experiment was about to begin. At that point, the participant was asked to return to the blue marker, and the experimenter started the experiment.

Participants then completed 120 trials. After each block, a 3-minute interview was recorded with the participant. The interview typically began by asking the participant whether they were comfortable. If the participant mentioned any form of discomfort, then the MISC scale was administered. Next, they were asked *"what did you think of this particular interface/situation?"*, and the participant was prompted to elaborate further on their answers. The line of questioning about the interface continued by asking the participant about preference between the red and green states, and whether their crossing decisions were based on the interface. Participants were corrected if they began commenting about the VR environment rather than the interfaces themselves.

Following all the trials, the participant was invited back to the table where they had signed the consent form. There they were presented with a sheet containing a table with screenshots of all the nine interfaces in both their yielding and non-yielding states side by side. The participant was asked to assign a rank from 1 to 9 next to each interface according to their preference, with 1 being the most preferred, and 9 being the least preferred. Each number could only be assigned once. When the participant had finished with the ranking,

they were thanked and rewarded for their time.

6.2.7 DATA LOGGING

The vehicle's position and speed were logged during the experiment at a frequency of 120 Hz, to produce a total of 3600 log files. For the gaze data analysis (logged at 50 Hz) and head-tracking data (logged at 120 Hz) the VR environment was segmented into a number of areas of interest:

- **Road1, Road2, Road3**: We used Road1 (near distance), Road2 (medium distance), and Road3 (far distance) segments (see Figure 6.6) to explore the distribution of the participants' attention as the AV approaches. Given that the AV approached from the right in all trials, no areas of interest to the left were created.
- **Road**: the rest of the road.
- **Car**: main body of the vehicle (Figure 6.6, inset).
- **Circle**: the region encompassed by the cyan attention-attractor circle.
- **AR Concept**: the region where the AR interface is projected.
- **HUDReg**: the region where the HUD interfaces (8 and 9) are projected. The HUD region follows the participant's head rotation and is always in front of the participant, 2.5 m ahead, and 0.25 m above the participant's head.
- **TrafficLightReg**: the region where the *Fixed Traffic Lights* would be projected, i.e., a static region in front of the pedestrian.
- **WindScrReg**: the windscreen region on the vehicle (Figure 6.6, inset).
- **Other**: any other non-segmented area in the CAVE, or outside of it.

These regions were active for all tested conditions, except the *AR Concept* region for the Baseline condition. Therefore, some segmented regions were prioritised over others when a gaze or head vector intersected in a region where multiple segmented layers may have been overlaid on one another. Priorities were as follows: The AR Concept region took precedence over all other regions, followed by the *Car*, *WindScrReg*, *TrafficLightReg*, *HUDReg*, *Road1*, *Road2*, *Road 3* ,*Circle* , *Road*, and *Other*, in the order presented.

6.2.8 PROCESSING OF POST-BLOCK INTERVIEWS

The post-block interviews were analysed using a novel approach, outlined by Tabone and De Winter (2023). Specifically, the 300 post-block voice recordings were transcribed using Otter.ai (2023), an AI tool which offers an audio to text transcription service. The transcripts were exported as text files where the text was automatically split by the different speakers, and timestamps by Otter.ai. Each transcript was then automatically submitted to OpenAI's ChatGPT-4 API (version: March 14, 2023). The API's temperature setting, which controls the level of the output's randomness, was set to 0.

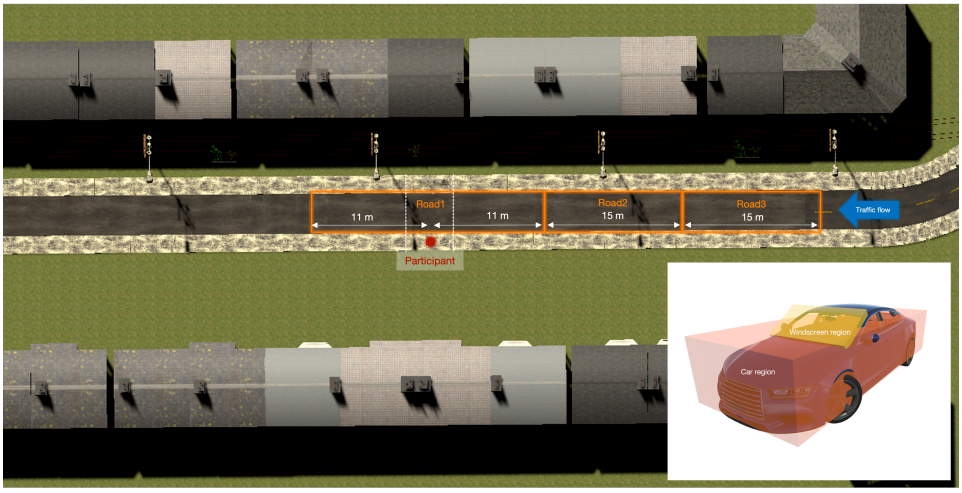


Figure 6.6: Top-down view of the environment. The road was segmented into three different regions for logging purposes. Inset: Two collider regions labelled as 'Car' and 'WindScrReg' were superimposed over the car 3D model in the environment. The labels were logged when the head-pose and eye-tracker vectors intersected with the colliders.

The following prompt was used to summarise each transcript file, encouraging ChatGPT to extract multiple strengths and weaknesses from the interview: *"Based on the participant's responses in this interview, report three strengths and three weaknesses about the AR interface. Start the strengths with "STRENGTHS:" and start the weaknesses with "WEAKNESSES:"*. Once all transcripts were summarised, the output summaries per interface (i.e., all summaries of an interface, combined) were submitted together to the ChatGPT API and summarised once again using the following prompt: *"Based on all the summaries below, summarise the strengths and weaknesses of the interface concisely. Do not overly mention general characteristics such as "the interface is intuitive", but report specific aspects that could aid designers. Keep it short, specific, and interesting, with a maximum of 4 sentences in total. Start the strengths with "STRENGTHS:" and start the weaknesses with "WEAKNESSES:"*.

6.2.9 DEPENDENT MEASURES

The following dependent measures were calculated:

- ***First glance at the AR interface***: this measure was calculated using the Tobii eye-tracker data. Ray tracing was applied to determine at first moment the gaze vector intersected with the AR interface. This measure was not available for the Baseline condition.

- ***Participant crossing initiation time***: the measure was computed for each trial in which the AV yielded. The crossing initiation time was calculated by subtracting the moment the participant's position in the CAVE environment exceeded a set threshold

(corresponding to the edge of the road) from the moment at which the attention-attractor circle disappeared.

- ***Intuitiveness ratings***: the self-reported intuitiveness rating was available for each trial.
- ***Interface ranking***: the mean rank, per participant per AR interface was calculated from the responses to the post-experiment ranking questionnaire. In addition, the distribution of the ranking was also analysed from a produced matrix (Appendix A3).
- ***Sentiment score***: the post-block questionnaire transcript files were also analysed to generate a sentiment score, using the same ChatGPT prompt which analysed the sentiment for the online questionnaire respondent open-question responses (Tabone & De Winter, 2023): *"Looking at the participant's responses, score the interface, from 1 to 100. Only report a number between 1 and 100, rounded to two decimals"*. Note that the prompt provides no dimension, such as 'intuitiveness', an approach which allows GPT-4 to provide a generic sentiment rating (Tabone & De Winter, 2023).

In addition to calculating the above dependent measures, we plotted a number of graphs to better understand how participants used the AR interfaces. In order to visualise where the participants looked at during the trials, all timings were centred and normalised around the moment that the attention-attractor circle disappeared. Hence, all time values were rounded to the nearest multiple of the sampling rate (120 Hz). Then, for each 0.01 s timestamp between -9.0 s and 9.0 s, the counts for each time a gaze region was gazed at (eye-tracking) were registered. These gaze counts against time were converted to percentages (%) and plotted against time (s) for a number of conditions (such as the yielding state, and the circle position). The produced plots would provide a visualisation of where participants distributed their gaze across areas of interest (AOI) in the CAVE.

6.2.10 STATISTICAL ANALYSIS

In order to judge the similarity of the simulator results with the obtained results from the online questionnaire study, the Pearson product-moment correlations of means for the AR interfaces were calculated. Moreover, to understand the effect of the attention-attractor circle on the dependent variables, a two-way repeated measures ANOVA of crossing initiation times and the time of the first glance towards the AR interface was conducted, with the AR interface and circle position as within-subject factors. Within-subject confidence intervals of the means were also generated according to a method presented by Morey (2008). Pairwise comparisons between the three circle positions were conducted using paired t-tests. To account for multiple comparisons, a Bonferroni correction was applied, resulting in an adjusted alpha value of 0.05 / 3 (given that there were three pairs for comparison).

6.3 RESULTS

ALL 3600 trials were completed successfully. There were only three instances where blocks had to be restarted because of technical disruptions (e.g., the eye-tracker switched off). However, in each case, the previous intuitiveness scores were retained, and any missing log data re-recorded. Hence, there was no missing data by the end of the experiment.

6.3.1 Objective 1: Replication of Intuitiveness Ratings: CAVE study vs. Online Study

Figure 6.7 shows scatter plots reporting the mean intuitiveness scores of each interface for both the CAVE simulator experiment, and the online questionnaire study (Tabone et al., 2023). Since the online study did not have a Baseline condition, its mean intuitiveness score (yielding vehicle: 4.80 [*SD* = 1.37], non-yielding vehicle: 5.48 [*SD* = 1.53]) was omitted from the plot.

The mean intuitiveness score from the HIKER study correlated strongly with the mean intuitiveness score from the online study ($r = 0.91$ for yielding vehicles, and $r = 0.90$ for non-yielding vehicles, respectively, n = 9). Similar to the online questionnaire study, the *Nudge HUD*, *Fixed pedestrian lights*, *Augmented zebra crossing*, *Pedestrian lights HUD*, and *Virtual fence* interfaces scored the highest, while the *Phantom car* yielded the lowest intuitiveness rating.

Table 6.2: Correlation matrix for measures relating to the online questionnaire study (Tabone et al., 2023), and the simulator study conducted in the HIKER

	1	2	3	4	5	6
1. HIKER: Intuitiveness (non-yielding vehicle)						
2. HIKER: Intuitiveness (yielding vehicle)	0.73					
3. HIKER: ChatGPT sentiment score	0.85	0.82				
4. HIKER: Mean preference rank	-0.77	-0.90	-0.94			
5. HIKER: Crossing initiation time	-0.51	-0.87	-0.59	0.63		
6. Online: Intuitiveness (non-yielding vehicle)	0.90*	0.83	0.86	-0.71	-0.77	
7. Online: Intuitiveness (yielding vehicle)	0.83	0.91*	0.90	-0.79	-0.90	0.94

Note: n = 10 for measures 1, 2, 3, and 5. n = 9 for measures 4, 6, and 7.
∗ These two correlation coefficients are depicted in scatter plots in Figures 6.7 and 6.8.

6.3.2 Objective 2: AR Interface/Attention-Attractor Position Interaction

Crossing Initiation Time

According to a repeated-measures ANOVA of the crossing initiation time, with interface condition and circle position as within-subject factors, there was a significant effect of interface condition, $F(9, 252) = 17.98$, $p < 0.001$, partial $\eta^2 = 0.39$, a significant effect of circle position, $F(2, 56) = 56.163$, $p < 0.001$, partial $\eta^2 = 0.67$, and a significant interface condition × circle position interaction, $F(18, 504) = 2.13$, $p < 0.001$, partial $\eta^2 = 0.22$.

Figure 6.9 shows the means and 95% confidence intervals of the crossing initiation times for all AR interfaces. Among the road-mapped interfaces, the *Fixed pedestrian lights* (which were always positioned on the opposite side of the street) demonstrated optimal performance when the attention-attractor circle was also presented centrally. Regarding the *Augmented*

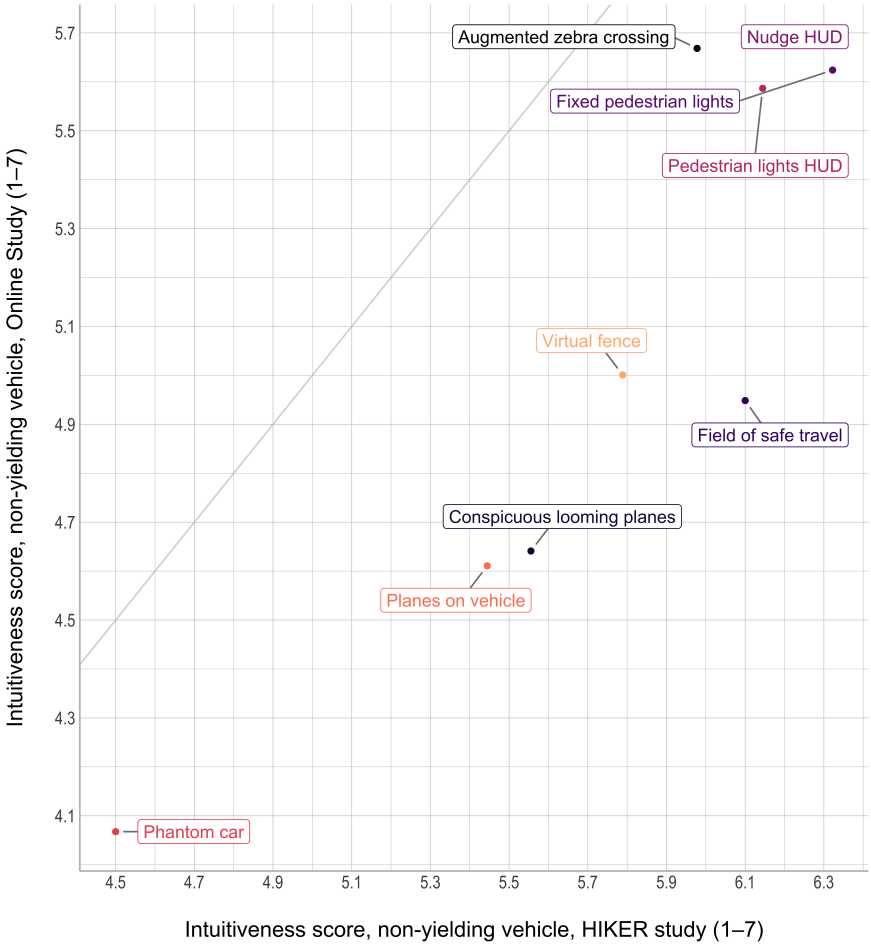


Figure 6.7: Scatter plot of the mean intuitiveness rating per AR interface in their non-yielding state, as reported in the current HIKER experiment and the previous online questionnaire study (Tabone et al., 2023).

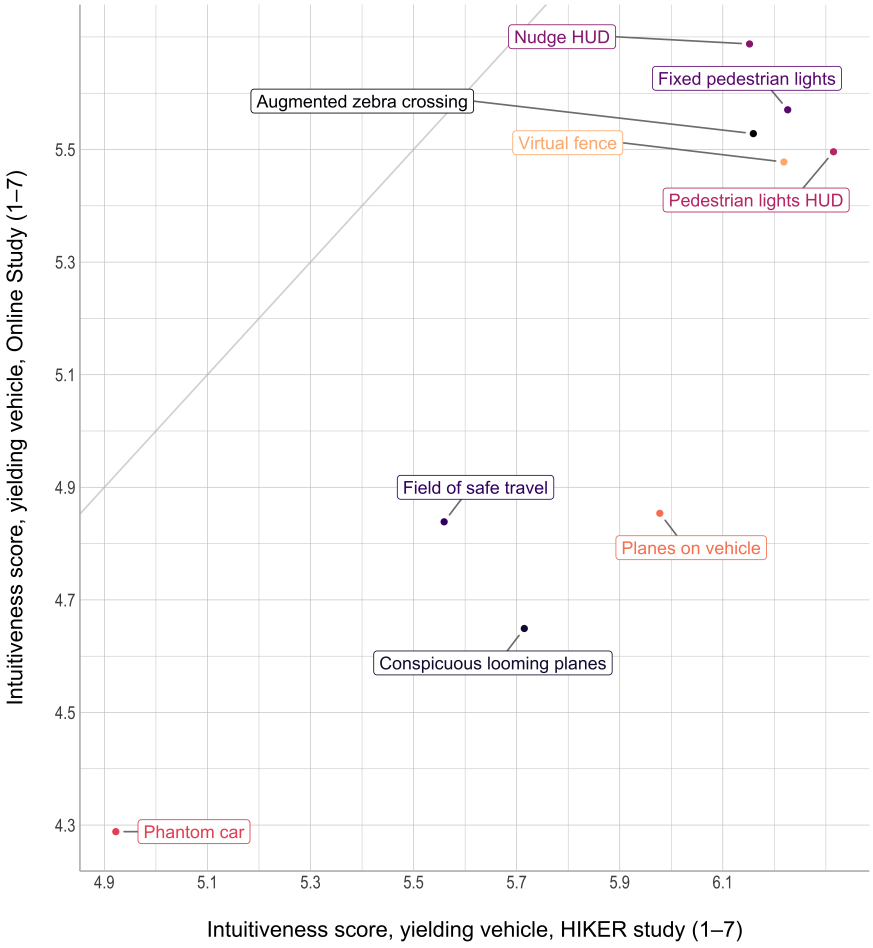


Figure 6.8: Scatter plot of the mean intuitiveness rating per AR interface in their yielding state, as reported in the current HIKER experiment and the previous online questionnaire study (Tabone et al., 2023).

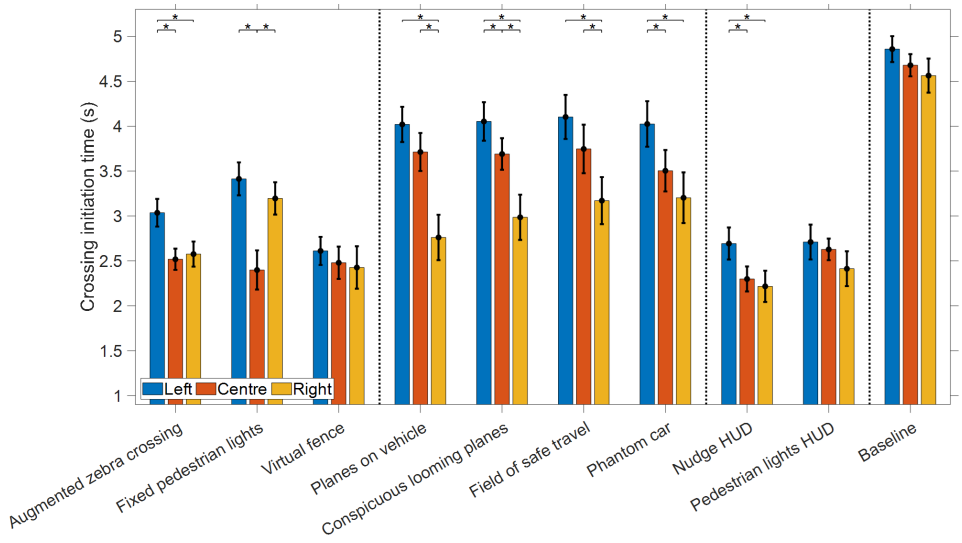


Figure 6.9: Bar plot of the mean crossing initiation time as affected by the attention-attractor position, for each AR interface. Vertical lines delineate the road-mapped interfaces, vehicle-mapped interfaces, HUDs, and baseline condition.

zebra crossing, the central and right circle presentation positions exhibited faster crossing initiation times as opposed to the left presentation, while the *Virtual Fence* performance remained relatively consistent irrespective of the circle presentation position.

Concerning the vehicle-mapped AR interfaces, a consistent trend emerged, wherein superior performance was observed when the circle was presented on the right—the direction from which the AV approached.

Lastly, with respect to the HUDs and *Baseline* condition, a discernible pattern emerged, where the left circle presentation was less advantageous compared to central or right presentations. However, this effect was not statistically significant for the *Baseline* condition and *Pedestrian lights HUD* and was considerably smaller in magnitude than the vehicle-mapped interfaces.

TIME OF FIRST GLANCE AT AR INTERFACE

According to a repeated-measures ANOVA of the time of the first glance at the AR interface, with interface condition, and circle position as within-subject factors, showed a significant effect of interface condition, $F(8, 200) = 21.23$, $p < 0.001$, partial $\eta^2 = 0.46$, a significant effect of circle position, $F(2, 50) = 36.58$, $p < 0.001$, partial $\eta^2 = 0.59$, and a significant interface condition × circle position interaction, $F(16, 400) = 15.30$, $p < 0.001$, partial $\eta^2 = 0.38$.

As a manipulation check for the experiment, the results presented in Figure 6.10 generally correspond with the crossing initiation times shown in Figure 6.9. Specifically, when the attention-attractor circle was centrally positioned, the *Fixed pedestrian lights* were the first to be glanced at, while the vehicle-mapped interfaces drew attention more quickly when the circle was displayed on the right as opposed to the centre or left. Regardless of the circle

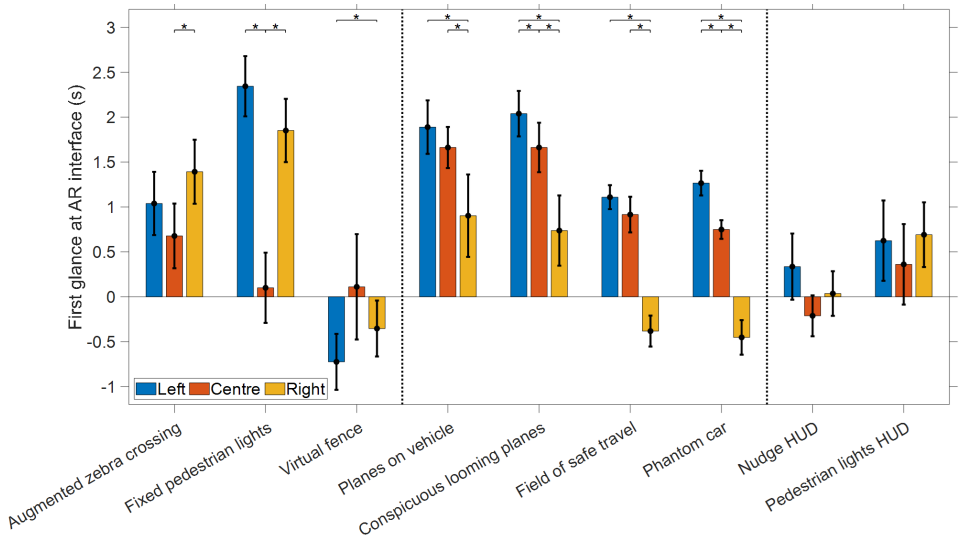


Figure 6.10: Bar plot of the moment in which the participants first glanced at the AR interface, as affected by the attention-attractor position, per AR interface condition. Vertical lines delineate the road-mapped interfaces, vehicle-mapped infaces, and HUDs. Out of a total of 3240 trials (9 AR interfaces × 12 trials × 30 participants), participants did not glance at the AR interface in 213 instances. This may potentially be attributed to inaccuracies in the eye-tracker, or to the genuine lack of attention directed toward the AR interface by the participants.

position, the HUDs were promptly noticed. Furthermore, the *Virtual fence* quickly captured attention, which can likely be attributed to its significant size.

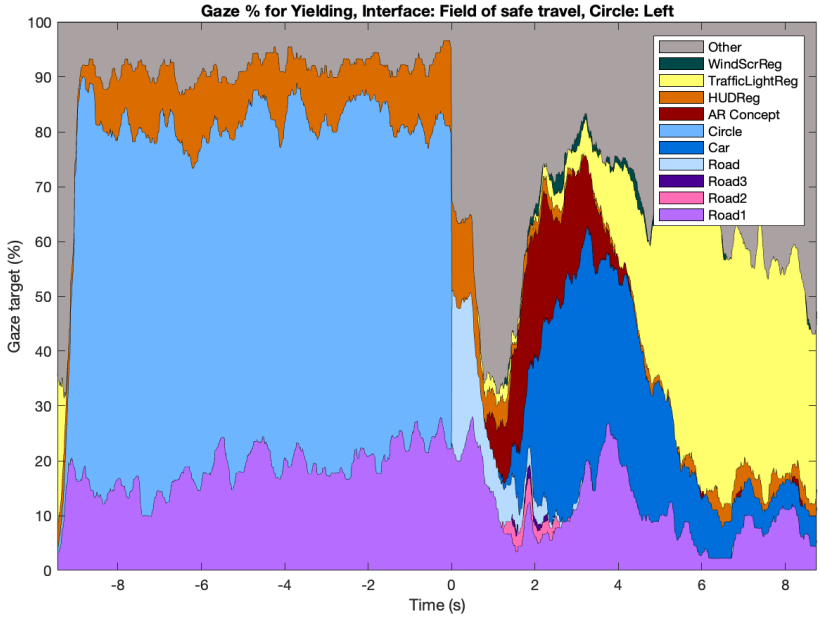
Distribution of Gaze across AOIs

Figure 6.11 shows a sample of four gaze plots, pertaining to the yielding states of the *Field of safe travel* (a, b), and the *Pedestrian lights HUD* (c, d). In each case, the area plot depicts the gaze distribution across time for the different areas of interest. The two interfaces were chosen as they clearly demonstrate the effect of interface placement on the participants' attention distribution.

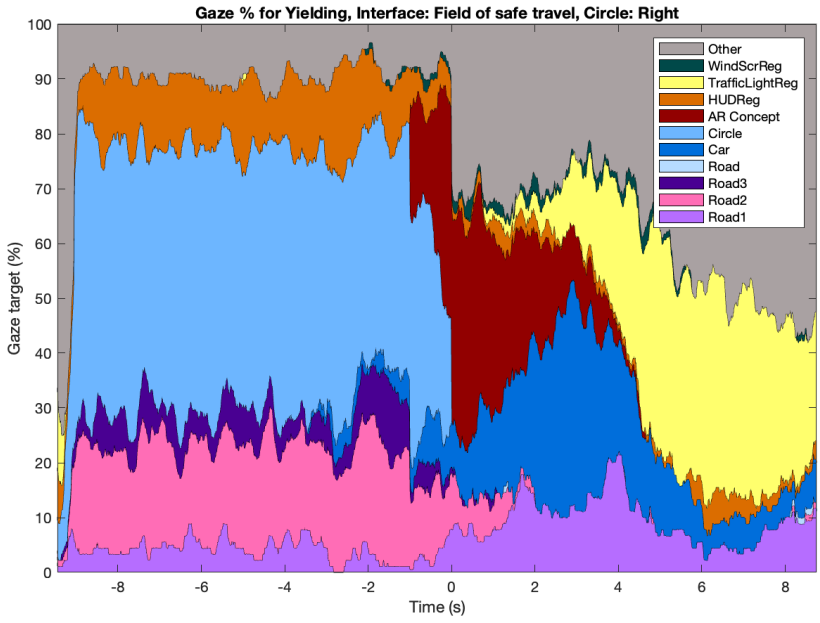
Figure 6.11 (a, c) shows the gaze distribution when the circle was presented on the left, while Figure 6.11 (b, d) corresponds to the circle presented on the right. When the circle was presented to the right, the participants were able to direct their gaze to the approaching vehicle ("Car") and the presented interface, i.e., *Field of safe travel* ("AR Concept") much earlier after time 0.0 s, compared to when the circle was presented to the left. In that case, there was a delay of over 1 second.

On the other hand, in the case of the *Pedestrian lights HUD*, the AR interface was gazed at before 0.0 s in both the left and right cases. These results confirm the advantage of the HUD, since it follows the user's gaze as a head-locked augmentation, whereas the *Field of safe travel* only has an advantage when a participant starts the trial while looking to the right.

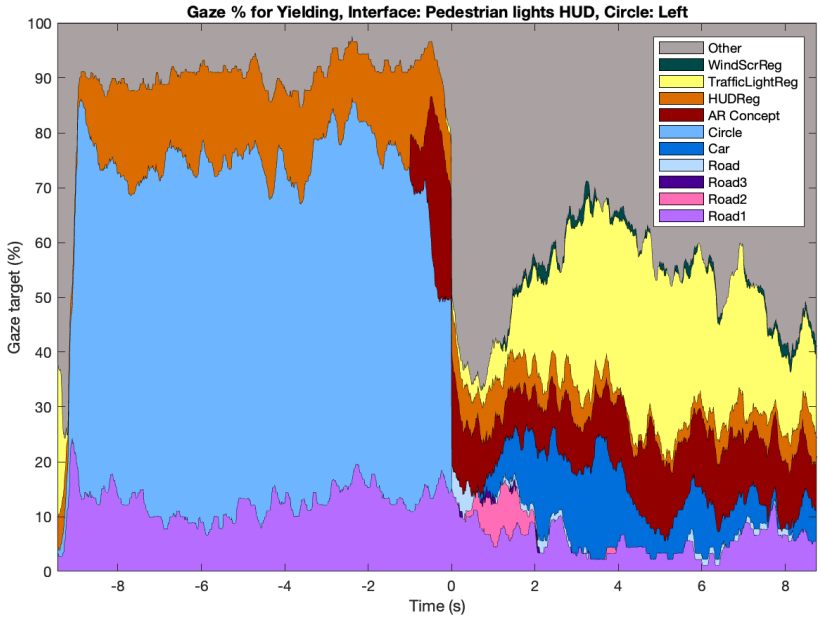
All the other gaze and head tracking plots for the other conditions can be found in the Supplementary Repository made available as a complement to this paper (see Appendix, section 6.5).


Gaze % for Yielding, Interface: Field of safe travel, Circle: Left
Gaze target (%)
Time (s)
Other
WindScrReg
TrafficLightReg
HUDReg
AR Concept
Circle
Car
Road
Road3
Road2
Road1

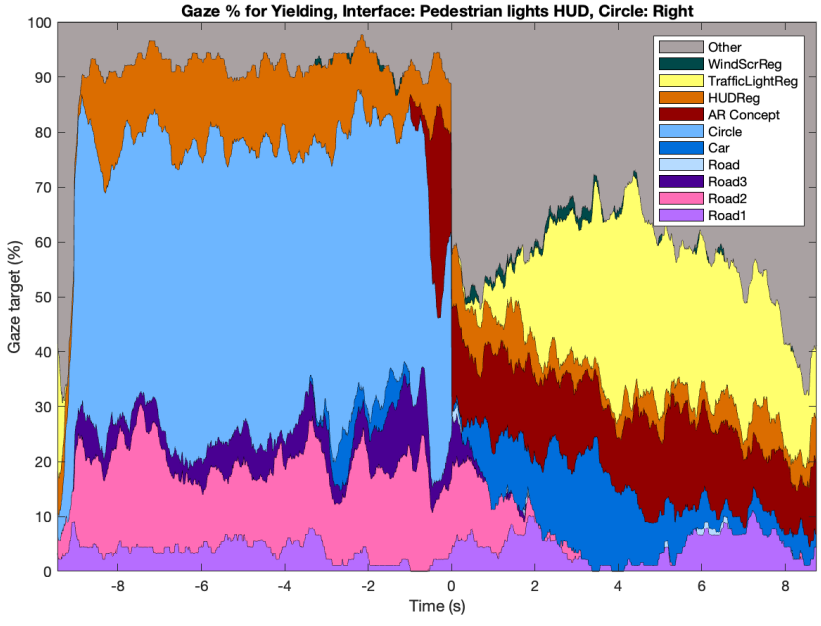
(a)


Gaze % for Yielding, Interface: Field of safe travel, Circle: Right
Gaze target (%)
Time (s)
Other
WindScrReg
TrafficLightReg
HUDReg
AR Concept
Circle
Car
Road
Road3
Road2
Road1

(b)



(c)



(d)

Figure 6.11: Plots of the gaze percentage for each area of interest region in the environment. The figures above are for the yielding-state condition of the *Field of safe travel* interface, with the attention-attractor circle projected to the left (a), and right (b); and for the *Pedestrian lights HUD* with the attention-attractor circle projected to the left (c), and right (d). Time = 0 corresponds to the moment that the circle disappeared.

6.3.3 POST-BLOCK INTERVIEWS

Table 6.3 presents the outputs of the prompts from the Chat GPT-4 analysis, where the model was asked to output the strengths, and weaknesses of each interface based on 30 transcripts. The initial summaries can be found in the Supplementary Repository (link in Appendix, section 6.5).

The overall summaries offered a fingerprint of what the participants said after each block. In general, interfaces mapped to the road (*Augmented zebra crossing*, *Fixed pedestrian lights*, and *Virtual fence*) were found to elicit confidence due to the participants' familiarity with common traffic designs. Moreover, the *Virtual fence*'s walls elicited a "sense of safety". On the other hand, the participants pointed out that such interfaces may cause users to focus too much on the design, rather than the surroundings, as the view behind the tunnel walls somewhat occludes the environment, despite being semi-translucent . Such familiar designs may also make the user overly reliant on the interface, or cause a false sense of security.

Interfaces mapped to the vehicle (*Planes on vehicle*, *Conspicuous looming planes*, *Field of safe travel*, *Phantom car*) were generally described as providing clear and distinguishable coloured cues. In fact, the use of looming in *Conspicuous looming planes* was described as effectively communicating urgency, and capturing the user's attention. However, these interfaces were also described as having the potential to not be immediately intuitive for the user, cause problems for colour-blind users in distinguishing the different states, confusion with animations (such as the scaling in *Conspicuous looming planes*), and ineffectiveness if not looking in the direction of the approaching vehicle. In general, participants felt that this group of interfaces had a large learning curve.

The HUD interfaces (*Nudge HUD*, *Pedestrian lights HUD*) were described as "providing a seamless experience" since the interface follows the user's gaze, which made the experience convenient, thus "boosting user confidence in crossing decisions". On the other hand, the HUDs were sometimes considered confusing, uncomfortable, and distracting, requiring a period of adjustment. Moreover, participants commented on the lack of customisation options for the HUD interfaces.

In the *Baseline* condition without any AR interface, participants stated that the interfaces helped provide "clear visual cues and signals that help users make informed decisions when crossing the road", which "can improve users' confidence and awareness while crossing". However, the participants also pointed out that users of the AR system might "become overly reliant on AR interfaces, which may decrease their ability to judge traffic situations independently", or the interface may not provide enough information or guidance in particular situations, which could ultimately cause confusion or hesitation.

Table 6.3: Interview summaries by ChatGPT-4. In the summaries for the Baseline condition, an 'AR interface' is still alluded to. The reason is that the ChatGPT prompt only mentioned 'AR interfaces'.

Interface	Strengths	Weaknesses
Augmented zebra crossing	The AR interface is generally easy to understand, with clear and familiar signals such as green and red colours. It provides a sense of confidence and safety for users, effectively communicating when it is safe or unsafe to cross the road.	The interface may cause users to focus too much on the signals and not enough on their surroundings. It is less effective when users are not looking straight ahead, and the green signal is often less visible or intuitive than the red signal.
Planes on vehicle	The AR interface generally provides clear and easily distinguishable visual cues, such as colour coding and symbols, which aid users in making informed decisions about crossing the road. The interface is often intuitive and can be learned over time, increasing user confidence and understanding of vehicle intentions.	The AR interface may not be immediately intuitive for all users, requiring an adjustment period. Some participants found certain aspects of the interface, such as the red signal, less effective or attention-grabbing. The interface may not account for all possible scenarios or adapt well to different contexts, potentially leading to confusion or misinterpretations.
Conspicuous looming planes	The AR interface effectively communicates urgency and safety through colour coding (red and green) and size changes, capturing users' attention and building trust over time. The interface is generally intuitive and easy to understand, with some users preferring it over other options.	The interface may not be immediately intuitive for first-time users, and its reliance on colour could be problematic for colorblind users or in situations with poor visibility. The green interface is often less effective and noticeable than the red one, and inconsistencies in design elements, such as scaling, can cause confusion and distraction for users.

Interface	Strengths	Weaknesses
Field of safe travel	The red signal in the AR interface is highly visible, easily understandable, and effective in conveying safety information. The colour-coded design is generally simple for users to comprehend, and the interface has the potential to improve pedestrian safety.	The green signal is less visible, less effective, and can cause confusion for users. The interface may require a learning curve for users to fully understand its purpose and meaning. The reliance on colour cues may not be effective for users with colour vision deficiencies, and the interface's effectiveness may vary in different lighting conditions or environments.
Fixed pedestrian lights	The AR interface is familiar, intuitive, and easy to understand, using recognizable traffic light symbols and colours. It provides clear signals and guidance for users to make informed decisions about crossing the road.	The interface may be distracting, difficult to notice, or require users to adjust their perspective, potentially impacting safety and decision-making. Users may become overly reliant on the interface, ignoring real-world cues, and the effectiveness of the interface may vary depending on lighting, weather conditions, and individual preferences.
Virtual fence	The AR interface effectively communicates safety through clear visual cues, such as red and green walls, and is generally intuitive and easy to understand. It provides a sense of safety and confidence for users when crossing the road and is adaptable to various situations.	The interface can be visually overwhelming and obstruct the view of vehicles, potentially causing confusion or a false sense of security. It may not cater to individual preferences or be suitable for color-blind users. Additionally, the reliance on the AR interface may lead to a lack of personal responsibility in making safe decisions while crossing the street.

Interface	Strengths	Weaknesses
Phantom car	The AR interface's colour coding (green for safe, red for danger) effectively communicates safety levels and aids decision-making. The red signs are particularly helpful and comfortable, providing insight for crossing the road. The interface can improve user alertness and engagement, potentially enhancing safety.	The interface can be initially confusing and may have a learning curve for new users. The green virtual vehicle is less intuitive and can be misleading, while the red car's display can be distracting and difficult to judge. The interface's effectiveness is dependent on the user's focus and direction, potentially causing confusion and reducing its overall utility.
Nudge HUD	The AR interface is generally intuitive, with clear colour coding (red and green) and text for guidance. It follows the user's eye movement, providing a seamless experience and boosting user confidence in crossing decisions.	The interface may cause initial discomfort or confusion, and users may still hesitate to trust it fully, relying on real-world cues. There is room for improvement in terms of simplicity, distinct visual cues, and catering to different user preferences and needs.
Pedestrian lights HUD	The AR interface is generally clear, easy to understand, and familiar, using traffic light symbols and colours. It effectively communicates safety information and follows the user's gaze, making it convenient and efficient.	Some participants found the interface initially confusing, distracting, or disorienting. There is a lack of customization options and the interface may not be universally preferred. Trust in the interface varies, with some users preferring to rely on their own judgement or finding certain aspects (e.g., colours or symbols) less intuitive.

Interface	Strengths	Weaknesses
Baseline	The AR interface is generally user-friendly, providing clear visual cues and signals that help users make informed decisions when crossing the road. It simulates real-life traffic situations and can improve users' confidence and awareness while crossing.	Users may become overly reliant on the AR interface, potentially decreasing their ability to judge traffic situations independently. The interface may not provide enough information or guidance in certain situations, leading to confusion or hesitation. Additionally, the simplicity of the simulation may not accurately represent complex real-world scenarios, limiting the effectiveness of the training.

6.4 DISCUSSION

THIRTY participants completed 120 trials each in a CAVE pedestrian simulator experiment that assessed nine AR interfaces for pedestrian-vehicle interaction, and a baseline condition without any interface. The participants were asked to cross the virtual road if they felt it was safe to do so, and to rate how intuitive the interfaces were in communicating their message of intent. In addition to the standard head-tracker used in the CAVE, participants were equipped with an eye-tracker, which recorded their point of gaze during the experiment. Additionally, participants were asked about each interface in an audio-recorded interview following exposure. The data were analysed to assess the intuitiveness of each interface, the willingness to cross the road, the gaze and head direction of participants, and the further information provided during the interviews.

The first objective of the experiment was to evaluate the effect of nine AR interfaces on pedestrian crossing behaviour and perceived intuitiveness, with the expectation of replicating the intuitive ratings observed in a previous online questionnaire (Tabone et al., 2023). The second objective was to examine the effect of the position of the attention attractor on pedestrian crossing behaviour and perceived intuitiveness across different AR interfaces. An AR interface/ attention-position interaction was hypothesised, where head-locked, infrastructure-locked, and vehicle-locked AR interfaces would exhibit varying levels of intuitiveness based on the positioning of the attention-attractor circle.

6.4.1 OBJECTIVE 1: REPLICATION OF INTUITIVENESS RATINGS: CAVE STUDY VS. ONLINE STUDY

The present study was conducted in a highly immersive CAVE simulator, in which the participant (pedestrian) was surrounded by images displayed on the floor and walls (Cruz et al., 1992), while the scene orients according to the head position of the user. The CAVE allowed participants to see their own body (in contrast to online studies, or when using a

HMD), allowing for a more natural interaction with the environment (Blissing & Bruzelius, 2018). In our experiment, none of the 30 participants suffered any discomfort related to motion sickness despite having gone through 120 trials each. Hence, when compared to HMDs or video-based study, a CAVE offers a more natural setting (Schneider et al., 2022).

Despite the high level of immersiveness offered by the CAVE, the AR interfaces achieved similar relative differences in mean intuitiveness scores as in the online questionnaire study. In fact, a correlation of $r = 0.91$ was observed between the two measures for yielding, and $r = 0.90$ for non-yielding, indicating that the intuitiveness scores highly replicated the online study. A correlation of 0.90 is strong, especially considering the limited sample size of only 30 participants, which implies imperfect statistical reliability. As in the online study, the interfaces that used traditional traffic elements (*Augmented zebra crossing*, and *Fixed pedestrian lights*), the HUDs (*Nudge HUD*, and *Pedestrian lights HUD*), and other interfaces mapped to the road (*Virtual fence*) were rated relatively highly, while interfaces mapped to the vehicle were deemed less intuitive (*Field of safe travel*, *Planes on vehicle*, *Conspicuous looming planes*, and *Phantom car*). The behavioural measure, i.e., crossing initiation time, also exhibited a substantial correlation with the online intuitiveness ratings ($r = -0.90$).

Unintended effects were once again observed for the *Field of safe travel*, and the *Phantom car*, which were designed to adhere to the principle of ecological interface design (Tabone, Lee, et al., 2021), and the principle of predictive aiding, respectively. However, similar to the online study, in the simulator study, the Phantom car may have failed to comply with the proximity compatibility principle, as participants found it difficult to separate the Phantom car from the 'real' car (as mentioned in Table 6.3). Moreover, both interfaces only displayed a coloured element without any icon or text, and may have therefore failed to comply with the principle of redundancy gain. Given that both interfaces lacked the incorporation of traditional design elements from traffic, the application of the top-down processing principle, which relies on the recognition of familiar symbols, may have posed another challenge for participants in terms of initially comprehending the designs intuitively. Conversely, the interfaces that achieved the highest intuitiveness scores demonstrated adherence to these aforementioned principles (Tabone et al., 2021).

The remarkably strong correlation between the outcomes of the online study and those of the CAVE simulator raises questions about the necessity of conducting experiments in a resource-intensive CAVE. Based on our results, we contend that if the objective is solely to gather subjective evaluations, such as average intuitiveness ratings, preference rankings, an online questionnaire could not only be sufficient but even preferable, considering the potential for larger sample sizes (refer to Schneider et al., 2022).

Nonetheless, it is acknowledged that certain discrepancies emerged between the subjective impressions of the online study and the current CAVE-based study. For example, the HUD interfaces, which tracked participants' head motion, were experienced in real-time, whereas these motions were preprogrammed in the online study. This allowed participants to provide feedback on their functioning, which was occasionally described as jittery, annoying, or confusing. Moreover, the *Virtual fence* might create a false sense of security and potentially obstruct the view of oncoming vehicles, which are features that could not be observed in the prior online study, which used video clips. Similarly, in the preference rankings, the *Virtual fence* received a high number of first-place rankings, which could be attributed to its visually appealing animation of opening doors and its large size, potentially appealing

to participants due to a "coolness" factor.

6.4.2 OBJECTIVE 2: AR INTERFACE/ATTENTION-POSITION INTERACTION

A main reason the study was conducted in the HIKER was due to its high field of view compared to an online experiment, allowing for the guidance of the participant's initial attention allocation. An attention-attractor circle was utilised to guide the participant's attention towards a specific region of the road environment before the arrival of the vehicle, as in real-world situations, pedestrians may initially neglect to observe the approaching vehicle due to finite attentional resources (Ralph & Girardeau, 2020; Wickens et al., 2004). Hence, the effect of the position of the attention-attractor circle on pedestrian crossing behaviour and perceived intuitiveness across different AR interfaces could be explored.

The position of the attention-attractor circle did in fact have an effect on the crossing initiation time of the participants. As detailed in Figure 6.9, there was a significant delay in crossing initiation when participants were presented with an interface mapped to the vehicle while the circle was presented to the left.

Interfaces mapped to the road, in particular the *Fixed pedestrian lights*, registered a faster crossing initiation time when the attention-attractor circle was presented in the centre position, while the HUDs had similar initiation times for each of the three attention-attractor positions. For the HUDs, this could be explained by the fact that it remained in the central field-of-view of the participant, regardless of the participant's head orientation. On the other hand, the outcomes for the road-mapped interfaces were supported by the post-trial interviews, where the *Augmented zebra crossing* was found to be "less effective when users are not looking straight ahead". A surprising outcome from the interviews was that the HUDs received criticism for 'lack of customisation' and were also found to be 'distracting and disorienting'. The unintended effect was revealed because, in the CAVE, it was possible to have the interfaces follow the participant's gaze in contrast to video-based studies. However, the positive comments on the HUDs did confirm the advantage they offer in following the users' gaze.

The *Virtual fence* was also unaffected by the attention-attractor position. Two possible explanations can account for this observed phenomenon. Firstly, the *Virtual fence* was large and visually conspicuous, thereby increasing the likelihood that participants covertly perceive its presence and colour through peripheral vision, even if they do not directly glance at it. The size of the *Virtual fence* was also mentioned by participants in the interviews. Secondly, the substantial size of the virtual fence made it more probable that participants could rapidly direct their foveal vision toward it. This assertion is supported by the findings illustrated in Figure 6.10 (time of first glance) and Figure 6.9 (crossing initiation times), which demonstrated that the *Virtual fenc*e performed well in these regards. In contrast, other AR interfaces were smaller, such as the *Fixed pedestrian lights*, or moving within the scene, such as the vehicle-locked interfaces, rendering them more challenging to detect with peripheral vision (i.e., before turning the eyes) and foveal vision (i.e., by moving the eyes).

The inferior performance of the interfaces mapped to the vehicle may have implications for eHMI design, taking into account that proposed eHMIs have mostly been placed on the vehicle itself. On the other hand, the performance of the HUDs, being less affected by user distraction, makes them ideal candidates for use in such traffic interactions, because

in everyday situations, the pedestrian cannot be expected to look at an approaching vehicle right away. More so, they are also the easiest to implement in terms of sensor requirements, since the HUDs would need minimal context from the environment (i.e., only AV intent), unlike some other interfaces, such as the *Planes on Vehicle*, which in reality would need to make use of computer vision methods to detect the vehicle (Tabone et al., 2021b). At the same time, it may be important to consider the potential risk of overreliance. A more detailed exploration of the eye-tracking data revealed that in the Baseline condition, participants neglected to glance at the AV in merely 9 out of 360 trials. In contrast, for the *Augmented Zebra Crossing*, *Fixed Pedestrian Lights*, *Nudge HUD*, and *Pedestrian Lights HUD*, the respective numbers were 76, 71, 70, and 72 out of 360. This observation indicates that the AR interfaces could potentially cause crossing without proper looking (for a similar concern in eHMI research, see Kaleefathullah et al., 2022).

An intriguing observation emerged regarding the crossing initiation times for the *Baseline* condition, which appeared to be only slightly affected by the position of the attention attractor. A plausible explanation for this phenomenon is that, although the attention attractor positioned on the left and centre caused a delayed response in glancing at the approaching AV (just as it did for the nine AR interface conditions), this temporal delay is not detrimental. In the Baseline condition, no valuable information would be missed, as it is challenging to discern whether a car has initiated braking from implicit communication (i.e., vehicle speed) alone. The AR interfaces, on the other hand, were activated 0.8 seconds before the AV started to slow down. Therefore, a pedestrian would miss valuable explicit communication if they do not glance promptly at the AR interface (and see De Winter & Dodou, 2022, for a discussion on eHMIs that provide anticipatory information).

6.4.3 Limitations and Future Work

Despite the improved ecological validity when compared to the online questionnaire study in Tabone et al. (2023a), there are still some limitations to the study presented in this paper. Firstly, the CAVE and the equipment itself were sources of distraction as there were points during the interviews where the participant was commenting about the VR environment rather than the interfaces themselves.

Other limitations arose from equipment issues. The eye-tracker had a limited battery life, and there were instances when the battery needed to be replaced between blocks. Moreover, on average, 4 or 5 participants were accommodated in a single experiment day, resulting in a cumulative running time of over 9 to 10 hours. This extended duration occasionally led to equipment overheating. Also, the *Field of Safe Travel* and *Phantom Car* interfaces experienced some rubberbanding issues, which manifested as jittery and inconsistent depth perceptions. This phenomenon may have impacted the reported intuitiveness ratings to a certain degree. In fact, some post-block interviews (such as P20 and P24; see Supplementary Material) for the *Phantom Car* reported instances of abnormal behaviour, including "one of the green cars reversed" and "on a couple of cases with the green, the car went past the hologram." These issues were due to occasional synchronisation problems across screens and computers, as well as limitations of the Unity networking system. Regardless of these issues, a high correlation between the simulator and online questionnaire results was observed.

Additionally, the limited screen area of the CAVE meant that some of the interfaces had to be scaled down to fit. In fact, the size of both the HUD interfaces were reduced

from their original dimension during piloting as they became too large to fit on the CAVE screens as the participant approached the wall across (end of the virtual road). In fact, in the post-experiment subjective ranking results there were some comments about the HUD interfaces being 'too small'. Such an unintended effect was therefore caused by technical limitations of a CAVE simulator.

The AR interfaces were still projected on a virtual environment, which ultimately rendered them as part of the VR simulation. Furthermore, despite the CAVE simulation being immersive, it still did not offer real risk, and participants were asked to cross repeatedly without any real purpose other than to complete the experiment. Hence, to better understand the trust and behaviour of participants with these AR interfaces, the ecological validity must once again be increased and the AR interfaces be tested overlaid over the real world layer.

6.5 CONCLUSION

A pedestrian simulator experiment was conducted to investigate the interaction between pedestrians and automated vehicles using nine different augmented reality interfaces presented within a virtual reality CAVE. The aim of the experiment was twofold: (1) to investigate whether results from an online questionnaire video-based study could be replicated in a CAVE simulator, and (2) to investigate whether the effectiveness of different AR designs is contingent upon the pedestrian's attention allocation. The statistical and qualitative findings indicate that the mean intuitiveness ratings correlated substantially with results from a prior online study ($r \approx 0.90$). Interfaces using traditional traffic design elements, and head-locked HUDs were rated as more intuitive, in contrast to vehicle-mapped interfaces.

The position of the attention-attractor circle affected crossing initiation time, with participants initiating crossing earlier for road-mapped and HUD interfaces in general, in contrast to interfaces that were mapped to the vehicle. For the *Virtual fence* and the HUDs, crossing initiation times were relatively unaffected by the attention-attractor position. These findings are likely due to the size of the *Virtual fence* and the omnipresence of the HUD interfaces in the participant's FOV.

The poor performance of vehicle-mapped AR interfaces may have implications for eHMIs, which are typically mounted on the vehicle. Lastly, the experiment outcomes also highlight the fact that a CAVE simulator might not be necessary if the goal is to investigate intuitiveness ratings, given the high correlation with the online study. However, the immersive nature of the CAVE proved essential for studying the distributed attention of the participants.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860410.

References

Alam, F., Mehmood, R., Katib, I., Albogami, N. N., & Albeshri, A. (2017). Data fusion and IoT for smart ubiquitous environments: A survey. IEEE Access, 5, 9533–9554. https://doi.org/10.1109/ACCESS.2017.2697839

Arena, F., Collotta, M., Pau, G., & Termine, F. (2022). An overview of augmented reality. Computers, 11(2), 28. https://doi.org/10.3390/computers11020028

Banks, V. A., Stanton, N. A., Burnett, G., & Hermawati, S. (2018). Distributed Cognition on the road: Using EAST to explore future road transportation systems. Applied Ergonomics, 68, 258–266. https://doi.org/10.1016/j.apergo.2017.11.013

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. Transportation Research Part F: Traffic Psychology and Behaviour, 67, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2020). External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'? IEEE International Conference on Systems, Man and Cybernetics (SMC), 3721–3728, Toronto, Canada. https://doi.org/10.1109/SMC42975.2020.9282998

Bhorkar, G. (2017). A survey of augmented reality navigation. arXiv. https://arxiv.org/abs/1708.05006 Blissing, B., & Bruzelius, F. (2018). Exploring the suitability of virtual reality for driving simulation. In Driving Simulation Conference 2018 (pp. 163–166).

Bos, J. E., MacKinnon, S. N., & Patterson, A. (2005). Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. Aviation, space, and environmental medicine, 76, 1111–1118.

Calvi, A., D'Amico, F., Ferrante, C., & Ciampoli, L. B. (2020). Effectiveness of augmented reality warnings on driving behaviour whilst approaching pedestrian crossings: A driving simulator study. Accident Analysis & Prevention, 147, 105760. https://doi.org/10.1016/j.aap.2020.105760

Cao, D., Wang, X., Li, L., Lv, C., Na, X., Xing, Y., Li, … & Wang, F. Y. (2022). Future directions of intelligent vehicles: Potentials, possibilities, and perspectives. IEEE Transactions on Intelligent Vehicles, 7, 7–10. https://doi.org/10.1109/TIV.2022.3157049

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. Multimedia tools and applications, 51, 341–377. https://doi.org/10.1007/s11042-010-0660-6

Colley, M., Eder, B., Rixen, J. O., & Rukzio, E. (2021). Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. Proceedings of the 2021 CHI Conference on Human Factors in Computing

Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445351

Cordeil, M., Cunningham, A., Dwyer, T., Thomas, B. H., & Marriott, K. (2017). ImAxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In Proceedings of the 30th annual ACM symposium on user interface software and technology, 71–83. https://doi.org/10.1145/3126594.3126613

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. Communications of the ACM, 35, 64–73. https://doi.org/10.1145/129888.129892

Currano, R., Park, S. Y., Moore, D. J., Lyons, K., & Sirkin, D. (2021). Little road driving hud: Heads-up display complexity influences drivers' perceptions of automated vehicles. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan. https://doi.org/10.1145/3411764.3445575

Dancu, A., Vechev, V., Ünlüer, A. A., Nilson, S., Nygren, O., Eliasson, S., Barjonet, J.-E., Marshall, J., & Fjeld, M. (2015). Gesture bike: Examining projection surfaces and turn signal systems for urban cycling. Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, Madeira, Portugal, 151–159. https://doi.org/10.1145/2817721.2817748

De Clercq, K., Dietrich, A., Núñez Velasco, J. P., De Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. Human Factors, 61, 1353–1370. https://doi.org/10.1177/00187208198363

De Winter, J., & Dodou, D. (2022). External human-machine interfaces: Gimmick or necessity. Transportation Research Interdisciplinary Perspectives, 15, 100643. https://doi.org/10.1016/j.trip.2022.100643

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., & Terken, J. (2020a). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives, 7, 100174. https://doi.org/10.1016/j.trip.2020.100174

Dong, W., Wu, Y., Qin, T., Bian, X., Zhao, Y., He, Y., Xu, Y., & Yu, C. (2021). What is the difference between augmented reality and 2D navigation electronic maps in pedestrian wayfinding? Cartography and Geographic Information Science, 48, 225–240. https://doi.org/10.1080/15230406.2021.1871646

Ginters, E. (2019). Augmented reality use for cycling quality improvement. Procedia Computer Science, 149, 167–176. https://doi.org/10.1016/j.procs.2019.01.120

Hesenius, M., Börsting, I., Meyer, O., & Gruhn, V. (2018). Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Barcelona, Spain, 261–268. https://doi.org/10.1145/3236112.3236148

Ishihara, S. (1917). Tests for colour-blindness. Tokyo: Kanehara Shuppan Co., Ltd.

Jerald, J. (2016). The VR book: Human-centered design for virtual reality. Morgan & Claypool.

Kalantari, A. H., Yang, Y., Merat, N., & Markkula, G. (2023). Modelling vehicle-pedestrian interactions at unsignalised locations: Road users may not play the Nash equilibrium. Manuscript submitted for publication.

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & De Winter, J. C. F. (2022). External human–machine interfaces can be misleading: An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human Factors, 64, 1070–1085. https://doi.org/10.1177/0018720820970751

Kadar, E. E., & Shaw, R. E. (2000). Toward an ecological field theory of perceptual control of locomotion. Ecological Psychology, 12, 141–180. https://doi.org/10.1207/S15326969ECO1202_02

Kim, H., Gabbard, J. L., Anon, A. M., & Misu, T. (2018). Driver behavior and performance with augmented reality pedestrian collision warning: An outdoor user study. IEEE Transactions on Visualization and Computer Graphics, 24, 1515–1524. https://doi.org/10.1109/TVCG.2018.2793680

Lanzer, M., Koniakowsky, I., Colley, M., & Baumann, M. (2023). Interaction Effects of Pedestrian Behavior, Smartphone Distraction and External Communication of Automated Vehicles on Crossing and Gaze Behavior. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3581303

Lee, J. D. (2008). Fifty years of driving safety research. Human Factors, 50, 521–528. https://doi.org/10.1518/001872008X2883

Li, T., Guo, F., Krishnan, R., Sivakumar, A., & Polak, J. (2020). Right-of-way reallocation for mixed flow of autonomous vehicles and human driven vehicles. Transportation research part C: emerging technologies, 115, 102630. https://doi.org/10.1016/j.trc.2020.102630

Li, Y., Kamalasanan, V., Batista, M., & Sester, M. (2022). Improving pedestrian priority via grouping and virtual lanes. arXiv. https://arxiv.org/abs/2205.08783

Mallaro, S., Rahimian, P., O'Neal, E. E., Plumert, J. M., & Kearney, J. K. (2017). A comparison of head-mounted displays vs. large-screen displays for an interactive pedestrian

simulator. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. https://doi.org/10.1145/3139131.3139171

Matviienko, A., Müller, F., Schön, D., Seesemann, P., Günther, S., & Mühlhäuser, M. (2022). BikeAR: Understanding cyclists' crossing decision-making at uncontrolled intersections using Augmented Reality. Proceedings of the CHI Conference on Human Factors in Computing Systems, New Orleans, LA. https://doi.org/10.1145/3491102.3517560

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Mukhopadhyay, A., Sharma, V. K., Tatyarao, P. G., Shah, A. K., Rao, A. M., Subin, P. R., & Biswas, P. (2023). A comparison study between XR interfaces for driver assistance in take over request. Transportation Engineering, 11, 100159. https://doi.org/10.1016/j.treng.2022.100159

Otter.ai (2023). Otter.ai - Voice meeting notes & real-time transcription. https://otter.ai

Pichen, J., Yan, F., & Baumann, M. (2020). Towards a cooperative driver-vehicle interface: Enhancing drivers' perception of cyclists through Augmented Reality. Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, 1827–1832. https://doi.org/10.1109/IV47402.2020.9304621

Pratticò, F. G., Lamberti, F., Cannavò, A., Morra, L., & Montuschi, P. (2021). Comparing state-of-the-art and emerging Augmented Reality interfaces for autonomous vehicle-to-pedestrian communication. IEEE Transactions on Vehicular Technology, 70, 1157–1168. https://doi.org/10.1109/TVT.2021.3054312

Qualtrics. (2022). Qualtrics XM // The leading experience management software. https://www.qualtrics.com

Ralph, K., & Girardeau, I. (2020). Distracted by "distracted pedestrians"?. Transportation research interdisciplinary perspectives, 5, 100118. https://doi.org/10.1016/j.trip.2020.100118

Rasouli, A., & Tsotsos, J. K. (2020). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems, 21, 900–918. https://doi.org/10.1109/TITS.2019.2901817

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. Frontiers in Psychology, 10, 2757. https://doi.org/10.3389/fpsyg.2019.02757

Saffer, D. (2010). Designing for interaction: Creating innovative applications and devices. Berkeley, CA: New Riders. Sewalkar, P., & Seitz, J. (2019). Vehicle-to-pedestrian communication for vulnerable road users: Survey, design considerations, and challenges. Sensors, 19, 358. https://doi.org/10.3390/s19020358

Schneider, S., Maruhn, P., Dang, N. T., Pala, P., Cavallo, V., & Bengler, K. (2022). Pedestrian crossing decisions in virtual environments: behavioral validity in CAVEs and head-mounted displays. Human factors, 64, 1210–1226. https://doi.org/10.1177/0018720820987446

Tabone, W., De Winter, J. C. F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B., & Stanton, N. A. (2021a). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transportation Research Interdisciplinary Perspectives, 9, 100293. https://doi.org/10.1016/j.trip.2020.100293

Tabone, W., Lee, Y. M., Merat, N., Happee, R., & De Winter, J. C. F. (2021b). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Leeds, UK, 209–218. https://doi.org/10.1145/3409118.3475149

Tabone, W., Happee, R., García, J., Lee, Y. M., Lupetti, L., Merat, N., & De Winter J. C. F. (2023). Augmented reality interfaces for pedestrian-vehicle interactions: An online study. Transportation Research Part F: Traffic Psychology and Behaviour, 94, 170–189. https://doi.org/10.1016/j.trf.2023.02.005

Tabone, W., & De Winter J. C. F. (2023). Using ChatGPT for Human-Computer Interaction Research: A Primer. Manuscript submitted for publication.

Tafidis, P., Farah, H., Brijs, T., & Pirdavani, A. (2022). Safety implications of higher levels of automated vehicles: a scoping review. Transport Reviews, 42, 245–267. https://doi.org/10.1080/01441647.2021.1971794

Toh, C. K., Sanguesa, J. A., Cano, J. C., & Martinez, F. J. (2020). Advances in smart roads for future smart cities. Proceedings of the Royal Society A, 476, 20190439. https://doi.org/10.1098/rspa.2019.0439

Tong, Y., & Jia, B. (2019). An Augmented-Reality-based warning interface for pedestrians: User interface design and evaluation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63, 1834–1838. https://doi.org/10.1177/1071181319631413

Tran, T. T. M., Parker, C., Wang, Y., & Tomitsch, M. (2022). Designing wearable augmented reality concepts to support scalability in autonomous vehicle–pedestrian interaction.

Frontiers in Computer Science, 4, 866516. https://doi.org/10.3389/fcomp.2022.866516

Tran, T. T. M., Parker, C., Hoggenmüller, M., Hespanhol, L., & Tomitsch, M. (2023). Simulating wearable urban augmented reality experiences in VR: Lessons learnt from designing two future urban interfaces. Multimodal Technologies and Interaction, 7, 21. https://doi.org/10.3390/mti7020021

Unity. (2022). Unity real-time development platform | 3D, 2D VR & AR engine. https://www.unity.com University of Leeds. (2022). Highly Immersive Kinematic Experimental Research (HIKER) pedestrian lab. https://uolds.leeds.ac.uk/facility/hiker-lab

Von Sawitzky, T., Wintersberger, P., Löcken, A., Frison, A. K., & Riener, A. (2020). Augmentation concepts with HUDs for cyclists to improve road safety in shared spaces. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI. https://doi.org/10.1145/3334480.3383022

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). Crossing the Street Across the Globe: A Study on the Effects of eHMI on Pedestrians in the US, Germany and China. In HCI in Mobility, Transport, and Automotive Systems: First International Conference, MobiTAS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, 21, 515–530. Springer International Publishing. https://doi.org/10.1007/978-3-030-22666-4_37

Wickens, C. D., Gordon, S. E., & Liu, Y. (2004). An introduction to human factors engineering. Upper Saddle River, NJ: Pearson Prentice Hall.

World Health Organisation. (2020). The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Appendix

Demographics Questionnaire

A preliminary question (Q0) was added to record the participant number, which was manually filled out by the experimenter beforehand. The questionnaire began by asking participants about their identified gender (Q1), age (Q2), nationality (Q3), the length of time they have lived in the UK (Q4), and their familiarity with traffic systems (i.e., left-hand, right-hand, or both) (Q5), along with their occupation. The subsequent question (Q6) assessed the participant's affinity for technological systems, using the Affinity for Technology Interaction (ATI) scale (Franke et al., 2019). This was followed by questions regarding previous use of VR headsets (Q7), AR apps (Q8), and participation in CAVE-based simulator experiments (Q9). Participants were then asked about their primary mode of transportation (Q10) and whether they possessed a driving licence (Q11). If they answered "yes" to the latter, two additional questions about the year they obtained their licence (Q12) and the number of miles driven in the past 12 months (Q13) were presented. The final question in the demographic questionnaire included a colour blindness test (Q14) (Ishihara, 1917; as used in Bazilinskyy et al., 2020 and Tabone et al., 2023).

General characteristics of the participants

General characteristics of the 30 participants were as follows:

- 26.7% (n = 8) indicated that they had lived in the UK for 1–5 years, 3.3% (n = 1) lived in the UK between 5 and 10 years, 30% (n = 9) for less than a year, and 40% (n = 12) for more than 10 years.
- 33.3% (n = 10) were familiar with both left-hand traffic (LHT) and right-hand traffic (RHT), 43.3% (n = 13) indicated that they were familiar with only LHT, and 23.3% (n = 7) with only RHT.
- The mean total Affinity for Technology Interaction (ATI) scale score was 4.25.
- 26.7% (n = 8) had never used VR, while 73.3% (n = 22) indicated that they had used VR.
- 43.3% (n = 13) had never used AR, while 56.7% (n = 17) did.
- 56.7% (n = 17) had never been in a CAVE, while 43.3% (n = 13) indicated that they did.
- 50.0% (n = 15) indicated that their daily walking ranged between 15 and 30 minutes, 10.0% (n = 3) stated 30–45 minutes, 23.3% (n = 7) indicated 45–60 minutes, and 16.7% (n = 5) stated 60 minutes and above.
- For primary transportation, 13.3% (n = 4) use cycling, 23.3% (n = 7) use private vehicles, 16.7% (n = 5) use public transportation, and 46.7% (n = 14) mostly walk.
- 86.7% (n = 26) possess a driving licence, while 13.3% (n = 4) do not.

- From those that had a driving licence, 77% (n = 20) obtained it in the last 10 years, while 23% (n = 6) before that, with 1993 being the earliest year of attainment, and 2020 the most recent.
- 34.6% (n = 9) of participants with a licence, drive everyday, 3.8% (n = 1) drive just in the weekdays, 7.7% (n = 2) drive just in the weekends, 26.9% (n = 7) never drive, 7.7% (n = 2) drive once per week, 11.5% (n = 3) drived only once, and 7.7% (n = 2) preferred not to respond.
- From the licensed sample, 11.5% (n = 3) had driven 0 miles in the preceding 12 months, 38.5% (n = 10) had driven 1–5,000 miles, 11.5% (n = 3) had driven 10,001–15,000 miles, 7.7% (n = 2) had driven 15,001–20,000 miles, 23.1% (n = 6) had driven 5,001–10,000 miles, and 7.7% (n = 2) preferred not to respond.
- 3.3% (n = 1) of the respondent pool was considered colourblind as they had submitted three or more incorrect responses (Bazilinskyy et al., 2020; and Tabone et al., 2023) for the six-item Ishihara colour blindness test.
- During the experiment, only one participant from 30 stated that they were uncomfortable at one point in the experiment. The MISC scale was presented to the participant, but they had voted '1', which corresponds to 'some discomfort, but no specific symptoms' (Bos et al., 2005). Therefore, the experiment continued, as the score was not above 3.

SCORES OF DEPENDENT MEASURES SUPPLEMENTARY MATERIAL

In the spirit of open science, and to assist the replicability of this study, a repository has been made available at https://doi.org/10.4121/73c200c9-ee9e-423e-95a6-a22ee4935617. It contains supplementary data to this paper, such as the processed comma-separated-value files, interview transcripts, ChatGPT outputs, gaze and head-tracking plots, and the code used to generate the results presented in this paper.

Table 6.4: Means of dependent measures or the nine AR interfaces and Baseline Condition

	HIKER: Intuitiveness (non-yielding vehicle)	**HIKER: Intuitiveness (yielding vehicle)**	**HIKER: ChatGPT sentiment score**	**HIKER: Mean preference rank**	**HIKER: Crossing initiation time**	**Online: Intuitiveness (non-yielding vehicle)**	**Online: Intuitiveness (yielding vehicle)**
Augmented zebra	5.98	6.16	81.82	4.34	2.71	5.67	5.53
Planes on vehicle	5.44	5.98	79.76	4.41	3.50	4.61	4.85
Conspicuous looming planes	5.56	5.71	76.37	5.62	3.55	4.64	4.65
Field of safe travel	6.10	5.56	76.76	5.07	3.67	4.95	4.84
Fixed pedestrian lights	6.32	6.23	80.54	5.03	3.00	5.62	5.57
Virtual fence	5.79	6.22	80.60	4.34	2.51	5.00	5.48
Phantom car	4.50	4.92	69.20	7.14	3.60	4.07	4.29
Nudge HUD	6.22	6.15	81.76	4.41	2.40	5.68	5.69
Pedestrian lights HUD	6.14	6.31	82.71	4.62	2.58	5.59	5.50
Baseline	5.48	4.80	77.36	—	4.69	—	—

Table 6.5: Post-experiment questionnaire: Interface rankings

Score	**Augmented zebra crossing**	**Planes on vehicle**	**Conspicuous looming planes**	**Field of safe travel**	**Fixed pedestrian lights**	**Virtual fence**	**Phantom car**	**Nudge HUD**	**Pedestrian traffic lights HUD**
1	5	3	3	1	4	10	0	3	0
2	5	3	3	4	0	1	1	6	6
3	2	8	1	2	2	2	1	5	6
4	6	2	2	6	8	1	1	1	2
5	1	5	3	5	1	5	1	4	4
6	2	2	4	1	5	1	5	3	6
7	2	1	5	4	4	2	6	3	2
8	4	2	3	5	4	4	5	1	1
9	2	3	5	1	1	3	9	3	2

7

Augmented reality for supporting the interaction between pedestrians and automated vehicles: An experimental outdoor study

This chapter was published as: Aleva, T*., *Tabone, W*., Dodou, D., & De Winter, J. C. F. (2024). Augmented reality for supporting the interaction between pedestrians and automated vehicles: An experimental outdoor study. Frontiers in Robotics and AI, 11, 1324060.

ABSTRACT

Communication from automated vehicles (AVs) to pedestrians using augmented reality (AR) could positively contribute to traffic safety. However, previous AR research for pedestrians was mainly conducted through online questionnaires or experiments in virtual environments instead of real ones. In this study, 28 participants conducted trials outdoors with an approaching AV and were supported by four different AR interfaces. The AR experience was created by having participants wear a Varjo XR-3 headset with see-through functionality, with the AV and AR elements virtually overlaid onto the real environment. The AR interfaces were vehicle-locked (Planes on vehicle), world-locked (Fixed pedestrian lights, Virtual fence), or head-locked (Pedestrian lights HUD). Participants had to hold down a button when they felt it was safe to cross, and their opinions were obtained through rating scales, interviews, and a questionnaire. The results showed that participants had a subjective preference for AR interfaces over no AR interface. Furthermore, the Pedestrian Lights HUD was more effective than no AR interface in a statistically significant manner, as it led to participants more frequently keeping the button pressed. The Fixed pedestrian lights scored lower than the other interfaces, presumably due to low saliency and the fact that participants had to visually identify both this AR interface and the AV. Hence, while users favour AR in AV-pedestrian interactions over no AR, its effectiveness depends on design factors like location, visibility, and visual attention demands. In conclusion, this work provides important insights into the use of AR outdoors. The findings illustrate that, in these circumstances, a clear and easily interpretable AR interface is of key importance.

7.1 INTRODUCTION

EVERY year, 300,000 pedestrian deaths occur worldwide, accounting for 23% of all road fatalities (World Health Organization, 2018). The United Nations has set a target within its Sustainable Development Goals to halve the number of traffic deaths by 2030 (United Nations, 2020). In this resolution (A/RES/74/299), it was also noted that "continuous progress of automotive and digital technologies could improve road safety, including through the progressive development of highly and fully automated vehicles in road traffic" (United Nations, 2020, p. 4).

One possibility to improve the safety of pedestrians is to equip automated vehicles (AVs) with better sensors and intelligence so that they can respond earlier, in order to reduce collisions with vulnerable road users (Jungmann et al., 2020; Paiva et al., 2021). Another possibility is the introduction of AV-to-pedestrian communication technologies. Previous research has shown that with displays on the outside of the AV, called external human-machine interfaces or eHMIs, pedestrians can make more effective road-crossing decisions (Bazilinskyy et al., 2021; Bindschädel et al., 2022; Dey et al., 2020). However, eHMIs have certain disadvantages in that they typically cannot address an individual pedestrian (Colley et al., 2020; Tran et al., 2023) and that they can be difficult to perceive in some cases, for example due to an occlusion by another object (Dey et al., 2022; Troel-Madec et al., 2019).

Recently, AV-to-pedestrian communication via wearable devices has been proposed (Gelbal et al., 2023; Hasan & Hasan, 2022; Lakhdhir et al., 2023). More specifically, augmented reality (AR), defined as a technology that overlays virtual information onto the real world, appears to be a promising solution for communicating with pedestrians (Tabone et

al., 2021b). On the one hand, the use of AR might sound unusual and undesirable; after all, it may be questioned whether pedestrians would want to rely on an expensive headset in order to move safely through traffic (e.g., Berge et al., 2022; Tabone et al., 2021a). At the same time, given that other AR technologies such as car head-up displays (HUDs) (Transparency Market Research, 2023) and Google Maps visual overlays on mobile phones (Google, 2020), once seen as futuristic, are now common, it is conceivable that in the future, pedestrians will receive visual assistance through AR glasses.

Although AR is a potentially promising technology for pedestrians, still little research exists on this topic. One of the problems is that AR is currently challenging to implement, with only the most recent headsets capable of creating a compelling and user-friendly experience (e.g., Microsoft Hololens 2, Varjo XR 3, Magic Leap 2). While exceptions exist (e.g., Kang et al., 2023; Zhang et al., 2023), much previous AR research for pedestrians has limitations: some studies focus solely on the orienting phase without any form of human-subject evaluation (Tabone et al., 2021b; Tong et al., 2021), others use questionnaires with photos or videos (Hesenius et al., 2018; Tabone et al., 2023a; Wilbrink et al., 2023), and still others test AR concepts in immersive virtual reality (VR) environments (Malik et al., 2023; Pratticò et al., 2021; Tabone et al., 2023b; Tran et al., 2022). A critical note that should be made in this last point is that—although valid results can be obtained in VR—this is strictly speaking not AR. After all, merely adding virtual displays within a completely virtual environment is still considered VR, rather than AR. AR is different from VR in that AR incorporates elements of the physical world into the user's experience, creating a blend of virtual and real-world elements (Rauschnabel et al., 2022).

In this study, we share the findings from an outdoor experiment involving participants using an AR headset. In addition to the AR interfaces, the approaching AV was also simulated. The use of a virtual AV, as opposed to a real one, ensured consistent experimental conditions and simplified the process of attaching AR elements to the moving AV. Our approach serves as a stepping stone to a future full AR experience, in which pedestrians can walk around untethered while the AR headset wirelessly communicates with an actual AV. Our use of AR, where virtual road users coexist with real humans in an outdoor space, has been applied by several others before, though not in the context of AV-pedestrian communication (Kamalasanan et al., 2022; Maruhn et al., 2020). Our methodology also corresponds with other realistic simulation methods, such as displaying virtual road users in a real car (Bokc et al., 2007; Butenuth et al., 2017; Feng et al., 2018; Hussain et al., 2013; Sheridan, 2016), a technique that supersedes driving simulators which inherently offer limited visual and motion cues. In our approach, while the environment remains the real world, it is enhanced with simulated experimental stimuli.

In the current experiment, four different AR interfaces and a no-AR-interface baseline condition were compared among human participants. The focus was on whether these AR interfaces made participants feel safe to cross, as well as subjective qualities including whether the AR interface was found to be intuitive and was accepted. The interfaces used in this study were adopted from three prior works: an AR concept design study (Tabone et al., 2021b), an online questionnaire with video clips in which participants provided subjective ratings (Tabone et al., 2023a), and a CAVE-based simulator study in which participants crossed a virtual road (Tabone et al., 2023b). Our goal was to investigate whether the results we found in the current experiment correspond with the earlier research among human

participants (Tabone et al., 2023a, 2023b).

The tested AR interfaces differed in terms of their anchoring techniques: the AR interface was either vehicle-locked, world-locked, or head-locked, three methods that bring fundamentally different demands regarding the user's visual attention (Lebeck et al., 2017; Lingam et al., 2023; Peereboom et al., 2023; Tabone et al., 2023b). A vehicle-locked AR interface entails, just like an eHMI, that implicit communication in the form of the speed of and distance to the AV and explicit AR cues are congruent in time and place. A head-locked AR interface, on the other hand, is always visible but requires that the AV is looked at to confirm the cues of the AR interface. Finally, a world-locked interface requires attention to be distributed between the AV and the AR interface, with the possible advantage that the AR interface can be presented at a fixed and familiar location.

For illustration, Lingam et al. (2023) conducted a study using a VR driving simulator to investigate whether eHMIs were better positioned on the roof of an approaching AV, or integrated into the road infrastructure, resembling a traffic light configuration. Their results showed that both solutions were more favoured by the car drivers than the absence of any signalling. However, each concept presented its own set of advantages and disadvantages. The advantage of the infrastructure solution (i.e., world-locked signal) was that the signals were positioned at a fixed and accessible location before the intersection, and the driver could see this from a distance. An eHMI on the AV (i.e., vehicle-locked signal), on the other hand, required that the driver first turn their head/eyes to see the AV with its eHMI; once identified, however, the driver could directly infer what the AV was going to do, as deduced from its speed and the eHMI signal.

The current study, conducted with pedestrians outdoors instead of in VR, aimed to determine whether similar strengths and weaknesses could be identified for various AR anchoring methods. We implement four AR interface concepts, taken from a CAVE study by Tabone et al. (2023b): (1) *Virtual fence*, a world-locked interface that is highly visible, yet may impart a false sense of security and which partially occludes the environment due to its semi-transparent walls, (2) *Fixed pedestrian lights*, another world-locked interface; it has a familiar design and is smaller in stature compared to the *Virtual Fence*, (3) *Pedestrian lights HUD*, a head-locked interface similar to a conventional traffic light, but always visible in the users field of view, and (4) *Planes on vehicle*, a vehicle-locked AR interface similar to an eHMI. We adopted the designs from Tabone et al. (2023a, 2023b) without design improvements. Note that several insurmountable differences exist between the AR implementation of this study and that of the online and VR studies Tabone et al. (2023a, 2023b), such as in timing and visibility in the external environment, thereby making an exact numerical comparison not our objective. Instead, we opted for a qualitative comparison, focusing on the effects of the anchoring method. In addition to comparing different AR interfaces with no AR interface, one of the objectives of this research was to document, accompanied by Unity source code, how AR research in the outdoor environment can be conducted.

7.2 METHODS

7.2.1 PARTICIPANTS

Participants were recruited through personal networks, without the offer of financial reimbursement or other incentives. The study included a total of 28 persons, comprising 23 males and 5 females, with ages ranging from 19 to 59 years. The average age was 27.2 years, with a standard deviation of 9.2 years. In response to the intake questionnaire item 'Do you use any seeing aids?', 19 participants answered 'no', 5 answered 'yes, glasses', and 4 answered 'yes, contact lenses'. We did not record whether participants wore glasses or contact lenses *during* the experiment. However, it was observed that at least one participant wore their glasses under the headset, without noticeable problems.

The intake questionnaire further indicated that 19 of the participants were students. Twenty-six participants indicated being Dutch, one Chinese, and the nationality of one participant was unknown. Twenty-six participants held a driver's licence, for an average of 8.0 (SD = 7.3) years. A total of 17 participants had used a VR headset before while 9 participants had used an AR headset before. Daily walking time was less than 15 min for 7 participants, 15 to 30 minutes for 12 participants, and more than 30 min for 9 participants. Cycling was the primary mode of transportation for most participants (22 out of 28). A test for colour blindness showed that one participant was colorblind. This person was not excluded from participation, since the AR interfaces featured redundancy gain in that they did not solely rely on red/green colours, but also used icons, movement, or stimulus location cues. Additionally, the inclusion of this participant was considered beneficial for gaining valuable insights.

Each participant provided written informed consent. The experiment procedure was approved by the TU Delft Human Research Ethics Committee, approval no. 3054.

7.2.2 MATERIALS AND SETTINGS

The experiment took place in a designated area of the Delft University of Technology campus, which was closed to normal traffic. It was conducted on an Alienware PC powered by an Intel®Core™i7-9700K CPU at 3.60 GHz, equipped with 16 GB RAM and a Nvidia GeForce RTX 2080 Ti GPU. The AR software ran in Unity 2021.3.13f1 (Unity, 2022), combined with the Varjo XR Plugin (Varjo Developer, 2023). A custom script was used that allowed the experimenter to select the experiment conditions from within Unity.

AR was displayed by means of a Varjo XR-3 headset (Varjo, 2023). The Varjo XR-3 provides a 90 Hz refresh rate and a 115° horizontal field of view. The focus area, of 27° × 27° , was rendered at 70 pixels per degree on a μOLED display, providing 1920 × 1920 pixels per eye. The peripheral area was rendered at about 30 pixels per degree on an LCD, producing 2880 × 2720 pixels per eye. A pole was used to route the cables from the PC to the participant (Figure 7.1).

The Varjo XR-3 presented a virtual AV and virtual AR interfaces while depicting the real world by means of video pass-through. Within the headset configuration software 'Varjo Base', the imaging and exposure settings were set to automatic, and the highest image quality settings were selected. This proved to be feasible for our application without glitches or delays in data storage or visual animations.

Our initial intention was to anchor the virtual objects to the real world by means of object

tracking and reference markers (Varjo Technologies, 2023). However, this approach proved to be unreliable in our outdoor setting, presumably due to the relatively empty environment and sunlight reflections on the markers. Instead, tracking for the headset was adopted, by means of two SteamVR Base Stations 2.0 (HTC VIVE, 2019), as depicted in Figure 7.1. Our solution was to position the *AR Origin* at a fixed position, based on the initial calibration position and orientation of the Varjo XR-3. Each time the hardware was set up, calibration of the HMD tracking was performed first. In Unity, an invisible *Ground place* was created, which functioned as a surface for the AV to drive on. The scale value of the *AR Origin* was set such that distances in the virtual world corresponded with distances in the real world. Unity was configured to allow participants to rotate their heads and look around, but not to translate or move through the environment. The height of the *Main Camera* above the ground plane was fixed at 1.70 m. Our outside-in tracking approach provides a simple and robust way to present virtual overlays onto the real world. However, the location of these overlays is not anchored to the real world, and can somewhat vary depending on the precise calibration of the setup.

A Logitech R400 wireless presenter was used as the remote control. The right arrow button was used by the participant as the button to indicate if the participant felt it was safe to cross. The receiver was connected via a USB port of the Alienware desktop.

The sound of the AV was transmitted through headphones that were plugged into the Varjo XR-3. The headphones did not include noise-cancelling, so surrounding real-world environment noise could still be heard.

7.2.3 PARTICIPANT INFORMATION AND TASK

Participants were provided with information about the procedure and tasks through a leaflet. They were informed that they would wear an AR headset, which uses cameras to capture the environment and can project virtual objects onto this real-world setting.

Participants were informed that they would stand at the side of a road and that their task was to press and hold the button when they felt it was safe to cross and to release it when they did not feel safe to cross. It was emphasised that participants should not physically cross the road. This method for measuring crossing intentions was previously introduced by De Clercq et al. (2019) and was considered favourable as it allows participants to respond almost immediately to changing conditions. In other research, we had participants actually cross (Tabone et al., 2023b) or communicate their crossing intention or 'critical gap' through hand gestures (Rodríguez Palmeiro et al., 2018) or by taking a single step (Epke et al., 2021). A drawback of these methods is the variability among participants in executing these actions, making it relatively challenging to extract their intentions or perceptions.

The leaflet also mentioned that, at the start of each trial, participants would see a circle (hereafter named 'attention-attractor circle') at one of three locations, i.e., either in front of them, to their left, or to their right. Participants were instructed to gaze at the circle for a duration of one second to initiate the trial. It was mentioned that a virtual vehicle would approach from the right, potentially stopping for the participant and possibly communicating this intent through various communication interfaces. Finally, it was mentioned that, after each trial, participants would be prompted with a question displayed in the AR environment, to which they were to provide a verbal response, and that this procedure would be repeated for a total of four different AR interfaces, plus a no-interface baseline condition. Partici-



Figure 7.1: Participant wearing the tethered Varjo XR-3 headset.

pants were also informed that, following every block with a particular AR condition, the experimenter would ask open-ended questions about their experiences with the presented condition.

In previous research, the attention-attractor circles were used to impose visual distraction. This was done to examine whether different AR interfaces exhibit different robustness to distraction while the AR interface was active (Tabone et al., 2023b). Amongst others, Tabone et al. (2023b) found that the *Pedestrian lights HUD* supported crossing decisions even though the participants were visually distracted and had not glanced at the AV yet, because the HUD moved with the participant's field of view . In the current study, the circles played a different role. In our case, the circles disappeared before the start of the trial, while the AR interface appeared 3.1 s later, which allowed participants to direct attention to the approaching AV before the activation of the AR interface. The purpose of the circles, in our case, was to reduce monotony of the task and to increase visual demands.

7.2.4 AUTOMATED VEHICLE BEHAVIOUR

In each trial, a virtual AV approached from the right, displaying one of two behaviours: non-yielding or yielding. The vehicle model asset had dimensions of 4.95 m in length, 2.10 m in width, and 1.35 m in height (Final Form Studio, 2021). The AV reflected light from

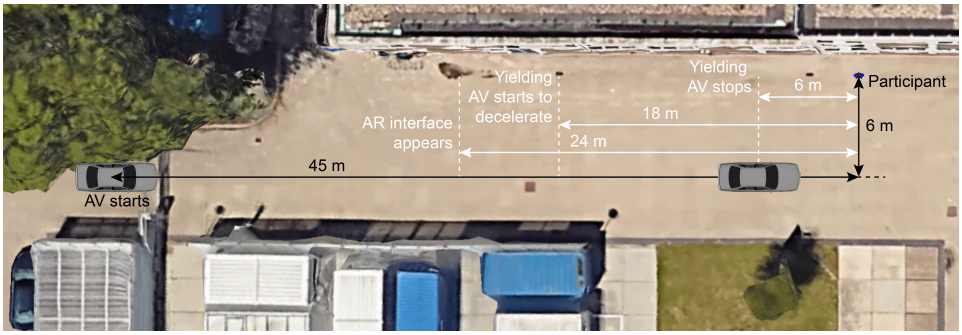


Figure 7.2: Top-down view of the experiment area (overlay drawn on an image from Google Earth, 2023).

a virtually positioned sun, and the wheels rotated in accordance with the forward speed. Shadows were not simulated. Audio was incorporated to produce a speed-sensitive engine sound, with a Doppler effect applied.

When the trial began, the AV appeared approximately 45 m away from the participant, with the distance measured parallel to the road. Simultaneously, an attention-attractor stimulus in the form of a cyan circle ring appeared either in front of, to the left of, or to the right of the participant. After the participant looked at the circle for 1 s, the AV began driving at a speed of 30 km/h.

Once 3.1 s had elapsed from the moment that the AV began driving, it passed an invisible trigger, prompting the AR interface to appear. In the yielding condition, the AV began decelerating 3.9 s from the start of its movement and came to a complete stop at an elapsed time of 7.8 s. The AV featured a slight forward pitch angle as it decelerated. In the non-yielding condition, the AV maintained a speed of 30 km/h and passed the participant 6.5 s from the beginning of its movement. Figure 7.2 provides an aerial view of the roadway, illustrating the sequence of events: the AV starting, the appearance of the AR interface, the point where the AV began decelerating, and where it came to a full stop.

7.2.5 AUGMENTED REALITY DESIGNS

The AR interfaces were adopted from earlier research in which these interfaces had been designed (Tabone et al., 2021b), and after minor modifications, presented to a large sample of online respondents (Tabone et al., 2023a), and subjected to experimental testing in an immersive virtual simulator (Tabone et al., 2023b). The nine AR interfaces were previously divided into three anchoring methods: world-locked, head-locked, and vehicle-locked (Tabone et al., 2023b). The present study focused on comparing four selected AR interfaces. Specifically, from each anchoring category, one was selected: Virtual fence (world-locked), *Pedestrian lights HUD* (head-locked), and *Planes on vehicle* (vehicle-locked). In addition, the *Fixed pedestrian lights* interface was selected. The reason for including this world-locked interface was to allow a comparison with the Pedestrian lights HUD. The four selected AR interfaces were as follows:

1. The *Virtual fence* consisted of a zebra crossing projected on the road combined with semi-translucent walls surrounding the zebra (Figure 7.3). The interface was 3.0 m

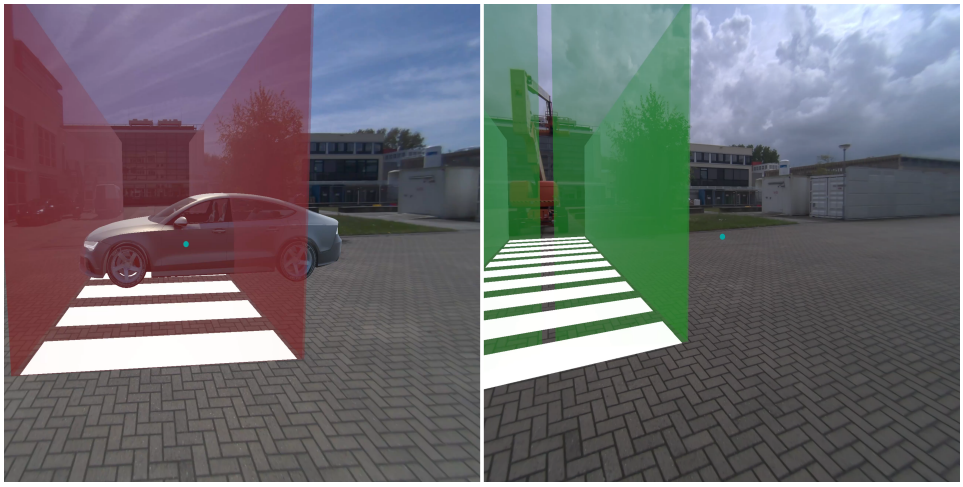


Figure 7.3: *Virtual fence* in the non-yielding (left) and yielding condition (right).

tall, 2.5 m wide and 7.5 m long. A semi-translucent gate positioned in front of the pedestrian opened in 1-s time after the interface appeared in the yielding condition, and remained closed in the non-yielding condition.

2. The *Fixed pedestrian lights* resembled existing traffic lights (Figure 7.4). It remained stationary across the street from where the pedestrian was positioned. The interface was 2.1 m tall and positioned 14 m from the pedestrian. It consisted of a pole with a box (0.20 m wide, 0.38 m tall) on top displaying either a lit-up red icon of a standing pedestrian, or a green icon of a walking pedestrian.

3. The *Pedestrian lights HU*D was similar to the *Fixed pedestrian lights* interface (Figure 7.5). However, instead of being world-locked, the same box was anchored to the user's head. It was placed at 1 m distance from the participant's head and off-centre by 30 cm upwards and to the right. The interface was rotated around the vertical axis in order to face the user.

4. The *Planes on vehicle* was a vehicle-mapped interface consisting of a red plane with a stop-hand icon or a green plane with an icon depicting a pedestrian crossing a zebra crosswalk (Figure 7.6). The 2.1 wide and 1.58 m tall plane hovered above the front of the AV so that the bottom of the plane aligned with the front bumper; it was tilted by 54° to be positioned parallel with the AV's windshield.

In all four AR interfaces, colour was used to provide a redundant cue; for the non-yielding AVs this was pure red, and for yielding AVs this was pure green. The *Virtual fence* and *Planes on vehicle* were semi-transparent, to ensure that the AV or possibly other relevant objects remained visible to the participants.



Figure 7.4: *Fixed pedestrian lights* in the non-yielding (left) and yielding condition (right).



Figure 7.5: *Pedestrian lights HUD* in the non-yielding (left) and yielding condition (right).



Figure 7.6: *Planes on vehicle* in the non-yielding (left) and yielding condition (right).

7.2.6 EXPERIMENT DESIGN

The experiment was of a within-subject design. Each participant was exposed to the four AR interfaces and a no-interface baseline condition, two AV behaviours (yielding and non-yielding), and three attention-attractor locations (left, middle, right).

Participants first completed two practice trials in the Baseline condition, one trial with a yielding AV one trial with a non-yielding AV. Each of the five interface conditions was presented in a separate block. Each block consisted of six trials: three with a yielding AV and three with a non-yielding AV. The three trials were conducted with the attention attractor at the left, right, or middle. This resulted in a total of 30 trials per participant (5 interface conditions × 2 yielding behaviours × 3 attention-attractor locations). The order of the five blocks, as well as the order of the six trials within each block, were counterbalanced using a Latin Square method.

The experiment lasted 45 to 60 minutes per participant, with the time spent wearing the headset amounting to approximately 30 minutes.

7.2.7 QUESTIONNAIRES AND RATING SCALES

After signing the consent form, participants completed an intake questionnaire, designed in Qualtrics (Qualtrics XM, 2023), to gather demographic information. The intake questionnaire also included items about affinity for technology (ATI scale, Franke et al., 2019), items about whether participants had experienced VR or AR headsets before, and a brief test for colour blindness (Ishihara, 1917). Some participants had completed the pre-experiment questionnaire before their scheduled experiment slot.

Next, participants were asked to read a leaflet with a short description of the experiment. A complementary oral explanation of the experiment was provided where needed. Subsequently, participants put on the Varjo XR-3 headset, were handed the remote button, and a multi-point eye-tracker calibration was conducted. After each trial, at exactly 10 seconds after the AV had started moving, a statement appeared in front of the participant: "*This*

interface/situation was intuitive for signalling: 'Please do cross the road'." for the yielding condition, or *"This interface/situation was intuitive for signalling: 'Please do not cross the road'."* for the non-yielding condition. Participants verbally indicated to what extent they agreed on a scale from *Fully disagree (1)* to *Fully agree (7)*.

After each block of six trials with a particular AR condition, a semi-structured interview was conducted regarding interface design qualities, the timing of the interface appearance, the preference between yielding and non-yielding state, and the participant's wellbeing according to the misery scale (MISC; Bos, 2015).

After all five blocks, participants completed a post-experiment questionnaire in Qualtrics. This questionnaire contained items related to the AR experience and about AR interfaces in general. Participants were also asked to rank the five interface conditions in terms of their preference. Additionally, they were asked, for the four AR interfaces, to answer items regarding the intuitiveness of the green and red interfaces, convincingness of the green and red interface, trust in the interface to make a decision (fully disagree, fully agree), size (too small, too large), timing (too early, too late), clarity/understandability, and visual attractiveness, as well as a 9-item acceptance scale (Van der Laan et al., 1997), identical to Tabone et al. (2023a). The items used 7-point scales, except for the acceptance scale and adoption questions which use a 5-point semantic differential scale, and the ranking item. Open questions were also asked per interface condition to allow participants to justify their responses.

7.2.8 DATA RECORDING AND ANALYSIS

The data was stored at a frequency of 50 Hz. Firstly, we determined per trial what percentage of the time participants kept the response button pressed. This was done from the moment the AR interface appeared until the vehicle came to a stop (yielding AVs) or passed (non-yielding AVs).

Additionally, from the post-experiment questionnaire, we determined a composite score per AR interface, identical to how it was done by Tabone et al. (2023a). We calculated a composite score because the self-reports were strongly correlated, and we found no evidence of multiple underlying constructs (Tabone et al., 2023a). This composite score was calculated based on participants' responses to 15 items, which included a 9-item acceptance scale and 6 additional items (1. Intuitiveness for nonyielding AVs, 2. Intuitiveness for yielding AVs, 3. Convincingness for non-yielding AVs, 4. Convincingness for yielding AVs, 5. Clarity/understandability, 6. Attractiveness). To calculate the composite score, we first concatenated the questionnaire results from the 28 participants across 4 AR conditions, resulting in a grand 112×15 matrix. We then standardised each of the 15 variables, so their mean became 0 and their standard deviation became 1. Next, we summed the standardised scores of the 15 items, resulting in a 112-element vector of total scores. These 112 scores were standardised again, providing a composite score for each participant and each AR interface. The preference rank, in which participants had to sort the five AR conditions from 1 (most preferred) to 5 (least preferred), was analysed as a separate item.

Finally, the post-trial intuitiveness ratings were averaged over 3 trials per participant so that for each AR interface and participant, and for both yielding and non-yielding AV, an intuitiveness score was available. Since the location of the attention-attractor circles did not appear to have a major influence on how the four AR interfaces were responded

to, the results for the three trials per AR interface condition and yielding condition were averaged. In the statistical analyses, repeated-measures ANOVAs were used, with the AR condition as the independent variable. The findings are shown as means, complemented by 95% confidence intervals for within-subjects designs (Morey, 2008). For each of the nine measures, post hoc comparisons between conditions were conducted using paired-samples *t*-tests. An alpha value of 0.005 was chosen, which is more conservative than the usual 0.05, due to a maximum of 10 possible combinations of AR conditions that can be compared (4 + 3 + 2 + 1).

The interview results for each AR interface and participant were condensed into brief highlights by one of the authors (the experimenter). These highlights from all 28 participants were subsequently analysed and summarised per AR condition by another author using thematic analysis, with a focus on the clarity of the AR interface and participant satisfaction. The resulting summaries were checked with the interview transcripts (available for 16 out of the 28 participants) to ensure representativeness and were adjusted as needed. For each AR interface, one or two quotes (translated from Dutch) were selected for presentation alongside the summary.

7.3 RESULTS

THE 28 participants performed a total of 840 trials. 4 of 840 log files were unavailable or could not be used due to an experimenter error. For the post-experiment questionnaire, responses for the Planes on vehicle were unavailable for 1 of 28 participants.

Misery scores were low, with 26 of 28 participants reporting no symptoms (0 or 1 on the 11-point scale) and 2 of 28 participants reporting maximum scores of 3 and 4, respectively, at any time during the experiment. One of these two participants indicated feeling ill already before the experiment.

Figure 7.7 shows the percentage of trials in which participants held down the response button. Interestingly, for both non-yielding AVs (Figure 7.7, left),and yielding AVs(Figure 7.7, right), a large portion of the participants (approximately 60%) did not hold down the response button at the beginning of the trial.The reasons for this are not clear, but it may be related to participants experiencing high workload or the participants not being reminded to keep the button pressed at the start of the trial. A technological explanation, involving the wireless connection of the button device, cannot be ruled out either. Therefore, the results in Figure 7.7 should be considered in a relative context, comparing the five conditions with each other, rather than in absolute terms.

For the non-yielding AV (Figure 7.7, left), participants, on average, released the response button as the AV came closer, indicating that they felt less and less safe to cross. However, about 13% of the participants did not have the response button released while the vehicle passed.

For the yielding AVs (Figure 7.7, right), participants started pressing the response button after the AR interface appeared, at 3.1 s. For the baseline condition, this happened somewhat later, which can be explained by the fact that the AV only started to slow down at 3.9 s (thus, 0.8 s after the AR interfaces became visible). That is, in the baseline condition, participants could not anticipate that the AV was going to stop until it actually started to decelerate. From the button press percentages in the grey interval, it can be seen that the four AR interfaces were slightly more effective than the baseline condition.

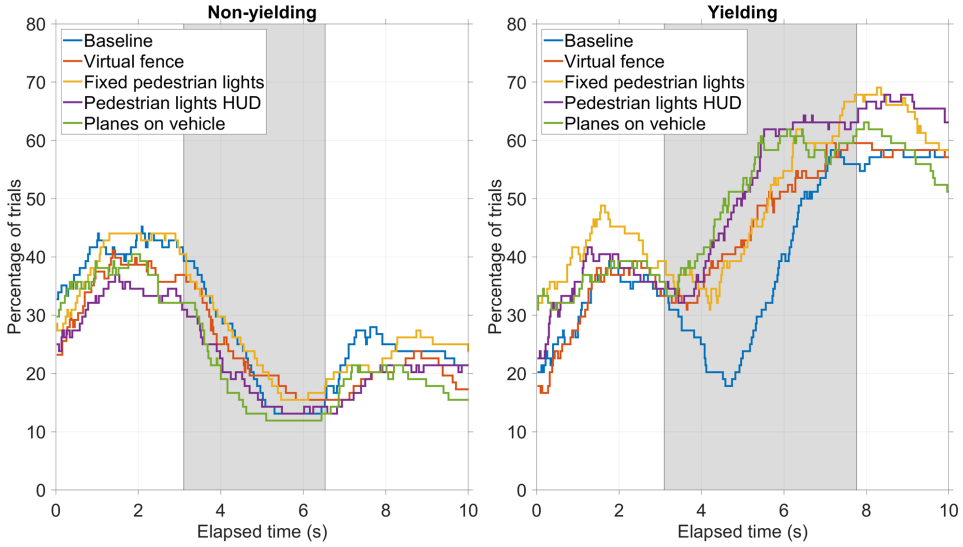


Figure 7.7: Button press percentage for the four AR interfaces and baseline condition, for non-yielding AVs (left), and yielding AVs (right). The grey background represents the interval between the moment the AR interface appeared (at 3.1 s) and the AV passed (non-yielding AVs, 6.5 s) or the AV came to a full stop (yielding AV, 7.8 s).

Figure 7.8 shows the button press percentagesin the selected time interval for non-yielding AVs (A) and yielding AVs (B), as well as the mean composite scores calculated from 15 items of the post-experiment questionnaire (C), intuitiveness scores measured after each trial for non-yielding AVs (D) and yielding AVs (E), the preference rank (F), trust (G), timing (H), and size (I). The green numbers indicate pairs of conditions that exhibit statistically significant differences following a paired-samples *t*-test with Bonferroni correction. Full results from the paired-samples *t*-tests are accessible in the Data Repository.

The results presented in Figure 7.8 indicate that the baseline condition scored relatively poorly, as demonstrated by a low button press percentagefor yielding AVs (B), low intuitiveness scores especially for non-yielding AVs (D) and yielding AVs (E), and on average the worst preference rank (F).

Among the four AR interfaces, there were no major differences, although the *Fixed pedestrian lights* scored relatively poorly. This is evident from the low composite score (c), low intuitiveness scores (D and E), and the poor preference rank (F) among the four AR interfaces. However, none of these effects were statistically significant. The largest effects between the four interface conditions were, however, observed regarding the perception of the physical variables timing and especially size. More specifically the *Fixed pedestrian lights* were perceived as clearly too small (I), and related to this, participants also believed that the traffic lights became visible too late (H); its timing was identical to that of the other three interfaces, but this assessment might be due to the fact that it took participants extra time to visually locate it, creating the impression that it became visible too late.

The results of the repeated-measures ANOVAs corresponding to the data presented in Figure 7.8 are shown in Table 7.1. The effect sizes (partial $\eta 2$) were not particularly strong

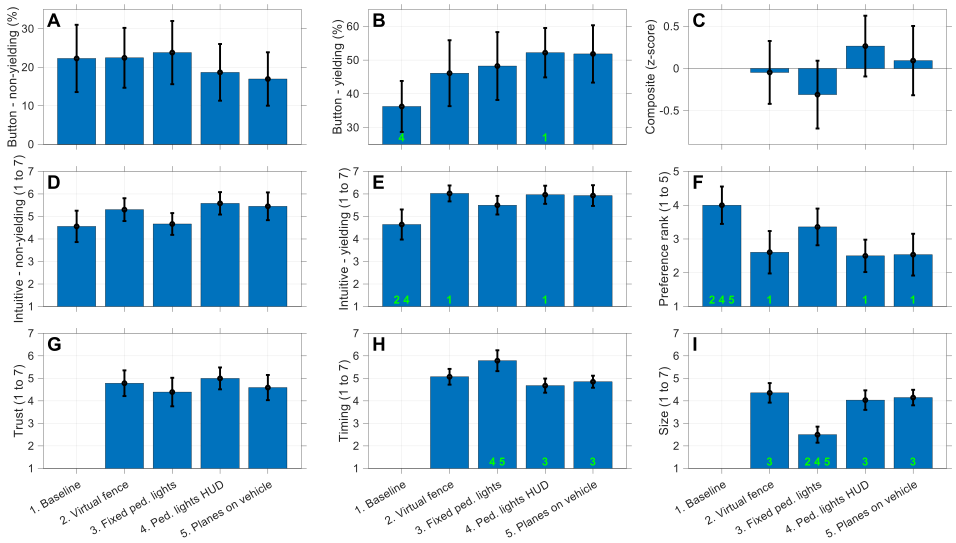


Figure 7.8: Means and 95% confidence intervals of participants' **(A)** button press percentage for non-yielding AVs (%), **(B)** button press percentage for yielding AVs (%), **(C)** composite scores based on post-experiment self-reports (z-score), **(D)** post-trial intuitiveness scores for non-yielding AVs (1: fully disagree, 7: fully agree), **(E)** post-trial intuitiveness scores for yielding AVs (1: fully disagree, 7: fully agree), **(F)** preference ranking (1: most preferred, 5: least preferred), **(G)** trust for decision-making (1: fully disagree, 7: fully agree), **(H)** interface trigger timing (1: too early, 7: too late), and **(I)** interface size (1: too small, 7: too large). The numbers in green indicate the conditions with which this condition shows a statistically significant difference after Bonferroni correction ($p<$ 0.005), according to paired-samples t-tests.

for the button presses (A and B), the composite score (C), or the trust score (G). However, they were fairly strong and statistically significant for the post-trial intuitiveness ratings (D and E), the preference rank (F), and as mentioned above, the timing (H) and size (I).

The results of the post-block interviews are indicated below:

- **Baseline.** Without an interface, many participants felt the situation was more akin to real-life scenarios. They indicated that the absence of clear signals, or feelings of insecurity, without the interface caused them to take longer in making decisions. The factors influencing their decision-making were the AVs deceleration and speed. Additionally, the changing pitch of the AV's sound during deceleration was reported to be a relevant cue.

  "Well, at some point you see that car slowing down. Then you still remain a bit apprehensive, thinking, 'Okay, it's really stopping.' And only then do I decide to start walking."

- **Virtual fence.** Participants mentioned that the size of the *Virtual Fence* made it stand out, and some even found it somewhat intimidating. The interface was generally perceived as clear, particularly the green signal. However, the red signal was confusing for some participants, who were unsure whether it warned them not to cross or

Table 7.1: Results of repeated-measures ANOVAs for the AR interface comparisons shown in Figure 7.8.

Dependent measure	Measurement moment	df	F	p	partial η^2
(A) Button press percentage, non-yielding AVs	During trials	4,108	0.57	0.684	0.02
(B) Button press percentage, yielding AVs	During trials	4,108	2.33	0.060	0.08
(C) Composite score	Post-experiment	3,78	1.65	0.185	0.06
(D) Intuitiveness, non-yielding AVs	Post-trial	4,108	2.88	0.026	0.10
(E) Intuitiveness, yielding AVs	Post trial	4,108	6.34	< 0.001	0.19
(F) Preference rank	Post-experiment	4,108	5.72	< 0.001	0.17
(G) Trust	Post-experiment	3,78	0.91	0.441	0.03
(H) Timing	Post-experiment	3,78	6.90	< 0.001	0.21
(I) Size	Post-experiment	3,78	19.1	< 0.001	0.42

Note: For (C), (G), (H), and (I), data for *Planes on vehicle* were missing for one participant.

indicated that the AV would stop. The zebra crossing within the interface also posed a dilemma: it seemed to encourage walking, but this was in conflict with the red walls. Additionally, some participants expressed concerns that the interface might obstruct the view of other potential road users.

"Yes, so I found it a bit difficult at first, because you initially have those walls, which are red. So then you think, oh, is it red for them? And are you protected by those walls or something? And you're also in a zebra crossing. When I see a zebra crossing, I think, oh, I need to walk. So I found that a bit counterintuitive. But once you get used to it, I think you see it faster."

- **Fixed pedestrian lights.** The *Fixed pedestrian lights* were a familiar concept to participants. However, its appearance, timing, and placement posed challenges. Some participants reported they failed to notice it, especially during their first trial. Furthermore, its sudden appearance made it hard for participants to rely on. Additionally, a recurrent concern among participants was the need to switch gaze between the AV and the traffic light. Finally, participants suggested making the distinction between red and green clearer.

"More unclear, because the pole only appears relatively late. So you're really only focused on the car that's approaching. And at first, I either saw the pole or I didn't.

You see the car approaching sooner, so then you start considering whether you're going to cross or not. And only then do you see the pole."

- **Pedestrian lights HUD.** The *Pedestrian lights HUD* also provided a familiar interface. Its upper-right positioning in their field of view was reported to be both an advantage and a disadvantage. Specifically, some participants reported that the HUD required them to roll their eyes, which felt unnatural to them, while others appreciated the constant visibility and fixed position within their view. Its sudden appearance sometimes led to initial startles or distractions, and some participants commented that it was obstructive and blocked a portion of their view. Finally, as with the *Fixed pedestrian lights*, according to some participants, the distinction between the red and green signals (lit-up vs. non-lit-up state) could have been clearer.

"Very clear, indeed. I'm not used to it, so that's why the first one might have been a bit unclear. Because I was thinking, should I now look or should I clearly know what the car is going to do? But after that, it's very clear, because it's just close by and you actually see it right away."

- **Planes on vehicle.** Participants initially found the *Planes on vehicle* novel and stated it took some time to become accustomed to it. The green signal and icon were perceived as clear. However, there was some uncertainty regarding the red signal, with participants unsure whether the AV would stop. Many felt that the *Planes on vehicle* improved trust because the communication came directly from the AV. One colour blind person, however, did not immediately understand the meaning of this interface. Furthermore, some participants remarked that in hypothetical situations involving multiple vehicles, the effectiveness of the interface might decrease. It was suggested that, for optimal effectiveness in busier scenarios, all cars might need to adopt such an interface.

"I don't know, it made me a bit nervous, because I thought okay... It seems very illogical that there's a very large sign in front of the car."

"Better than the previous one. I liked that you only have to pay attention to one thing. So you just have to look at the car and it's clear whether you can cross or not. Instead of having to look somewhere else for a signal. It was just clearly on the car."

Finally, a recurring topic regarding the timing of the AR interfaces was that they were activated somewhat late (even though this was still 0.8 s before the vehicle began to brake). Participants suggested that the AR interfaces could appear earlier, possibly in a default state, so that they would know where to look in advance to make their crossing decision. Another theme that emerged from the interviews is that participants tended to verify the advice of the AR interfaces with the movement of the vehicle.

7.4 DISCUSSION

THIS study tested four AR interfaces for supporting the interaction with AVs among 28 human participants, complemented with a baseline condition without AR interface. For the research, we used actual AR in an outdoor environment, in contrast to previous research that used AR-in-VR. In our paradigm, participants stood outside and had to press a button as long as they felt safe to cross. The experiment was set up so that the AV was virtual; this way, we could offer the same vehicle movement and AR activation timings in every trial.

Some of the results did not match our previous studies. Specifically, a previous experiment in a CAVE environment (Tabone et al., 2023b) and a large-scale online survey (Tabone et al., 2023a) showed that among the AR interfaces tested, the world-locked *Fixed pedestrian lights* and head-locked *Pedestrian lights HUD* yielded relatively high intuitiveness ratings while the vehicle-locked Planes on vehicle received lower ratings, though still net positive. However, in the current study, the intuitiveness of the *Planes on vehicles* interface was rated highly, with average scores of 5.5 (SD = 1.6) for non-yielding AVs and 5.9 (SD = 1.2) for yielding AVs on a scale of 1 to 7, whereas the *Fixed pedestrian lights* interface received lower ratings, at 4.7 (SD = 1.7) for non-yielding AVs and 5.5 (SD = 1.2) for yielding AVs, respectively (see Figures 7.8D & 7.8E).

Although the relative ratings of the AR interfaces did not immediately correspond with prior research, the information-processing mechanisms did. For example, previous online research using animated video clips (Tabone et al., 2023a) and research using a virtual pedestrian simulator (Tabone & De Winter, 2023) also found that a zebra crossing combined with the colour red can cause confusion, and that walls of the *Virtual fence* may obstruct the view of other road users. Additionally, the fact that there can be an egocentric vs. exocentric perspective confusion when a car emits a red signal is also known (Bazilinskyy et al., 2020), and the problem that world-locked interfaces, like the *Fixed pedestrian lights* in our case, cause divided attention, has also already been documented (e.g., Peereboom et al., 2023). The explanation for the relatively poor performance of the *Fixed pedestrian lights* in the present study seems to lie in more practical factors: We had placed it relatively far away from the participant, at 14 m. Furthermore, for most of the participants, an orange-coloured aerial work platform was present, located behind the *Fixed pedestrian lights*. This visual clutter could have further increased the difficulty in identifying the traffic light and discerning its status. This research thus shows that basic design decisions related to salience can have large effects on the effectiveness of AR interfaces.

The *Pedestrian lights HUD* also suffered from 'practical limitations'; it was positioned slightly off-centre, leading to some discomfort as individuals had to adjust their gaze (see Plabst et al., 2022, for a similar phenomenon). Note that participants could only glance at the HUD if they rotated their eyes, as the HUD always followed the user's head movement. The icons of the traffic light were not particularly salient in their illuminated 'on' state compared to their non-illuminated 'off' state, causing some participants to glance at the traffic light HUD instead of relying purely on their peripheral vision. Positioning the HUD towards a more central position would likely increase the risk of occlusion of relevant objects, which are typically centrally located in users' fields of view. Participants experienced fairly low sickness scores, which differs from the higher scores observed in Peereboom et al.'s (2023) HUD study for pedestrians in VR. This difference may be attributed to our HUD being placed at a larger virtual distance (1.0 m, compared to their 0.36 m in Peereboom et al.),

reducing the accommodation-vergence conflict. However, it isanticipated that poorer task performance and/or nausea can occur in more dynamic situations, where the AR information is not locked to, or embedded in, the world. This concerns the use of HUDs while the user is walking through the environment, resulting in a moving background relative to static AR stimuli in the field of vision (Fukushima et al., 2020) or when standing still while the AR stimuli move in the field of view (Kaufeld et al., 2022). For future research, we suggest creating a HUD that does not require the user's direct focus. For instance, the onset of a more bright coloured light, which can be unambiguously identified through peripheral vision, might offer a more comfortable experience for the user (e.g., Chaturvedi et al., 2019).

The *Fixed pedestrian lights* had the disadvantage that it required divided attention. This same disadvantage was mentioned in earlier experimental research in a CAVE environment, where participants took a relatively long time to visually identify the traffic lights (Tabone et al., 2023b). It might be the case that in the current real-world study, the task of identifying the traffic lights and the AV was even more challenging than in VR, because there is more visual clutter in the real world, such as other pedestrians and parked vehicles.

Indeed, we noticed that participants seemed to have difficulty with the experiment. For example, some participants had difficulty localising the attention-attractor circles or had to be reminded to look into these circles; participants also occasionally had to be reminded to press the response button if they felt safe to cross. One explanation for this forgetting is that we did not explicitly provide 'press now' instructions before each trial (e.g., Peereboom et al., 2023). Another explanation for the low press percentages is that the button, which operated through infrared light, might not have worked reliably in the outdoor environment. It was observed that when participants held the button in front of their body, instead of next to their body, missing values could result[1]. The button press percentages, as shown in Figure 7.7, resemble online crowdsourcing research on pedestrian-crossing decisions (e.g., Bazilinskyy et al., 2020, 2021; Sripada et al., 2021), where a portion of crowdworkers were apparently inattentive. Another possible explanation for the poor quality of button press data is that mental demands are higher with AR in the real world compared to AR-in-VR, because the real world is more cluttered, and the participant has more to keep track of, including perceiving real objects as well as maintaining postural stability and safety in the physical world. In turn, these observations suggest that simplicity and clarity of AR communication are likely even more important than in VR or online experiments.

A limitation of our study is that it used simulated AVs; in future traffic, actual AVs would need to communicate their stopping intentions wirelessly to the pedestrian's AR headset. Furthermore, instead of outside-in tracking using base stations that emit infrared light, there would likely need to be a form of inside-out tracking, where the pedestrian's headset detects objects in the environment (e.g., Bhakar et al., 2022). Although the Varjo XR-3 is a state-of-the-art AR device, there were some technical hiccups, such as a small number of occasions of loss of tracking, which caused VR objects to display an abrupt rotation, leading to some confusion or disorientation. Moreover, while the headset offered a large FOV, it was still more limited than natural vision without a headset. Finally, in the current study, only a single AV came from the right. Using multiple vehicles would increase realism, something that is especially relevant for vehicle-locked AR interfaces. On the

[1]Excluding trials where no button press was recorded is not advisable. This could introduce bias, as it is unclear whether participants genuinely did not feel safe crossing or if there was a technical issue.

other hand, the environment was realistic, with university employees and students walking around occasionally, as well as some maintenance vehicles being present. Adding false positives, like a green-coloured signal in combination with an AV that does not stop, could be useful to investigate (over-)reliance on the AR interfaces (see also Holländer et al., 2019; Kaleefathullah et al., 2022).

7.5 CONCLUSION

IN this study, different AR interfaces for AV-pedestrian interactions were assessed in an outdoor setting, distinguishing it from previous online and AR-in-VR research. The results showed that having an AR interface was generally preferred to no AR interface. Additionally, the results of post-trial interviews replicated previous information-processing-related findings, such as that a red-coloured surface can be confusing, since this cue can pertain to the AV or to the pedestrian. The experiment also showed that visual attention mechanisms are of key importance, with a world-locked traffic light causing challenges since pedestrians need to distribute attention between the approaching AV and the traffic light. Our findings also demonstrated the importance of practical design considerations, such as placement and salience, in determining the effectiveness of AR interfaces. Our results and observations further suggest that participants found the present task challenging, which, given the complex and cluttered nature of the real world as opposed to virtual environments, points to the need for simplicity and clarity in AR-based communication. Future research is thus advised to implement simple AR solutions, such as through a single coloured light visible in the periphery. Finally, it is recommended to explore forms of AR for pedestrians where the pedestrian can move around wirelessly. Such research could be a step towards AR for vulnerable road users, with real application.

SUPPLEMENTARY MATERIAL

Raw data, MATLAB scripts used for the analyses, and material from the experiment are available at the public data repository: https://doi.org/10.4121/a1f9f15c-1213-4657-8e4d-a154a725d747

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 860410 (project name: SHAPE-IT).

REFERENCES

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2020). External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'? Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Toronto, Canada, 3721–3728. https://doi.org/10.1109/SMC42975.2020.9282998

Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2021). How should external Human-Machine Interfaces behave? Examining the effects of colour, position, message, activation distance, vehicle yielding, and visual distraction among 1,434 participants. Applied Ergonomics, 95, 103450. https://doi.org/10.1016/j.apergo.2021.103450

Berge, S. H., Hagenzieker, M., Farah, H., & De Winter, J. C. F. (2022). Do cyclists need HMIs in future automated traffic? An interview study. Transportation Research Part F: Traffic Psychology and Behaviour, 84, 33–52. https://doi.org/10.1016/j.trf.2021.11.013

Bhakar, S., Bhatt, D. P., Dhaka, V. S., & Sarma, Y. K. (2022). A review on classifications of tracking systems in augmented reality. Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012077). IOP Publishing. http://doi.org/10.1088/1742-6596/2161/1/012077

Bindschädel, J., Krems, I., & Kiesel, A. (2022). Two-step communication for the interaction between automated vehicles and pedestrians. Transportation Research Part F: Traffic Psychology and Behaviour, 90, 136–150. https://doi.org/10.1016/j.trf.2022.08.016

Bokc, T., Maurer, M., & Farber, G. (2007). Validation of the vehicle in the loop (vil); a milestone for the simulation of driver assistance systems. Proceedings of the 2007 IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, 612–617. https://doi.org/10.1109/IVS.2007.4290183

Bos, J. E. (2015). Less sickness with more motion and/or mental distraction. Journal of Vestibular Research: Equilibrium and Orientation, 25, 23–33. https://doi.org/10.3233/ves-150541

Butenuth, M., Kallweit, R., & Prescher, P. (2017). Vehicle-in-the-loop real-world vehicle tests combined with virtual scenarios. ATZ Worldwide, 119, 52–55. https://doi.org/10.1007/s38311-017-0082-4

Chaturvedi, I., Bijarbooneh, F. H., Braud, T., & Hui, P. (2019). Peripheral vision: a new killer app for smart glasses. Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Ray, CA, 625–636. https://doi.org/10.1145/3301275.3302263

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, Honolulu, HI. https://doi.org/10.1145/3334480.3382865

De Clercq, K., Dietrich, A., Núñez Velasco, P., De Winter, J. C. F., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. Human Factors, 61, 1353–1370. https://doi.org/10.1177/0018720819836343

Dey, D., Ackermans, S., Martens, M., Pfleging, B., & Terken, J. (2022). Interactions of automated vehicles with road users. In A. Riener, M. Jeon, & I. Alvarez (Eds.), User experience design in the era of automated driving (pp. 533–581). Cham: Springer. https://doi.org/10.1007/978-3-030-77726-5_20

Dey, D., Holländer, K., Berger, M., Eggen, B., Martens, M., Pfleging, B., & Terken, J. (2020). Distance-dependent eHMIs for the interaction between automated vehicles and pedestrians. Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Virtual Event DC, 192–204. https://doi.org/10.1145/3409120.3410642

Epke, M. R., Kooijman, L., & De Winter, J. C. F. (2021). I see your gesture: a vr-based study of bidirectional communication between pedestrians and automated vehicles. Journal of Advanced Transportation, 2021, 5573560. https://doi.org/10.1155/2021/5573560

Feng, Y., Yu, C., Xu, S., Liu, H. X., & Peng, H. (2018). An augmented reality environment for connected and automated vehicle testing and evaluation. Proceedings of the 2018 IEEE Intelligent Vehicles Symposium, Changshu, China, 1549–1554. https://doi.org/10.1109/IVS.2018.8500545

Final Form Studio. (2021). Sedan car - 01 [3D land vehicle model]. Unity Asset Store. Retrieved from https://assetstore.unity.com/packages/3d/vehicles/land/sedan-car-01-190629

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. International Journal of Human–Computer Interaction, 35, 456–467. https://doi.org/10.1080/10447318.2018.1456150

Fukushima, S., Hamada, T., & Hautasaari, A. (2020). Comparing world and screen coordinate systems in optical see-through head-mounted displays for text readability while walking. Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality, 649–658, Porto de Galinhas, Brazil. https://doi.org/10.1109/ISMAR50242.2020.00093

Gelbal, S. Y., Cantas, M. R., Guvenc, B. A., Guvenc, L., Surnilla, G., & Zhang, H. (2023). Mobile safety application for pedestrians. arXiv. https://doi.org/10.48550/arXiv.2305.17575

Google. (2020). A new sense of direction with Live View. https://blog.google/products/maps/

new-sense-direction-live-view

Google Earth. (2023). Top-down view of part of the TU Delft Campus. https://earth.google.com/web/@52.00008632,4.37098049,0.35726392a,80.22226634d,35y,336.33134285h,0t,0r/data=OgMKATA

Hasan, R., & Hasan, R. (2022). Pedestrian safety using the Internet of Things and sensors: Issues, challenges, and open problems. Future Generation Computer Systems, 134, 187–203. https://doi.org/10.1016/j.future.2022.03.

Hesenius, M., Börsting, I., Meyer, O., & Gruhn, V. (2018). Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Barcelona, Spain, 261–268. https://doi.org/10.1145/3236112.3236148

Holländer, K., Wintersberger, P., & Butz, A. (2019). Overtrust in external cues of automated vehicles: An experimental investigation. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 211–221. https://doi.org/10.1145/3342197.3344528

HTC VIVE. (2019). SteamVR Base Station 2.0. https://www.vive.com/eu/accessory/base-station2

Hussain, K. F., Radwan, E., & Moussa, G. S. (2013). Augmented reality experiment: drivers' behavior at an unsignalized intersection. IEEE Transactions on Intelligent Transportation Systems, 14, 608–617. https://doi.org/10.1109/TITS.2012.2226239

Ishihara, S. (1917). Tests for colour-blindness. Tokyo: Kanehara Shuppan Co., Ltd.

Jungmann, A., Lang, C., Pinsker, F., Kallweit, R., Taubenreuther, M., & Butenuth, M. (2020). Artificial intelligence for automated driving – quo vadis? In T. Bertram (Ed.), Automatisiertes Fahren 2019: Von der Fahrerassistenz zum autonomen Fahren 5. Internationale ATZ-Fachtagung (pp. 117–134). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-27990-5_11

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & De Winter, J. C. F. (2022). External Human-Machine Interfaces can be misleading: An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human Factors, 64, 1070–1085. https://doi.org/10.1177/0018720820970751

Kamalasanan, V., Mukbil, A., Sester, M., & Müller, J. P. (2022). Mixed reality agent-based framework for pedestrian-cyclist interaction. Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct, Singapore, 363–368. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00079

Kang, Y., Choi, S., An, E., Hwang, S., & Kim, S. (2023). Designing virtual agent human–machine interfaces depending on the communication and anthropomorphism levels in augmented reality. Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ingolstadt, Germany, 191–201. https://doi.org/10.1145/3580585.3606460

Kaufeld, M., Mundt, M., Forst, S., & Hecht, H. (2022). Optical see-through augmented reality can induce severe motion sickness. *Displays*, 74, 102283. https://doi.org/10.1016/j.displa.2022.102283

Lakhdhir, S., Somanath, S., & Sharlin, E. (2023). Wearing awareness: Designing pedestrian-wearables for interactions with autonomous vehicles. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany. https://doi.org/10.1145/3544549.3585655

Lebeck, K., Ruth, K., Kohno, T., & Roesner, F. (2017). Securing augmented reality output. Proceedings of the 2017 IEEE Symposium on Security and Privacy, San Jose, CA, 320–337. https://doi.org/10.1109/SP.2017.13

Lingam, S. N., De Winter, J. C.F. , Dong, Y., Tsapi, A., Van Arem, B., & Farah, H. (2023). eHMI on the vehicle or on the infrastructure? A driving simulator study. ResearchGate. https://www.researchgate.net/publication/362751474_eHMI_on_the_Vehicle_or_on_the_Infrastructure_A_Driving_Simulator_Study

Malik, J., Kim, N.-Y., Parr, M. D. N., Kearney, J. K., Plumert, J. M., & Rector, K. (2023). Do simulated augmented reality overlays influence street-crossing decisions for non-mobility-impaired older and younger adult pedestrians? Human Factors, 0018720823115
1280. https://doi.org/10.1177/0018720823115

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing pedestrian behavior in augmented reality — Proof of concept. Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA. https://doi.org/10.1109/vr46266.2020.00051

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Paiva, S., Ahad, M. A., Tripathi, G., Feroz, N., & Casalino, G. (2021). Enabling technologies for urban smart mobility: Recent trends, opportunities and challenges. Sensors, 21, 2143. https://doi.org/10.3390/s21062143

Peereboom, J., Tabone, W., Dodou, D., & de Winter, J. C. F. (2023). Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for

pedestrian safety in occluded scenarios. ResearchGate. https://www.researchgate.net/publication/371509780_Head-locked_world-locked_or_conformal_diminished-reality_An_examination_of_different_AR_solutions_for_pedestrian_safety_in_occluded_scenarios

Plabst, L., Oberdörfer, S., Ortega, F. R., & Niebling, F. (2022). Push the red button: Comparing notification placement with augmented and non-augmented tasks in AR. Proceedings of the 2022 ACM Symposium on Spatial User Interaction, Online, CA. https://doi.org/10.1145/3565970.3567701

Pratticò, F. G., Lamberti, F., Cannavò, A., Morra, L., & Montuschi, P. (2021). Comparing state-of-the-art and emerging augmented reality interfaces for autonomous vehicle-to-pedestrian communication. IEEE Transactions on Vehicular Technology, 70, 1157–1168. https://doi.org/10.1109/TVT.2021.3054312

Qualtrics XM. (2023). Qualtrics. https://www.qualtrics.com

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. Computers in Human Behavior, 133, 107289. https://doi.org/10.1016/j.chb.2022.107289

Rodríguez Palmeiro, A. R., van der Kint, S., Vissers, L., Farah, H., de Winter, J. C. F., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. Transportation Research Part F: Traffic Psychology and Behaviour, 58, 1005–1020. https://doi.org/10.1016/j.trf.2018.07.020

Sheridan, T. B. (2016). Recollections on presence beginnings, and some challenges for augmented and virtual reality. Presence: Teleoperators and Virtual Environments, 25, 75–77. https://doi.org/10.1162/PRES_e_00247

Sripada, A., Bazilinskyy, P., & De Winter, J. C. F. (2021). Automated vehicles that communicate implicitly: Examining the use of lateral position within the lane. Ergonomics, 64, 1416–1428. https://doi.org/10.1080/00140139.2021.1925353

Tabone, W., & De Winter, J. C. F. (2023). Using ChatGPT for human-computer interaction: A primer. Royal Society Open Science, 10, 231053. https://doi.org/10.1098/rsos.231053

Tabone, W., De Winter, J. C. F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B., & Stanton, N. A. (2021a). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transportation Research Interdisciplinary Perspectives, 9, 100293. https://doi.org/10.1016/j.trip.2020.100293

Tabone, W., Happee, R., García, J., Lee, Y. M., Lupetti, M. C., Merat, N., & De Winter, J. C. F. (2023a). Augmented reality interfaces for pedestrian-vehicle interactions: An online

study. Transportation Research Part F: Traffic Psychology and Behaviour, 94, 170–189. https://doi.org/10.1016/j.trf.2023.02.005

Tabone, W., Happee, R., Yang, Y., Sadraei, E., García, J., Lee, Y. M., Merat, N., & De Winter, J. C. F. (2023b). Immersive insights: Evaluating augmented reality interfaces for pedestrians in a CAVE-based experiment. ResearchGate. https://www.researchgate.net/publication/370160064_Immersive_Insights_Evaluating_Augmented_Reality_Interfaces_for_Pedestrians_in_a_CAVE-Based_Experiment

Tabone, W., Lee, Y. M., Merat, N., Happee, R., & De Winter, J. C. F. (2021b). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. 13th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, 209–218, Leeds, UK. https://doi.org/10.1145/3409118.3475149

Tong, Y., Jia, B., & Bao, S. (2021). An augmented warning system for pedestrians: user interface design and algorithm development. Applied Sciences, 11, 7197. https://doi.org/10.3390/app11167197

Tran, T. T. M., Parker, C., & Tomitsch, M. (2023). Scoping out the scalability issues of autonomous vehicle-pedestrian interaction. Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ingolstadt Germany, 167–177. https://doi.org/10.1145/3580585.3607167

Tran, T. T. M., Parker, C., Wang, Y., & Tomitsch, M. (2022). Designing wearable augmented reality concepts to support scalability in autonomous vehicle-pedestrian interaction. Frontiers in Computer Science, 4. https://doi.org/10.3389/fcomp.2022.866516

Transparency Market Research. (2023). Automotive Head-up Display (HUD) market. https://www.transparencymarketresearch.com/automotive-head-up-display-market.html

Troel-Madec, M., Boissieux, L., Borkoswki, S., Vaufreydaz, D., Alaimo, J., Chatagnon, S., & Spalanzani, A. (2019). eHMI positioning for autonomous vehicle/pedestrians interaction. Adjunct Proceedings of the 31st Conference on l'Interaction Homme-Machine, Grenoble, France. https://doi.org/10.1145/3366551.3370340

United Nations. (2020). United Nations A/RES/74/299 (A/RES/74/299). Resolution adopted by the General Assembly on 31 August 2020. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/226/30/PDF/N2022630.pdf?OpenElement

Unity. (2022). Unity 2021.3.13. https://unity.com/releases/editor/whats-new/2021.3.13

Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies, 5, 1–10. https://doi.org/10.1016/s0968-090x(96)00025-3

Varjo. (2023). Varjo XR-3 - The industry's highest resolution mixed reality headset. https://varjo.com/products/xr-3

Varjo Developer. (2023). Getting started with Varjo XR Plugin for Unity. https://developer.varjo.com/docs/unity-xr-sdk/getting-started-with-varjo-xr-plugin-for-unity

Varjo Technologies. (2023). Mixed reality – Varjo.com. https://varjo.com/use-center/get-to-know-your-headset/mixed-reality/

Wilbrink, M., Burgdorf, F. K. G., & Oehl, M. (2023). AR designs for eHMI–Communication between automated vehicles and pedestrians using augmented reality. Proceedings of the 8th Humanist Conference, Berlin, Germany.

World Health Organization. (2018). Global status report on road safety 2018: Summary. https://apps.who.int/iris/handle/10665/277370

Zhang, X., Wu, X., Cools, R., Simeone, A. L., & Gruenefeld, U. (2023). ARcoustic: A mobile augmented reality system for seeing out-of-view traffic. Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ingolstadt, Germany, 178–190. https://doi.org/10.1145/3580585.3606461

III

Part III: Diminished Reality

8

Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios

This chapter was submitted as: Peereboom, J*., *Tabone, W*.*, Dodou, D., & De Winter, J. C. F. (2023). Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios.

ABSTRACT

Introduction: *Many collisions between pedestrians and cars are caused by poor visibility, such as occlusion by a parked vehicle. Augmented reality (AR) could help to prevent this problem, but it is unknown to what extent the augmented information needs to be embedded into the world.* ***Methods:*** *In this virtual reality experiment with a head-mounted display (HMD), 28 participants were exposed to AR designs, in a scenario where a vehicle approached from behind a parked vehicle. The experimental conditions included a head-locked live video feed of the occluded region, meaning it was fixed in a specific location within the view of the HMD (VideoHead), a world-locked video feed displayed across the street (VideoStreet), and two conformal diminished reality designs: a see-through display on the occluding vehicle (VideoSeeThrough) and a solution where the occluding vehicle has been made semi-transparent (TransparentVehicle). A Baseline condition without augmented information served as a reference. Additionally the VideoHead and VideoStreet conditions were each tested with and without the addition of a guiding arrow indicating the location of the approaching vehicle. Participants performed 42 trials, 6 per condition, during which they had to hold a key when they felt safe to cross.* ***Results:*** *The keypress percentages and additional questionnaires showed that the diminished-reality TransparentVehicle and VideoSeeThrough designs came out most favourably, while the VideoHead solution caused some discomfort and dissatisfaction. An analysis of head yaw angle showed that VideoHead and VideoStreet caused divided attention between the screen and the approaching vehicle. The use of guiding arrows did not contribute demonstrable added value.* ***Conclusion:*** *AR designs that display a high level of local embeddedness are beneficial for addressing occlusion problems when crossing. However, the head-locked solutions should not be immediately dismissed because, according to the literature, such solutions can serve tasks where a salient warning or instruction is beneficial.*

8.1 INTRODUCTION

PEDESTRIAN safety represents a major issue, with approximately 310,000 pedestrian fatalities being recorded annually across the globe, a figure that accounts for 23% of total road traffic fatalities (World Health Organization, 2023). Pedestrian collisions frequently involve a motor vehicle as the impacting entity, contributing to 63% of such incidents in a developed country like the Netherlands (SWOV, 2020). Investigative analysis of accident casualties within the United States and Europe illustrates that pedestrian behaviour is implicated in approximately two-thirds of pedestrian-vehicle collisions, and that 10 to 15% of collisions stem from obstructed views (Bálint et al., 2021; European Road Safety Observatory, 2018; Hunter et al., 1996; Yue et al., 2020).

Improvements in pedestrian safety have been made through vehicle technology like radar and camera systems, which are used to detect approaching pedestrians, alert the driver, and autonomously initiate braking or steering. Improvements to these systems involve pedestrian path prediction, to identify pedestrians entering the road as early as possible (Rudenko et al., 2020), thereby increasing the time budget for the vehicle to respond. However, detecting a pedestrian or knowing a pedestrian's intention, may not always be possible, especially when the pedestrian steps onto the road from between parked vehicles (Palffy et al., 2023).

Additionally, pedestrian warning systems have been developed centred around mobile

devices, such as smartphones and smartwatches, to send alerts to pedestrians ahead of potential collisions (Bastani Zadeh et al., 2018; Liu et al., 2015; Won et al., 2020). In the current study, we extend this concept by investigating whether augmented reality (AR), which can be presented via a head-mounted display (HMD), could potentially offer a solution for pedestrians dealing with occlusion of approaching vehicles. More specifically, we explored: (i) Whether providing a live video stream of the occluded area can improve pedestrians' behaviour and perception, (ii) The impact of diminished reality on pedestrian behaviour and perception, where environmental occlusion is substituted with direct visibility.

A fundamental question in the design of any AR system is how the augmented feedback—in this case, video feed—should be positioned. There are different reference frames to be considered. Firstly, the information could be presented at a fixed position on the screen of the head-mounted display (HMD), an approach also known as a head-locked presentation (Lebeck et al., 2017). It is also possible to present the information at a fixed distance from the user's torso, an approach also referred to as a body-locked (Klose et al., 2019) or surround-fixed (Feiner et al., 1993) presentation. An alternative is to tie the information to the world, an approach also known as world-fixed (Feiner et al., 1993) or world-locked (Lebeck et al., 2017) presentation.

An advantage of head-locked and body-locked AR is the presence of information at an accessible position, which may induce a prompt response from the user, especially when information in the real world lies outside the user's immediate attention (Ghasemi et al., 2021; Schinke et al., 2010; Smith et al., 2021; Tabone et al., 2023). However, a risk is that users might respond to the cues presented without successfully integrating these with task-relevant real-world cues. This could occur either because the real-world cues have not been visually identified yet, or due to the challenges in switching cognitive and accommodative attention between the augmented and real-world cues (Chen et al., 2023; Dixon et al., 2014; Kerr et al., 2012). For similar reasons, head-locked information, or information that is otherwise not clearly locked to the world, can be difficult to use while walking and might induce discomfort, depending on whether the information is hard-locked or soft-locked to the user's head (Fukushima et al., 2020; Kaufeld et al., 2022; MagicLeap, 2020).

World-locked AR, on the other hand, offers the advantage of projecting cues at a contextually relevant location. Specifically, when a world-locked display aligns closely with the task at hand, it can offer an intuitive and seamless mode of information processing (Bauerfeind et al., 2021; Kim et al., 2018; Robertson et al., 2008; Schankin et al., 2017; Wickens, 2021; Zhao et al., 2023). One potential drawback of world-locked interfaces, however, is that the process of locating the world-locked information might consume more time for the user compared to head- or body-locked displays, where the information is directly accessible (De Oliveira Faria et al., 2020; Lee & Woo, 2023; Tabone et al., 2023).

It is noted that the classification into head-locked, body-locked, and world-locked reference frames does not completely capture the breadth of AR designs. Some propose that it is more beneficial to envisage a continuum of 'naturalism' (Pijnenburg, 2017) or 'local presence' (Rauschnabel et al., 2022), ranging from head-locked overlays, such as textual instructions or warning symbols, to the conformal presentation of virtual information that is seamlessly integrated into the real world (Kim et al., 2016; Wickens, 2021). The placement on this continuum is influenced not only by the choice of reference frame but also by the quality of depth rendering of virtual objects, amongst others (Rauschnabel et al., 2022).

Diminished reality (DR) refers to the process of removing or modifying specific elements from the environment, usually in real-time (Cheng et al., 2022; Mann & Fung, 2002; Mori et al., 2017). DR can be considered a subtype of AR: While AR is typically used to enhance the user's perception of the real world by overlaying virtual elements onto it, DR concerns the removal or reduction of certain elements from the physical world. There are several methods for achieving DR. One is inpainting, which involves removing an object from the current image and then filling in the gap with plausible background details to 'guess' what the background should look like (Elharrouss et al., 2020). Another approach is video see-through. This method involves using a remote video camera to capture the scene beyond the occluding object, and processing the video feed, before it is displayed to the user (e.g., Meerits & Saito, 2015; Rameau et al., 2016). In automotive settings, a number of see-through displays have been incorporated. In Samsung's Safety Truck, for example, live video images were displayed on the back of the truck, effectively granting trailing drivers the ability to 'see through' the truck (Samsung, 2015; see also Zhang et al., 2018). Gomes et al. (2012) and Rameau et al. (2016) developed a 'see-through cars' system, which provided drivers an unobstructed view of the road. Other researchers have used the concept of transparency, primarily in virtual environments. In this regard, Yasuda and Ohama (2012) tried to solve the problem of poorly visible intersections by making a wall semi-transparent, so that the approaching car could be seen through the wall. Lindemann et al. (2019) examined a semi-transparent cockpit which provided drivers with the ability to see parts of the environment that are usually blocked by the car body.

In the current study, we investigated a series of AR solutions aimed at solving the occlusion problem of pedestrians. The investigated designs ranged from a head-locked video feed of the occluded area (thus low local presence according to Rauschnabel et al., 2022's framework), a solution in which the same video feed was at a fixed distance from the user on the opposite side of the road (medium local presence), to two diminished-reality solutions that blended in with the environment (high local presence). Inspired by the above-described DR solutions in the automotive domain, two DR implementations were chosen, namely a see-through video feed of the occluding vehicle and a solution in which the occluding vehicle was made semi-transparent. Given that the different positions on the local-presence dimension may carry distinct advantages and disadvantages, as delineated above, we refrained from formulating explicit hypotheses. Instead, our primary interest lies in discerning which AR condition in the pedestrian-crossing scenario shows the most and least favourable results in terms of perceived pedestrian safety, workload, comfort, and acceptance.

Due to the challenges associated with implementing AR in real-world settings, we have chosen to test the proposed solutions in a virtual environment using a high-end HMD. While this approach may have its drawbacks, such as potential disparities in attentional switching as compared to AR evaluated in real-world settings (Gabbard et al., 2019), the use of virtual reality offers various advantages. These include improved experimental control and the mitigation of technical issues such as difficulties in precisely anchoring objects to the world (Walter et al., 2019; Wiesner, 2019).

As indicated above, there is a challenge of divided attention between the task-intrinsic content in the world (i.e., the approaching car) or on the augmented feedback (i.e., the video feed). Prior studies have focused on directing the user's view by means of arrows (Schinke et al., 2010) or attracting it by means of bounding boxes (Chen et al., 2015; Orlosky et al.,

2019), attention funnels (Renner & Pfeiffer, 2017), flickering (Schmitz et al., 2020; Waldin et al., 2017), and contrasting (Lu et al., 2012) around points of interest. In the present study, we additionally investigated whether cues in the form of a 3D arrow that continuously indicates where the task-intrinsic information (i.e., approaching car) is located provide added value compared to just using the video feed.

8.2 METHOD

8.2.1 PARTICIPANTS

A total of 28 individuals, 23 of whom were males, aged between 19 to 32 years (M = 25.0, SD = 3.1), participated in the experiment. Participants were students and doctoral candidates from different faculties at Delft University of Technology. The recruitment process did not offer incentives and welcomed individuals regardless of their driving experience, nationality, driving-side orientation, or age. The study was approved by the Delft Human Research Ethics Committee, approval no. 1817. Each participant provided written informed consent before the start of the experiment.

Demographic characteristics were recorded by means of a pre-experiment questionnaires. The sample predominantly comprised Dutch nationals (n = 19), but also included individuals of Indian, Italian, Belgian, Russian, and Maltese nationality. Most participants held a driver's licence (n = 25), and among these, the average driving experience was 7.12 years (SD = 3.0). Furthermore, most participants were regular pedestrians in urban environments, with 12 participants walking every day and 10 walking 4–6 days per week. In terms of digital entertainment, 10 participants reported playing video games several times a week, 8 playing approximately once a month, and 10 rarely playing or not playing anymore.

8.2.2 MATERIALS

The experiment was executed using Unity 2019.4.3f1 on an Alienware PC, paired with a Varjo VR-3 HMD. The PC was equipped with an Intel i7-9700K CPU and a NVIDIA GeForce RTX 2080 Ti GPU. Participants used the spacebar on the PC's keyboard to indicate their readiness to cross. The Varjo VR-3 HMD featured a high-resolution display, offering a 90 Hz refresh rate and a 115° horizontal field of view. The focus area, spanning 27° × 27°, was rendered at 70 pixels per degree on a μOLED display, providing 1920 × 1920 pixels per eye. Meanwhile, the peripheral area was rendered at about 30 pixels per degree on an LCD, producing 2880 × 2720 pixels per eye. Additionally, the Varjo VR-3 offered foveated rendering via integrated eye-tracking.

Four SteamVR base stations enabled the positional tracking of the Varjo HMD. The frame rate of the simulation was set to 30 frames per second. Audio was delivered through a Jabra Evolve stereo headset. After each experimental session, all equipment was sanitised with alcohol wipes. The experimental setup is depicted in Figure 8.1.

8.2.3 PARTICIPANT TASK

In the experiment, participants were instructed to press a response key when they felt safe to cross the road. More specifically, the participants were given the following task instructions: whenever they perceived crossing as safe, they were to press and hold down the spacebar on the provided keyboard. This action was to be sustained for the duration of perceived safety.



Figure 8.1: The experimental setup.



Figure 8.2: The virtual street in which the experiment took place.

Should they feel it was no longer safe to cross, they were to stop pressing the spacebar. They were allowed to engage and disengage the key as many times as they deemed necessary. Before each trial, participants received the verbal auditory instruction “Press now” to indicate that they should press and hold the key.

8.2.4 VIRTUAL ENVIRONMENT

The study used an open-source simulator (Bazilinskyy, 2020; Bazilinskyy et al. (2020b), designed using the Unity game development platform. The simulated setting is an urban city centre, with two-lane roads and static elements like buildings, parked cars, and trees (Figure 8.2).

In this study, the participant assumed the role of a pedestrian, placed on a curb behind a parked Nissan 1400 ‘Bakkie’ and a Ford Mustang GTO. This positioning largely impeded the view of the road, with an approaching vehicle, specifically a Smart Fortwo, approaching from the pedestrian’s left.

The in-game camera, representing the pedestrian’s perspective, was set at fixed height of 1.67 m, which is close to the 1.65 m average eye height for Dutch adults between 20 and 30 years old (DINED, 2020). The pedestrian was stationed on the 0.22 m-high curb and 2.5 m away from the road edge orthogonally. A fixed camera position was used; only the rotation of the head affected what participants viewed. This approach was adopted to maintain a consistent perspective and degree of occlusion of the approaching vehicle for all trials and participants. The pedestrian was represented by an avatar; this avatar was visible when the participant looked down. However, the avatar was static and did not respond to the participant’s movements.

The road spanned 10 m in width. The obstructive vehicle, a Nissan pickup truck, had dimensions of 3.8 m in length, 1.7 m in height, and 1.67 m in width. The pedestrian was located 4.25 m behind the rear of the pickup truck and remained stationary throughout the duration of the simulation. A top-down review of the distances is provided in Figure 8.3.

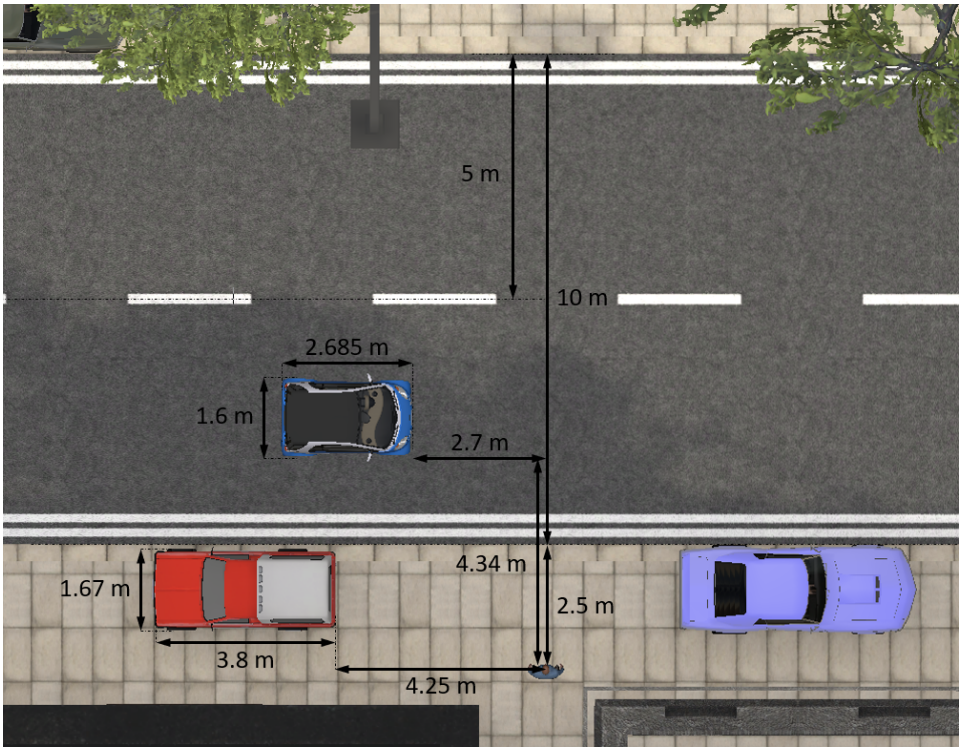


Figure 8.3: Top-down overview of the virtual environment. In this instance, the vehicle is positioned at the stopping point in the yielding scenario.

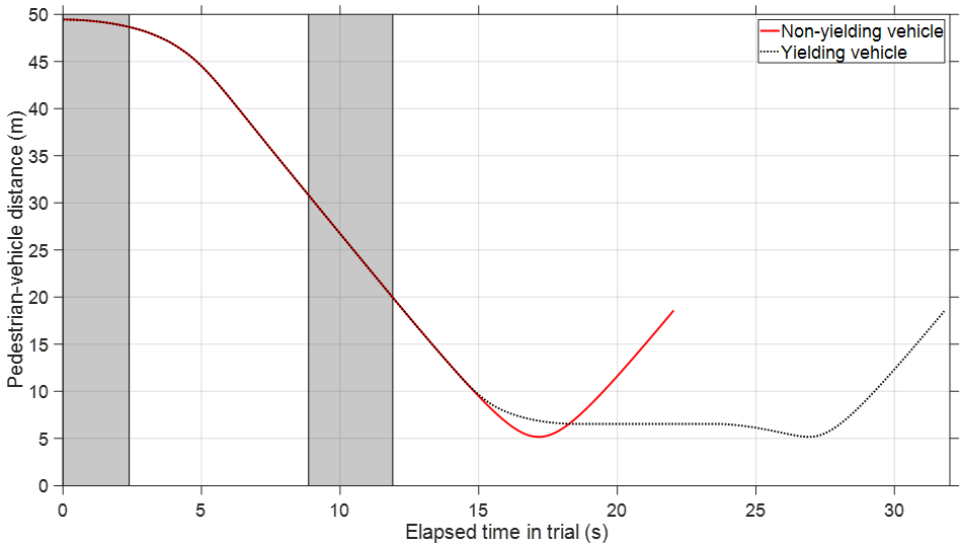


Figure 8.4: Distance between the approaching vehicle and pedestrian as a function of time. Grey backgrounds represent intervals where the vehicle could not be seen in the *Baseline* condition because it was fully occluded by the parked vehicle. Note that the yielding vehicle started to decelerate at an elapsed time of 14.3 s, came to a full stop at 18.4 s, drove away at 23.4 s, and passed the pedestrian at 26.9 s. The non-yielding vehicle passed the pedestrian at an elapsed time of 17.2 s. Distances were calculated based on the Euclidean norm, using the pedestrian's location and the centre point of the vehicle.

In all trials, the car started from a standstill. It accelerated and made a 90-degree left turn to then approach the participant; after 4.6 s, the car had completed this turn and had reached an approach speed of 15 km/h. We have opted for a low vehicle speed of 15 km/h, as a lower speed affords the pedestrian more time to respond, thus allowing for a more effective comparison of experimental conditions. Furthermore, a low speed introduces a degree of ambiguity regarding whether the AV will stop. In comparison, if the speed of the vehicle is high, then it is evident that crossing in front of this vehicle is not a safe option, and additional explicit signals will therefore have relatively little influence on the pedestrian's crossing intentions (see Dey et al., 2021 and Onkhar et al., 2022 for a similar argumentation).

In non-yielding scenarios, the vehicle maintained this 15 km/h speed until the end of the trial. In yielding scenarios, the vehicle initiated deceleration at a rate of 1 m/s^2 at an elapsed time of 14.3 s, 11.64 m from the pedestrian. The vehicle stopped at 18.4 s, positioned 6.55 m from the pedestrian. Following the halt, the vehicle remained stationary for 5 s before recommencing motion. Yielding trials consumed a total duration of 31.8 s, whereas non-yielding trials took 22.0 s to complete. A representation of the pedestrian-vehicle distance versus time relationship is shown in Figure 8.4.

Each trial began with the sound of a starting engine from the AV. As the vehicle drove, it produced a humming sound of a combustion engine. The sound perceived by the participant depended on the distance to the AV, including the Doppler Effect.

The approaching vehicle was rendered in cyan, as this colour bears no established connotations in signalling yielding or non-yielding behaviours to pedestrians (Bazilinskyy

et al., 2020a). To mimic a dart-out crossing scenario more closely, a zebra crossing was omitted from the design. The inclusion of such a crossing would implicitly suggest safety for the pedestrian to cross (in the Netherlands, traffic law mandates stopping for pedestrians poised at the curb).

8.2.5 AUGMENTED REALITY DESIGNS

The experiment included a total of six AR designs, as presented in Table 8.1, with an additional condition without additional functioning as the baseline for comparisons against the other designs.

Two DR solutions were designed, a see-through display (*VideoSeeThrough*) and a semi-transparent parked vehicle (*TransparentVehicle*), along with two video feed interfaces, namely a head-locked (*VideoHead*) and a body- and world-locked version (*VideoStreet*). An additional feature was implemented in the *VideoHead* and *VideoStreet* designs to assist participants' views, using a continuous 'waypoint arrow' pointing to the moving vehicle. The arrow was positioned in the central top portion of the video feeds. Figure 8.5 provides screenshots of the six AR designs and the Baseline tested in this experiment. Additionally, the online data repository includes a screen capture video, illustrating the various AR designs from the participant's perspective.

According to Rauschnabel et al. (2022)'s local-presence dimension, *VideoHead* exhibits a low local presence; the video feed is projected into the field of view but is not embedded in the world. *VideoStreet* can be described as having a medium local presence; it is present as a floating screen at a fixed position on the other side of the street, thus not truly part of the world. The two DR designs demonstrate a high local presence as the information provided is strongly embedded in the real world: *VideoSeeThrough* is presented as allowing visibility through the parked vehicle, while the transparent car also allows the participant to see through the parked vehicle.

Table 8.1: Overview of the six augmented reality designs.

Name	Anchoring	Video feed	Diminished reality	Guiding arrow	Local presence
VideoHead	Head-locked	Yes	No	No	Low
VideoHead+	Head-locked	Yes	No	Yes	Low
VideoStreet	Body- and world-locked	Yes	No	No	Medium
VideoStreet+	Body- and world-locked	Yes	No	Yes	Medium
VideoSeeThrough	World-locked	Yes	Yes	No	High
TransparentVehicle	World-locked	No	Yes	No	High

The positioning of the *VideoHead* display was in the upper section of the field of view, as per recommendations by Klose et al. (2019). This upper placement enabled unobstructed viewing of the road and other objects at the lower field of view. With dimensions of 8 cm both in height and width, the display was positioned 36 cm away from the midpoint between



Figure 8.5: The six augmented reality designs (screenshot resolution 2880 × 2720, consistent with the resolution of the displays in Varjo VR-3).

the pedestrian's eyes. This arrangement yielded a field of view angle, both vertical and horizontal, of 12.7°.

The *VideoStreet* display was similarly positioned above road level and objects on the curb, including a bench and a waste bin, ensured an unobstructed view. The choice of positioning was also influenced by prior research into AR in road-crossing scenarios (Tabone et al., 2021) and the contemporary traffic system, wherein traffic signals and pedestrian crosswalks typically appear perpendicular to or across the road, was a further motivation for this positioning. We consider the *VideoStreet* presentation as world-locked because it occupies fixed coordinates in the virtual world, as well as body-locked, because the participant in our experiment does not translate within the environment, and thus, the VideoStreet display always remains at a constant distance from the participant. The *VideoStreet* screen had a height of 300 cm and width of 400 cm, and was positioned 12.7 m distant from the pedestrian, leading to field of view angles of 13.5° vertically and 17.9° horizontally.

The *VideoSeeThrough* display, incorporated into the rear of the pickup truck, was 36 cm in height and 89 cm in width. When the pedestrian was positioned 4.25 m from the rear of the truck, it encompassed a vertical and horizontal field of view with angles of 4.9° and 12.0°, respectively. As for the *TransparentVehicle*, the transparency coefficient (alpha value) was set to 0.49 to generate a semi-transparent visual presentation.

The video feed was captured by a stationary camera unit centrally positioned and mounted on the top of the windshield of the pickup truck (Figure 8.6, top left). This camera was oriented to face forward in alignment with the road, and provided a field of view spanning 60°. Figure 8.6 shows examples of the camera view in various conditions.

8.2.6 EXPERIMENTAL DESIGN

This study used a within-subjects design, with each participant performing all seven experimental conditions. Conditions were grouped into blocks, each containing six randomly allocated yielding and non-yielding scenarios to offset expectancy effects. The sequence of block presentations was managed using Bradley's (1958) balanced Latin Square method. In total, the experiment included 1176 trials, namely 28 participants who each conducted 6 trials for each of the 7 experimental conditions (6 AR designs + *Baseline*).

8.2.7 QUESTIONNAIRES

Participants were provided with information regarding the study purpose, the procedural layout of the experiment, the data management process, and their right to abstain at any point, via an informed consent document. This was followed by the administration of a pre-experimental questionnaire, soliciting information on general demographics, commuting behaviour, and prior experiences with VR and gaming.

Upon the completion of each trial block, participants were prompted to provide a measure of their current state of well-being, using the Misery Scale (MISC; Bos et al., 2005). In instances where a participant's MISC score was observed to be 4 or greater, a two-minute break was offered. Following this, the NASA-TLX questionnaire (Hart & Staveland, 1988) was administered as a means of assessing workload. This questionnaire incorporated six variables: (1) mental demand, (2) physical demand, (3) temporal demand, (4) performance, (5) effort, and (6) frustration, evaluated via a 21-point scale ranging from 'perfect' to 'failure' for performance, and 'very low' to 'very high' for the other five items. Lastly, acceptance of



Figure 8.6: Top left: Windscreen camera placement. Top right: Close-up view of the *VideoHead+* concept. Bottom left: Close-up view of the *VideoSeeThrough* concept. Bottom right: Close-up view of the *TransparentVehicle* concept.



Figure 8.7: Post-block questionnaires, displayed above the road in front of the participant. *Note:* This screenshot shows a zoomed-out external view. During the experiment, the participant could turn their body and head to direct visual attention to one of the three questionnaires.

the seven experimental conditions in terms of usefulness and satisfaction was evaluated using a questionnaire developed by Van der Laan et al. (1997). In this questionnaire, participants were asked to rate nine semantic-differential items on a five-point Likert scale. Usefulness was calculated as the mean of the following five items: 1. useful–useless; 3. bad–good; 5. effective–superfluous; 7. assisting–worthless; and 9. raising alertness–sleep-inducing, and satisfaction was calculated based on the following four items: 2. pleasant–unpleasant; 4. nice–annoying; 6. irritating–likeable; and 8. undesirable–desirable.

Due to the logistical impracticalities associated with asking participants to disengage from the VR environment in order to complete questionnaires, all post-block questionnaires were conducted within the virtual environment itself, with the participant still wearing the HMD. Participants were directed to verbally provide their responses, which were then recorded by the experimenter via Google Forms on a laptop. Figure 8.7 depicts the questionnaires used to evaluate well-being, workload, and acceptance.

8.2.8 DATA ANALYSIS

Data was collected at a frequency of 50 Hz using a data logging script within Unity. No trials were excluded from the analysis.

For the analysis of keypress data, plots were created depicting the mean percentage of trials in which the key was pressed over time. The keypress percentage per AR design was computed by calculating the mean over the period from 8.86 to 14.30 s, namely from the moment the approaching vehicle was completely occluded by the parked vehicle until the approaching vehicle in the yielding condition began to brake. The initial seconds of the trials are less relevant to our analysis, as the approaching vehicle accelerated and executed a turn, and participants still had to press the key in response to the ‘press now’ instruction. The

percentages for each participant were averaged across the six trials to create a single score per participant. This averaging helps meet the assumptions of independence and normality, which are needed for valid statistical analysis, unlike the approach of using scores from each trial individually.

For the NASA-TLX responses, the 21-point scores were converted into percentage values. A composite score was then obtained by averaging the scores across the six items (Byers et al., 1989). For the acceptance questionnaire, the responses for Items 1, 2, 4, 5, 7, and 9 were mirrored, so that a higher score corresponds to higher usefulness/satisfaction. Next, scores were offset from the 1 to 5 scale (see Figure 8.7) to correspond with the original scale of -2 to +2.

The seven interface conditions were compared per dependent variable using a linear mixed effects model using Mathworks MATLAB R2023b. In this model, the interface condition is a categorical fixed effect, the interface block number (i.e., a number from 1 to 7 indicating whether that condition was presented first, second, third, fourth, fifth, sixth, or seventh for that participant) is a fixed effect and covariate. The participant number (1 to 28) was submitted as a random effect. The model was estimated using the maximum likelihood method, and the number of degrees of freedom was determined using the residual method. The covariate 'block number' was included to determine whether there was a learning/experience effect for the participants in our counterbalanced design; including this covariate enables a more powerful determination of the differences between the seven interface conditions.

Differences between pairs of conditions were assessed by evaluating the overlap of 95% confidence intervals. To calculate these confidence intervals, a method for within-subjects designs was used, as described by Morey (2008). The scores per dependent variable were first linearly detrended to correct for the above-mentioned learning/experience effects over the seven blocks.

8.3 RESULTS

ALL 28 participants completed the experiment. The MISC scores were overall low, with averages of 0.50, 0.82, 1.14, 1.25, 0.96, 1.18, and 1.21 for Blocks 1 through 7, respectively. Out of the 196 trials completed (28 participants × 7 conditions), a MISC score of 4 occurred 14 times and a MISC score of 5 occurred once. These scores were attributed to 7 of the 28 participants. Among them, 5 participants agreed to take a break after completing a trial, with the breaks lasting 3–5 minutes.

Figure 8.8 shows the key presses as a function of time for the seven conditions in the non-yielding scenario. Upon the approach of the vehicle, the majority of participants felt safe to cross, indicated by more than 80% of them keeping the key pressed. It can be seen that there are substantial differences between conditions, with participants feeling safest to cross in the *TransparentVehicle* condition, and least safe in the *Baseline* condition. In the Baseline condition, participants tended to release the response key as soon as the vehicle was behind the parked Nissan. For completeness, the same figures for the entire trial and for yielding and non-yielding vehicles separately are available in the Supplementary Material (Figures 8.11 and 8.12).

Table 8.2 presents the results of the linear mixed effects model, distinguishing between the effect of interface condition and the learning/experience effect. Figure 8.9 also shows

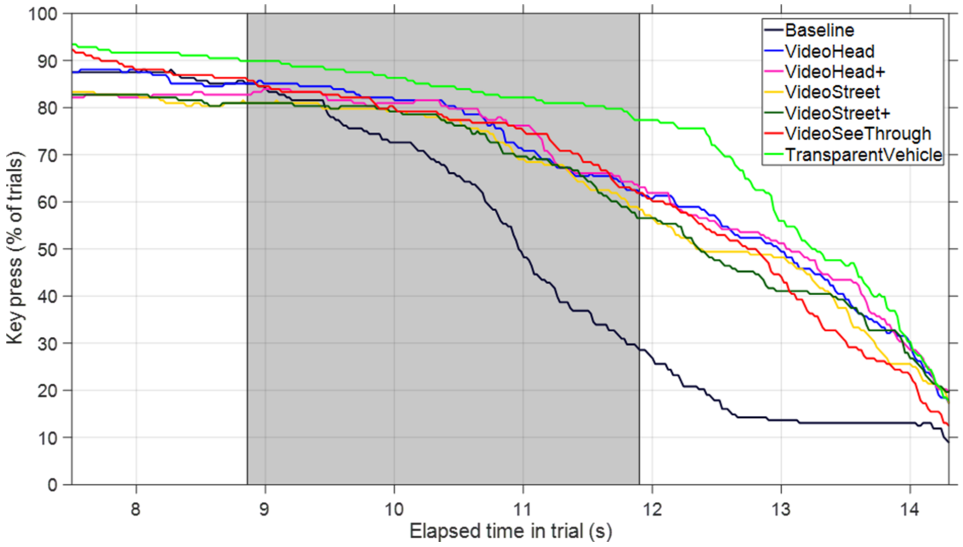


Figure 8.8: Key-press percentages for the seven conditions in the non-yielding scenario. The grey background represents the interval where the approaching vehicle could not be seen in the *Baseline* condition because it was fully occluded by the parked vehicle. Across the depicted interval, the vehicle drove 15 km/h.

the means and 95% confidence intervals, after applying a linear detrending to correct for the learning/experience effect. The effects of the interface condition are described below.

- The perception of safety, as measured through key-presses across the 8.86–14.30 s interval showed significant differences between the seven experimental conditions, with the highest score for the *TransparentVehicle*, and the lowest score for the *Baseline* condition (Figure 8.9, top left). The performance of the *VideoHead* and *VideoStreet* designs, including their guidance-included variations, was equivalent on this metric.

- Usefulness (Figure 8.9, top middle) and satisfaction (Figure 8.9, top right) showed significant differences between conditions. The *TransparentVehicle* was found to be the most useful and satisfying, followed by the *VideoSeeThrough*. A negligible difference was observed in the scores between the standard *VideoStreet* design and the one equipped with the guiding arrow, both of which attained positive scores on the scales of satisfaction and usefulness. A similar negligible difference is noted in the scores between the *VideoHead* design and the guidance-included version, although these two conditions score negatively on the satisfaction scale. Importantly, there was no substantial difference in the usefulness scores between *VideoHead* and *VideoStreet*, both with and without the guiding arrow. The *Baseline* condition attained the lowest scores, with net negative values on both the satisfaction and usefulness scales.

- Self-reported workload (Figure 8.9, bottom left) also showed significant differences between conditions, with the *TransparentVehicle* showing the lowest workload, and the *VideoHead* the highest.

- The MISC scores (Figure 8.9, bottom middle) showed significant differences between conditions, with the highest discomfort for the two *VideoHead* conditions.
- Upon completion, the participants ranked the AR solutions. The mean ranks (Figure 8.9, bottom right) showed significant differences between conditions. Participants preferred the *Baseline* condition over the *VideoHead* display, while the *TransparentVehicle* emerged as the most favoured by a considerable margin, followed by the *VideoSeeThrough* display. The *VideoHead* designs were the least preferred.

Table 8.2: Results of linear mixed model analysis for the six dependent variables.

	Interface condition		**Block number, (covariate)**		**Block number sign of effect**
	F	*p*	*F*	*p*	
Key-press (%)	14.0	3.53×10^{-13}	19.1	2.07×10^{-5}	+
Usefulness (-2 to 2)	23.8	6.91×10^{-21}	0.48	0.488	-
Satisfaction (-2 to 2)	36.0	7.90×10^{-29}	57.26	0.008	-
Workload (%)	7.22	5.96×10^{-7}	3.97	0.048	+
MISC (0 to 10)	4.06	7.44×10^{-4}	18.7	2.44×10^{-5}	+
Preference rank (1 to 7)	26.2	1.50×10^{-22}	2.52	0.114	+

Note. For Interface: df1 = 6, df2 = 188; for Block number: df1 = 1, df2 = 188. The block number sign indicates whether the scores on the dependent variable increased or decreased with the block number.

It should also be noted that the seven mean values shown in Figure 8.9 were found to correlate with each other. For example, the mean key-press percentage correlated strongly with mean self-reported usefulness (r = 0.87). Additionally, the mean perceived satisfaction strongly correlated with mean perceived usefulness (r = 0.82), mean perceived workload (r = -0.92), and the mean preference rank (r = -0.97).

8.3.1 HEAD MOVEMENT ANALYSIS

In conditions featuring high local presence, namely *VideoSeeThrough* and *TransparentVehicle*, as well as with the *Baseline* condition, participants predominantly glanced leftward. This is demonstrated by a yaw angle approximating 125°, corresponding to the location of the DR interfaces within the virtual environment Figure 8.10, left).

On the contrary, engagement with the *VideoStreet* displays, which were positioned across the road, resulted in participants mainly focusing straight ahead with a yaw angle near 175°, while intermittently looking leftward to spot the approaching vehicle (Figure 8.10, right).

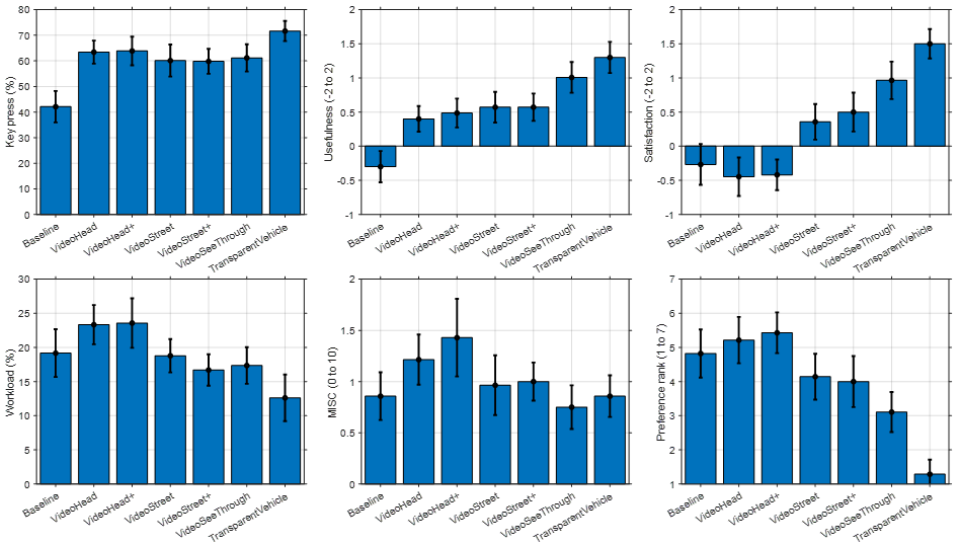


Figure 8.9: Means over the participants with 95% confidence intervals for the seven experimental conditions, for the dependent variables of this study. A linear trend was removed before calculating means and confidence intervals to compensate for block presentation order.

The histogram plot of the *VideoHead* conditions reveals less defined peaks, with a peak at 130°suggesting that participants predominantly focused on the parked vehicle or directly across the road at 180°(Figure 8.10, middle).

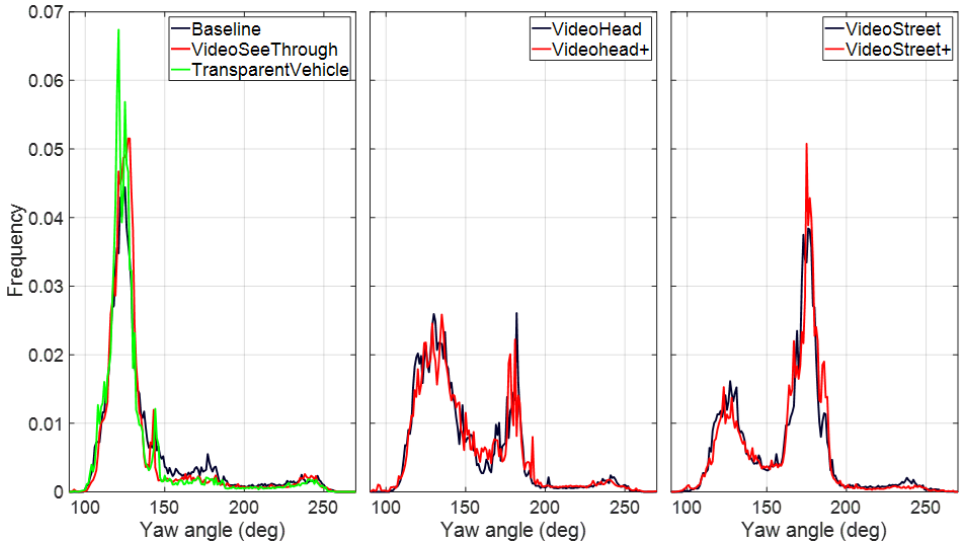


Figure 8.10: Distribution of yaw head rotation angles analysis during the vehicle's approaching phase (0 to 14.3 s). Note that a yaw angle of 180 degrees indicates the participant gaze across the road, with angles lesser or greater than 180 degrees suggesting a leftward or rightward gaze, respectively.

8.4 DISCUSSION

THIS study used a virtual reality setup with a HMD to examine the impact of six distinct AR designs on participants' perceived safety when crossing. The results show that pedestrians' perceived safety can increase by eliminating obstructions to their view. Particularly, in the *Baseline* condition, participants released the key earlier than with the other conditions, reflecting a diminished sense of safety. This may be attributed to participants potentially feeling cautious, given the view-blocking parked vehicle. It should be acknowledged here that the subjective feeling of safety does not necessarily imply objective safety. Nevertheless, the results clearly demonstrate that eliminating occlusion through innovative augmented reality solutions has the potential to positively impact participants' perceptions, as measured both during the experimental trials via the response key and subsequent to trial completion via a questionnaire.

All six AR designs can be considered effective compared to the *Baseline*, in terms of objective key-presses and subjective usefulness (see non-overlapping confidence intervals in Figure 8.9). That is participants, to a greater or lesser extent, made use of the additional information that made the occluded vehicle visible. However, when considering the results on the dimension of low local presence (*VideoHead*), medium local presence (*VideoStreet*), to high local presence (*VideoSeeThrough*, *TransparentVehicle*) (Rauschnabel et al., 2022), higher local presence was more advantageous. The transparent car, which was fully conformal and did not introduce new elements to the environment, received particularly high scores. This result is consistent with the proximity compatibility interface-design principle (Wickens & Carswell, 1995), which posits that when task elements need to be processed concurrently, it is advantageous to present corresponding cues together and integrated, rather

than separately. In this respect, it is useful that the estimation could be made by directly looking at the vehicle rather than elsewhere, such as on a separate screen.

The head-locked interfaces did not receive satisfactory ratings, and they appeared to be less preferred when compared to the *Baseline*. This observation can be linked to the literature in the introduction which indicated that head-locked displays may cause discomfort. In our case, the substantial degree of accommodation necessary for the *VideoHead* screen, which was presented at a close distance, could be an auxiliary explanatory factor.

Our analysis of head movements supports the notion that AR designs with a high degree of local presence permit participants to focus their attention on a single location, thereby obviating the need for attention division. While the *VideoHead* conditions somewhat alleviate the issue of divided attention between the video feed and the car (given that the video feed remains constantly in view), attentional switching was still evident. While displaying information on the opposite side of the road constitutes a familiar and useful approach, evidenced, for example, by pedestrian traffic signals (Tabone et al., 2023), a limitation is that it requires pedestrians to allocate their attention between the information across the road and the oncoming vehicle. The result is a pronounced bimodal distribution in the frequency of head movements (see Figure 8.10, right).

Our study established that information fully integrated within the environment yielded superior results compared to supplementary displays in the environment. This outcome appears to contradict a study by Tabone et al. (2023), where they found that a head-locked display was preferred over cues projected on the road or the approaching vehicle itself. A plausible explanation is that, in the study by Tabone and colleagues, the head-locked information comprised an unambiguous and dependable message (text: “danger! vehicle is approaching”/“safe to cross” or a red/green pedestrian traffic light) which pedestrians could use to determine whether or not to cross the road. Such explicit information could be particularly advantageous when the pedestrian is visually distracted or has not yet identified the approaching vehicle in the environment. In our case, however, the extra screens did not provide any information that was not available in the DR solutions. Another factor is that the head-locked displays in Tabone et al. (2023) were presented in a CAVE-based simulator rather than a head-mounted display, which reduces the likelihood of discomfort (Kim et al., 2012; Pala et al., 2021).

Although previous research has demonstrated that directional arrows in AR may have a beneficial effect on performance, this has primarily been established in the context of tasks requiring navigation or spatial orientation (Gabbard et al., 2019; Liu et al., 2021; Markov-Vetter et al., 2020). In our experiment, the guiding arrow did not have any meaningful impact on the dependent variables. A plausible explanation is that the vehicle always came from the left, and the participant, who conducted a total of 42 trials, eventually became aware of the vehicle’s origin. A second explanation concerns divided attention. As mentioned above, the augmented screens entailed that the participant had to divide attention between the screen and the approaching traffic. The introduction of an arrow, in addition to the screen, would require even more need for divided attention, something participants may have preferred to avoid by ignoring the arrow.

The present experiment was executed in a virtual environment to ensure a controlled comparison of conditions and to minimise the risk of technological glitches. Designs of AR solutions with minimal local presence, such as the *VideoHead* design, would demand

the least complex technology for real-life applications (Rauschnabel et al., 2022). More specifically, a live video stream would need to be relayed and then displayed on the HMD. This would require that (parked) vehicles will be equipped with streaming cameras, which is a reasonable assumption given that automated vehicles already use cameras to perceive their surroundings. The *VideoStreet* design additionally requires visual tracking technology. Here, an HMD-embedded camera would need to determine certain world features like the ground surface and road layout, and the screen could be presented at a predetermined height above this surface. In turn, the implementation of the two DR solutions requires more complex tracking technology, where the parked vehicle is recognised by the HMD camera, following which the vehicle is either rendered transparent or augmented with a screen. This may require inpainting techniques (e.g., Ardino et al., 2021; Elharrouss et al., 2020; Liao et al., 2020) to estimate the image behind the parked car. The approaching vehicle can then be displayed or simulated on this inpainted image, where the position and orientation of this vehicle can be obtained from a wireless broadcast from the vehicle. Despite the fact that previous DR applications have been demonstrated in real-world automotive contexts, and the concept of essentially rendering objects invisible presents an intriguing notion, the technological prerequisites along with the potential risks of visual artefacts, such as jitter or perspective distortions, are substantial (Overmeyer et al., 2023; Rameau et al., 2016; Wilmott et al., 2022).

A limitation of our study is that it examined fixed display sizes and positions. Investigating displays of different dimensions could provide additional insight, as the current video feed displays were relatively small. Furthermore, a more optimal positioning of the screen, especially in regard to the distance from the user, may aid in reducing feelings of discomfort. Additionally, in our experiment, the pedestrian remained stationary on the curb and could only visually scan the environment. For future research, it is recommended to consider either a CAVE-based simulator (Kaleefathullah et al., 2022) or a motion suit combined with an HMD (Kooijman et al., 2019) as alternatives to the current fixed camera position used in the present study. These setups may offer less constrained environments and allow for more comprehensive measurements of pedestrian behaviour, including walking and crossing actions, as part of the perception-action cycle. It should be noted that in these cases, extra thought needs to be given to the design of the *VideoStreet* and *VideoSeeThrough* designs, in regard to the perspective of the screens.

Another limitation is that the current study investigated attention distribution using head-tracking because eye-tracking data was not stored. The head-tracking data seemed sufficiently sensitive to differentiate between conditions (see Figure 8.10), although our previous research with a Varjo VR-2 HMD showed that eye-tracking provides sharper peaks in the distribution of the yaw angle than head-tracking data (Mok et al., 2022). This can be explained by the fact that participants can rotate their eyes to focus on a target, and therefore head orientation is only a proxy of where participants focus their attention.

A final limitation is that only one approaching vehicle was used in the current experiment. Further increasing the realism of the crossing situation could involve incorporating other vehicles. Such vehicles would increase visual demands and require divided attention.

8.5 CONCLUSION

THIS study shows the promise of removing visual occlusions through AR in increasing pedestrians' feeling of safety. The findings indicate a user preference for AR solutions that exhibit a high degree of 'local presence'—in this case, diminished reality solutions using see-through video and transparent vehicle renderings. These designs also demonstrated the least demand for divided attention. It is important to consider, however, the technological feasibility and complexity of implementing such solutions in a real-world scenario.

The study also revealed relatively high workload and discomfort from the head-locked AR display, as participants juggled between the close-by AR screens and real-world cues. The guiding arrows in video feeds did not improve performance. Finally, although the virtual reality experiment ensured control and safety, it might not reflect the complexities of implementing AR in real-world settings.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 860410 (project name: SHAPE-IT).

DATA AVAILABILITY

A video capture of an experiment, and data and scripts that reproduce the figures and tables presented in the paper are available at the public data repository: https://doi.org/10.4121/e0fd7ca5-7cb6-4ce2-8d42-7a62e867e1d1 . The repository with the version of the simulation used in the study is available at https://github.com/bazilinskyy/coupled-sim/tree/diminished-reality-experiment.

REFERENCES

Ardino, P., Liu, Y., Ricci, E., Lepri, B., & De Nadai, M. (2021). Semantic-guided inpainting network for complex urban scenes manipulation. 2020 25th International Conference on Pattern Recognition, 9280–9287, Milan, Italy. https://doi.org/10.1109/ICPR48806.2021.9412690

Bálint, A., Labenski, V., Köbe, M., Vogl, C., Stoll, J., Schories, L., Amann, L., Sudhakaran, G. B., Huertas Leyva, P., Pallacci, T., Östling, M., Schmidt, D., & Schindler, R. (2021). Use case definitions and initial safety-critical scenarios (Report No. D2.6. Project SAFE-UP).

Bastani Zadeh, R., Ghatee, M., & Eftekhari, H. R. (2018). Three-phases smartphone-based warning system to protect vulnerable road users under fuzzy conditions. IEEE Transactions on Intelligent Transportation Systems, 19, 2086–2098. https://doi.org/10.1109/TITS.2017.2743709

Bauerfeind, K., Drüke, J., Schneider, J., Haar, A., Bendewald, L., & Baumann, M. (2021). Navigating with augmented reality – How does it affect drivers' mental load? Applied Ergonomics, 94, 103398. https://doi.org/10.1016/j.apergo.2021.103398

Bazilinskyy, P. (2020). coupled-sim. https://github.com/bazilinskyy/coupled-sim

Bazilinskyy, P., Dodou, D., & De Winter, J. (2020a). External human-machine interfaces: Which of 729 colors is best for signaling 'please (do not) cross'? Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics, 3721–3728, Toronto, ON. https://doi.org/10.1109/SMC42975.2020.9282998

Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2020b). Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. Proceedings of the Driving Simulation Conference Europe, Antibes, France. https://repository.tudelft.nl/islandora/object/uuid:e14ae256-318d-4889-adba-b0ba1efcca71

Bos, J. E., MacKinnon, S. N., & Patterson, A. (2005). Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. Aviation, Space, and Environmental Medicine, 76, 1111–1118.

Bradley, J. V. (1958). Complete counterbalancing of immediate sequential effects in a Latin Square design. Journal of the American Statistical Association, 53, 525–528. https://doi.org/10.1080/01621459.1958.10501456

Byers, J. C., Bittner, A. C., Jr., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), Advances in

industrial ergonomics and safety, Vol. I (pp. 481–485). London: Taylor & Francis.

Chen, C. J., Hong, J., & Wang, S. F. (2015). Automated positioning of 3D virtual scene in AR-based assembly and disassembly guiding system. The International Journal of Advanced Manufacturing Technology, 76, 753–764. https://doi.org/10.1007/s00170-014-6321-6

Chen, W., Song, J., Wang, Y., Wu, C., Ma, S., Wang, D., Yang, Z., & Li, H. (2023). Inattentional blindness to unexpected hazard in augmented reality head-up display assisted driving: The impact of the relative position between stimulus and augmented graph. Traffic Injury Prevention, 24, 344–351. https://doi.org/10.1080/15389588.2023.2186735

Cheng, Y. F., Yin, H., Yan, Y., Gugenheimer, J., & Lindlbauer, D. (2022). Towards understanding diminished reality. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA. https://doi.org/10.1145/3491102.3517452

De Oliveira Faria, N., Gabbard, J. L., & Smith, M. (2020). Place in the world or place on the screen? Investigating the effects of augmented reality head-up display user interfaces on drivers' spatial knowledge acquisition and glance behavior. Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, 762–763. Atlanta, GA. https://doi.org/10.1109/VRW50115.2020.00232

Dey, D., Matviienko, A., Berger, M., Pfleging, B., Martens, M., & Terken, J. (2021). Communicating the intention of an automated vehicle to pedestrians: The contributions of eHMI and vehicle behavior. it-Information Technology, 63, 123–141. https://doi.org/10.1515/itit-2020-0025

DINED. (2020). Anthropomorphic database. https://dined.io.tudelft.nl/en/database/tool

Dixon, B. J., Daly, M. J., Chan, H. H. L., Vescan, A., Witterick, I. J., & Irish, J. C. (2014). Inattentional blindness increased with augmented reality surgical navigation. American Journal of Rhinology & Allergy, 28, 433–437. https://doi.org/10.2500/ajra.2014.28.4067

Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Akbari, Y. (2020). Image inpainting: A review. Neural Processing Letters, 51, 2007–2028. https://doi.org/10.1007/s11063-019-10163-0

European Road Safety Observatory. (2018). Traffic safety basic facts 2018 - pedestrians. European Commision. https://ec.europa.eu/transport/road_safety/system/files/2021-07/bfs2018_pedestrians.pdf

Feiner, S., MacIntyre, B., Haupt, M., & Solomon, E. (1993). Windows on the world: 2D windows for 3D augmented reality. Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology, 145–155, Atlanta, GA. https://doi.org/10.1145/168642.168657

Fukushima, S., Hamada, T., & Hautasaari, A. (2020). Comparing world and screen coordinate systems in optical see-through head-mounted displays for text readability while walking. Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality, 649–658, Porto de Galinhas, Brazil. https://doi.org/10.1109/ISMAR50242.2020.00093

Gabbard, J. L., Smith, M., Tanous, K., Kim, H., & Jonas, B. (2019). AR DriveSim: An immersive driving simulator for augmented reality head-up display research. Frontiers in Robotics and AI, 6, 98. https://doi.org/10.3389/frobt.2019.00098

Ghasemi, Y., Singh, A., Kim, M., Johnson, A., & Jeong, H. (2021). Effects of head-locked augmented reality on user's performance and perceived workload. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65, 1094–1098. https://doi.org/10.1177/1071181321651169

Gomes, P., Olaverri-Monreal, C., & Ferreira, M. (2012). Making vehicles transparent through V2V video streaming. IEEE Transactions on Intelligent Transportation Systems, 13, 930–938. https://doi.org/10.1109/TITS.2012.2188289

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in Psychology, 52, 139–183. http://doi.org/10.1016/S0166-4115(08)62386-9

Hunter, W. W., Stutts, J. C., Pein, W. E., & Cox, C. L. (1996). Pedestrian and bicycle crash types of the early 1990's (No. FHWA-RD-95-163). McLean, VA: Federal Highway Administration.

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & De Winter, J. C. F. (2022). External Human-Machine Interfaces can be misleading: An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human Factors, 64, 1070–1085. https://doi.org/10.1177%2F0018720820970751

Kaufeld, M., Mundt, M., Forst, S., & Hecht, H. (2022). Optical see-through augmented reality can induce severe motion sickness. Displays, 74, 102283. https://doi.org/10.1016/j.displa.2022.102283

Kerr, S. J., Rice, M. D., Lum, G. T. J., & Wan, M. (2012). Evaluation of an arm-mounted augmented reality system in an outdoor environment. Proceedings of the 2012 Southeast Asian Network of Ergonomics Societies Conference, Langkawi, Malaysia. https://doi.org/10.1109/SEANES.2012.6299589

Kim, H., Gabbard, J. L., Anon, A. M., & Misu, T. (2018). Driver behavior and performance with augmented reality pedestrian collision warning: An outdoor user study. IEEE Transactions on Visualization and Computer Graphics, 24, 1515–1524. https://doi.org/10.1109/TVCG.2018.279368

Kim, H., Miranda Anon, A., Misu, T., Li, N., Tawari, A., & Fujimura, K. (2016). Look at me: Augmented reality pedestrian warning system using an in-vehicle volumetric head up display. Proceedings of the 21st International Conference on Intelligent User Interfaces, 294–298, Sonoma, CA. https://doi.org/10.1145/2856767.2856815

Kim, K., Rosenthal, M. Z., Zielinski, D., & Brady, R. (2012). Comparison of desktop, head mounted display, and six wall fully immersive systems using a stressful task. Proceedings of the 2012 IEEE Virtual Reality Workshops, 143–144, Costa Mesa, CA. https://doi.org/10.1109/
VR.2012.6180922

Klose, E. M., Mack, N. A., Hegenberg, J., & Schmidt, L. (2019). Text presentation for augmented reality applications in dual-task situations. Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces, 636–644, Osaka, Japan. https://doi.org/10.1109/VR.2019.8797992

Kooijman, L., Happee, R., & De Winter, J. C. F. (2019). How do eHMIs affect pedestrians' crossing behavior? A study using a head-mounted display combined with a motion suit. Information, 10, 386. https://doi.org/10.3390/info10120386

Lebeck, K., Ruth, K., Kohno, T., & Roesner, F. (2017). Securing augmented reality output. Proceedings of the 2017 IEEE Symposium on Security and Privacy, 320–337, San Jose, CA. https://doi.org/10.1109/SP.2017.13

Lee, H., & Woo, W. (2023). Exploring the effects of augmented reality notification type and placement in AR HMD while walking. Proceedings of the 2023 IEEE Conference Virtual Reality and 3D User Interfaces, 519–529, Shanghai, China. https://doi.org/10.1109/VR55154.2023.00067

Liao, M., Lu, F., Zhou, D., Zhang, S., Li, W., & Yang, R. (2020). DVI: Depth guided video inpainting for autonomous driving. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), Computer Vision – ECCV 2020. Cham: Springer. https://doi.org/10.1007/978-3-030-58589-1_1

Lindemann, P., Eisl, D., & Rigoll, G. (2019). Acceptance and user experience of driving with a see-through cockpit in a narrow-space overtaking scenario. Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces, 1040–1041, Osaka, Japan. https://doi.org/10.1109/VR.2019.8798069

Liu, B., Ding, L., & Meng, L. (2021). Spatial knowledge acquisition with virtual semantic landmarks in mixed reality-based indoor navigation. Cartography and Geographic Information Science, 48, 305–319. https://doi.org/10.1080/15230406.2021.1908171

Liu, Z., Pu, L., Meng, Z., Yang, X., Zhu, K., & Zhang, L. (2015). POFS: A novel pedestrian-oriented forewarning system for vulnerable pedestrian safety. Proceedings of the 2015 International Conference on Connected Vehicles and Expo, 100–105. Shenzhen, China. https://doi.org/10.1109/ICCVE.2015.63

Lu, W., Duh, B.-L. H., & Feiner, S. (2012). Subtle cueing for visual search in augmented reality. Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality, 161–166. Atlanta, GA. https://doi.org/10.1109/ISMAR.2012.6402553

MagicLeap. (2020). 5.1 Head-locked content - Unity. https://ml1-developer.magicleap.com/en-us/learn/guides/head-locked-content-tutorial-unity

Mann, S., & Fung, J. (2002). EyeTap devices for augmented, deliberately diminished, or otherwise altered visual perception of rigid planar patches of real-world scenes. Presence: Teleoperators and Virtual Environments, 11, 158–175. https://doi.org/10.1162/105474602147060 3

Markov-Vetter, D., Luboschik, M., Islam, A. T., Gauger, P., & Staadt, O. (2020). The effect of spatial reference on visual attention and workload during viewpoint guidance in augmented reality. Proceedings of the 2020 ACM Symposium on Spatial User Interaction, 10, Virtual Event. https://doi.org/10.1145/3385959.3418449

Meerits, S., & Saito, H. (2015). Real-time diminished reality for dynamic scenes. Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality Workshop, 53–59. Fukuoka, Japan. https://doi.org/10.1109/ISMARW.2015.19

Mok, C. S., Bazilinskyy, P., & De Winter, J. C. F. (2022). Stopping by looking: A driver-pedestrian interaction study in a coupled simulator using head-mounted displays with eye-tracking. Applied Ergonomics, 105, 103825. https://doi.org/10.1016/j.apergo.2022.103825

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4, 61–64. https://doi.org/10.20982/tqmp.01.1.p042

Mori, S., Ikeda, S., & Saito, H. (2017). A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. IPSJ Transactions on Computer Vision and Applications, 9, 1–14. https://doi.org/10.20982/tqmp.04.2.p061

Onkhar, V., Bazilinksyy, P., Dodou, D., & De Winter, J. C. F. (2022). The effect of drivers' eye contact on pedestrians' perceived safety. Transportation Research Part F: Traffic Psychology and Behaviour, 84, 194–210. https://doi.org/10.1016/j.trf.2021.10.017

Orlosky, J., Liu, C., Kalkofen, D., & Kiyokawa, K. (2019). Visualization-guided attention direction in dynamic control tasks. Proceedings of the 2019 IEEE International

Symposium on Mixed and Augmented Reality Adjunct, 372–373. Beijing, China. https://doi.org/10.1109/ISMAR-Adjunct.2019.000-9

Overmeyer, L., Jütte, L., & Poschke, A. (2023). A real-time augmented reality system to see through forklift components. CIRP Annals. https://doi.org/10.1016/j.cirp.2023.03.010

Pala, P., Cavallo, V., Dang, N. T., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2021). Is the street-crossing behavior with a head-mounted display different from that behavior in a CAVE? A study among young adults and children. Transportation Research Part F: Traffic Psychology and Behaviour, 82, 15–31. https://doi.org/10.1016/j.trf.2021.07.016

Palffy, A., Kooij, J. F. P., & Gavrila, D. M. (2023). Detecting darting out pedestrians with occlusion aware sensor fusion of radar and stereo camera. IEEE Transactions on Intelligent Vehicles, 8, 1459–1472. https://doi.org/10.1109/TIV.2022.3220435

Pijnenburg, J. (2017). Naturalism: Effects of an intuitive augmented reality interface property in the display of automated driving status (MSc thesis). Delft University of Technology.

Rameau, F., Ha, H., Joo, K., Choi, J., Park, K., & Kweon, I. S. (2016). A real-time augmented reality system to see-through cars. IEEE Transactions on Visualization and Computer Graphics, 22, 2395–2404. https://doi.org/10.1109/TVCG.2016.2593768

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. Computers in Human Behavior, 133, 107289. https://doi.org/10.1016/j.chb.2022.107289

Renner, P., & Pfeiffer, T. (2017). Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. Proceedings of the 2017 IEEE Symposium on 3D User Interfaces, 186–194. Los Angeles, CA. https://doi.org/10.1109/3DUI.2017.7893338

Robertson, C. M., MacIntyre, B., & Walker, B. N. (2008). An evaluation of graphical context when the graphics are outside of the task area. Proceedings of the 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 73–76. Cambridge, UK. https://doi.org/10.1109/ISMAR.2008.4637328

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrila, D. M., & Arras, K. O. (2020). Human motion trajectory prediction: A survey. The International Journal of Robotics Research, 39, 895–935. https://doi.org/10.1177/0278364920917446 Samsung. (2015). The Safety Truck could revolutionize road safety. https://news.samsung.com/global/the-safety-truck-could-revolutionize-road-safety

Schankin, A., Reichert, D., Berning, M., & Beigl, M. (2017). The impact of the frame of reference on attention shifts between augmented reality and real-world environment.

Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, 25–30. Nantes, France. https://doi.org/10.1109/ISMAR-Adjunct.2017.24

Schinke, T., Henze, N., & Boll, S. (2010). Visualization of off-screen objects in mobile augmented reality. Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, 313–316, Lisbon, Portugal. https://doi.org/10.1145/1851600.1851655

Schmitz, A., MacQuarrie, A., Julier, S., Binetti, N., & Steed, A. (2020). Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, 63–72. Atlanta, GA. https://doi.org/10.1109/VR46266.2020.00024

Smith, M., Gabbard, J. L., Burnett, G., Hare, C., Singh, H., & Skrypchuk, L. (2021). Determining the impact of augmented reality graphic spatial location and motion on driver behaviors. Applied Ergonomics, 96, 103510. https://doi.org/10.1016/j.apergo.2021.103510

SWOV. (2020). Pedestrians [Fact sheet]. https://swov.nl/en/fact-sheet/pedestrians

Tabone, W., Happee, R., Yang, Y., Sadraei, E., García, J., Lee, Y. M., Merat, N., & De Winter, J. (2023). Immersive insights: Evaluating augmented reality interfaces for pedestrians in a CAVE-based experiment [preprint]. ResearchGate. https://www.researchgate.net/publication/370160064_Immersive_Insights_Evaluating_Augmented_Reality_Interfaces_for_Pedestrians_in_a_CAVE-Based_Experiment

Tabone, W., Lee, Y. M., Merat, N., Happee, R., & De Winter, J. (2021). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 209–218, Leeds, United Kingdom.. https://doi.org/10.1145/3409118.3475149

Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies, 5, 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Waldin, N., Waldner, M., & Viola, I. (2017). Flicker observer effect: Guiding attention through high frequency flicker in images. Computer Graphics Forum, 36, 467–476. https://doi.org/10.1111/cgf.13141

Walter, M., Wendisch, T., & Bengler, K. (2019). In the right place at the right time? A view at latency and its implications for automotive augmented reality head-up displays. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), Proceedings of the 20th Congress of the International Ergonomics Association (pp. 353–358). Cham: Springer. https://doi.org/10.1007/978-3-319-96074-6_38

Wickens, C. (2021). Attention: Theory, principles, models and applications. International Journal of Human–Computer Interaction, 37, 403–417. https://doi.org/10.1080/10447318.2021.1874741

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37, 473–494. https://doi.org/10.1518/001872095779049408

Wiesner, C. A. (2019). Increasing the maturity of the Augmented Reality Head-Up-Display (Doctoral dissertation). Technische Universität München.

Wilmott, J. P., Erkelens, I. M., Murdison, T. S., & Rio, K. W. (2022). Perceptibility of jitter in augmented reality head-mounted displays. Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality, 470–478, Singapore. https://doi.org/10.1109/ISMAR55827.2022.00063

Won, M., Shrestha, A., Park, K.-J., & Eun, Y. (2020). SaferCross: Enhancing pedestrian safety using embedded sensors of smartphone. IEEE Access, 8, 49657–49670. https://doi.org/10.1109/ACCESS.2020.2980085

World Health Organization. (2023). Pedestrian safety: a road safety manual for decision-makers and practitioners, 2nd ed. https://www.who.int/publications/i/item/9789240072497

Yasuda, H., & Ohama, Y. (2012). Toward a practical wall see-through system for drivers: How simple can it be? Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality, 333–334. Atlanta, GA. https://doi.org/10.1109/ISMAR.2012.6402600

Yue, L., Abdel-Aty, M., Wu, Y., Zheng, O., & Yuan, J. (2020). In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention. Journal of Safety Research, 73, 119–132. https://doi.org/10.1016/j.jsr.2020.02.

Zhang, B., Wilschut, E. S., Willemsen, D. M. C., Alkim, T., & Martens, M. H. (2018). The effect of see-through truck on driver monitoring patterns and responses to critical events in truck platooning. In N. Stanton (Ed.), Advances in Human Aspects of Transportation. AHFE 2017 (pp. 842–852). Cham: Springer. https://doi.org/10.1007/978-3-319-60441-1_81

Zhao, Y., Stefanucci, J., Creem-Regehr, S., & Bodenheimer, B. (2023). Evaluating augmented reality landmark cues and frame of reference displays with virtual reality. IEEE Transactions on Visualization and Computer Graphics, 29, 2710–2720. https://doi.org/10.1109/TVCG.2023.3247078

SUPPLEMENTARY MATERIAL

The vehicle commenced from a stationary position, accelerated to 15 km/h, and executing a 90° left turn to approach the pedestrian. In the *TransparentVehicle* condition, the vehicle was visible from the start of the trial. In contrast, for the *Baseline* condition, the vehicle was not fully visible at the start of the trial. For the remaining conditions, the vehicle was visible but may have appeared small, as its depiction was confined to the virtual displays and the vehicle was still far away. These differences in the designs may explain variations in key-press percentages early in the trial (i.e., before 5 s), with the *Baseline* and *TransparentVehicle* conditions eliciting the fewest and most key presses, respectively. It was also the case that for 38 of the 84 yielding trials in the *Baseline* condition, due to a programming error, the vehicle accelerated faster at the beginning of the trial (5 m/s^2) than in the other trials (1 m/s^2), as a result of which it reached its target speed more quickly. This offset has been corrected in the figure below, and these trials have been excluded in the visualisation of the yielding trials between 0 and 7.5 s. This anomaly does not apply to the non-yielding trials.

Figures 8.11 and 8.12 show that there were no discernible differences between conditions in pressing the key after the vehicle had passed the occluding vehicle, i.e., from 15 s onwards. A plausible explanation is that the vehicle was consistently visible and served as the dominant cue for determining whether or not to press the key. Also, the video displays were designed such that the vehicle was not shown after it had passed by (i.e., the displays simulated a non-rotating camera mounted on the parked pickup truck and did not provide a rear-facing view).

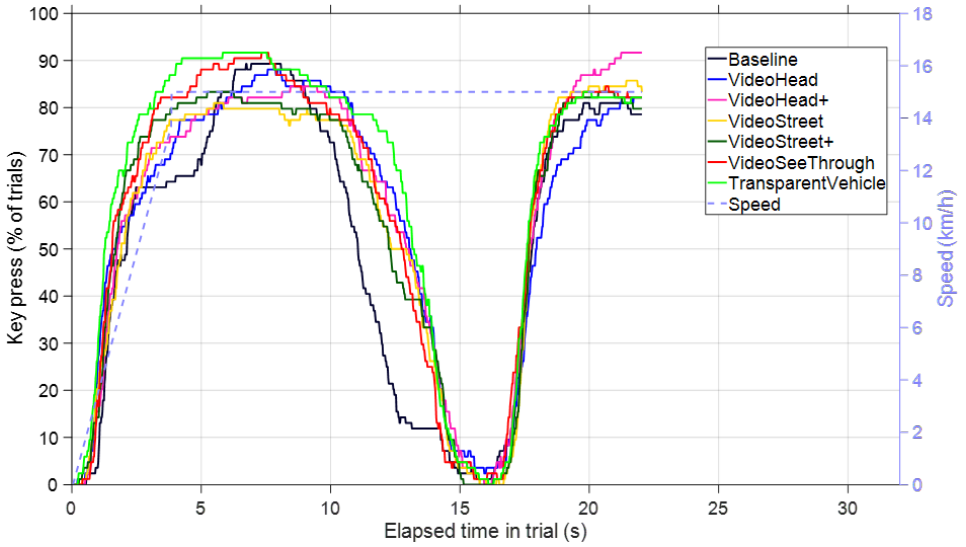


Figure 8.11: Key-press percentages for the seven conditions in the non-yielding scenario, with vehicle speed.

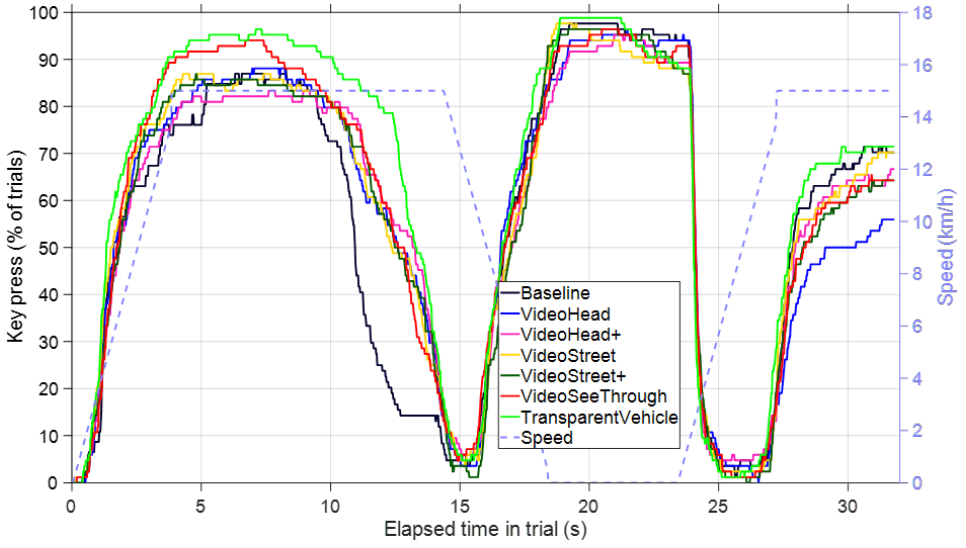


Figure 8.12: Key-press percentages for the seven conditions in the yielding scenario, with vehicle speed.

IV

Part IV: Conclusion

9

Conclusions and Recommendations

This thesis presents the work conducted to achieve the main aim of designing and evaluating augmented reality interfaces which promote safe, transparent, and trustworthy interactions between pedestrians and automated vehicles. In this chapter, a summary of the key findings from each chapter is presented, followed by a number of reflections, design recommendations based on the findings, recommendations for future work, and ultimately a presentation of a conclusive vision of the future.

9.1 Summary of Key Findings

Chapter 2: Vulnerable road users and the coming wave of automated vehicles: Expert perspectives

Chapter 2 presents the results of an interview study with 16 researchers discussing automated vehicles (AVs) and their interaction with vulnerable road users (VRUs) in future cities. Topics like smart infrastructure, external human-machine interfaces (eHMIs), and the potential of augmented reality (AR) were explored. Experts largely agreed that full AV autonomy is not imminent, with shared control being the short-term variable option for the time being. The viability of smart infrastructure was questioned due to the costs associated with it, but its potential was recognised. Some experts even expressed a need for AV-VRU segregation but were also concerned about cost and maintenance.

Regarding eHMIs, researchers recommended avoiding text-based and instructive designs, and also against the use of surface anthropomorphism, such as having virtual eyes or smiling faces on the car. Instead favouring allocentric designs which complement implicit vehicle communication like speed and trajectory.

The potential of AR in aiding VRUs was acknowledged, yet experts saw its current application limited to scientific experiments due to privacy concerns, invasiveness, user-friendliness, technological feasibility (eg., brightness, image stability), and inclusiveness (access to the technology). Despite their concerns, the researchers presented several ideas of how they could envision AR being implemented for AV-VRU interactions. Possibilities included the use of holographic traffic lights and signs, the removal of irrelevant information in the world (so-called 'diminished reality', a concept explored in Chapter 8), an indicator nudging the user's attention towards an AV, the projection of safety zones or coloured road surfaces, or a fence or barrier indicating that one should not cross.

In addition, the experts recognised the value of AR in overcoming eHMI limitations but similar to eHMIs, viewed AR as a secondary cue to implicit communication in the real layer. The reason being that not everyone can be expected to wear an AR device. Furthermore, the majority of experts saw AR as a suitable research tool for simulating and testing interfaces, and for training in a realistic context such as by utilising augmented cars.

Chapter 3: Towards future pedestrian-vehicle interactions: Introducing

In Chapter 3, a process for designing novel AR interfaces for pedestrian-vehicle interactions was presented. Specifically, a 'genius-based' (i.e., experience-based) design method was used for the overall design process, split into a four-phase process. In the first phase, a number of ideas suggested by the researchers in Chapter 2 were identified (7 quotes in total) and extracted. In the second phase, relevant literature was collected to serve as a theoretical

foundation for the design phase. For each identified quote in the first phase, supported theories and related concepts were extracted and collated from seminal works in the field of Human Factors and Ergonomics.

The third phase consisted of theoretically informed brainstorming sessions, where the designers performed an online collaborative affinity mapping exercise for each idea quote. A set of design principles for human-computer interaction, and a LiDAR scan of the road environment which would be used to demonstrate the finalised interfaces were provided on the online collaborative board for reference during the process. In total, nine different concepts were designed, each with two states, namely 'vehicle is not yielding, and 'vehicle is yielding, which would be depicted in red and green respectively since these colours have a high intuitiveness rating for signalling 'please (do not) cross'. The interfaces were separated into three categories, namely: interfaces which are mapped to the road, mapped to the vehicle, and heads-up displays (HUDs).

An AR heuristic evaluation was conducted on the proposed interfaces, which were subsequently implemented in AR during the fourth phase using the Unity engine with the Mixed and Augmented Reality Studio (MARS) extension package and demoed on an iPad Pro 2020 equipped with a LiDAR sensor. The LiDAR allowed for interfaces to be dynamically mapped to the real environment's topology in real time.

Following the design process, the next step was to conduct a user evaluation of the interfaces in order to gauge which design elements, and anchoring positions are most suited to eliciting transparent interactions between pedestrians and automated vehicles.

CHAPTER 4: AUGMENTED REALITY INTERFACES FOR PEDESTRIAN-VEHICLE INTERACTIONS: AN ONLINE STUDY

Chapter 4 presented the results of an online questionnaire answered by 992 respondents across 5 countries (Germany, The Netherlands, Norway, Sweden, and The United Kingdom) which evaluated the nine AR interfaces that were developed as part of the experiment presented in Chapter 3. The aim of the questionnaire was to get a first impression of which interfaces respondents found intuitive (how clear its conveyed message is), and convincing (in the message to either cross or not cross the road). Moreover, respondents rated the trigger speed, size, understandability, and aesthetic quality of the interfaces on a descriptor scale, and also a number of facets related to usefulness and satisfaction on a 9-item acceptance scale. A further free-text area was provided so that participants could elaborate further. The nine interfaces were implemented in a VR environment and videos of each interface operating in both states were created. The interfaces were presented in random order, and each questionnaire page contained both a video of the non-yielding and yielding states.

Results from the experiment indicated that each measured rating correlated highly (r = 0.998 just between intuitiveness and convincingness). A composite score of the correlated items was calculated, with interfaces that were head-locked or used traditional traffic elements scoring higher than their world-locked, and more 'creative design' counterparts. These results highlighted the limitations of the 'genius' design method and the importance of involving the user earlier in the process.

Responses on the general use of interfaces favoured those which were mapped to the street rather than vehicles. Most participants also expressed a desire for customizable AR interfaces, which they viewed as beneficial for future traffic communication.

The online questionnaire study provided preliminary insights into the most effective AR interfaces, their preferred placement and anchoring methods in the environment, and which design elements are more suitable for pedestrian-vehicle interactions. However, there were limitations related to the ecological validity dimension of the study. In order to better understand the behaviour of potential users of the system, the ecological validity was increased, and elements of the user evaluation were replicated in a CAVE pedestrian-simulator experiment. The findings of the latter experiment, which also examined the participants' attention distribution were presented in Chapter 6.

Chapter 5: Using ChatGPT for Human-Computer Interaction Research: A Primer

Chapter 5 presents a validation study on the use of Large Language Models (LLMs) for evaluating qualitative data since the method was used in Chapter 6. The study investigated the potential of ChatGPT as a text analysis tool in Human-Computer Interaction research. It was applied across three datasets—questionnaire responses, interviews, and think-aloud transcripts. ChatGPT was respectively utilised to generate sentiment scores, and meta-summaries, and highlight the differences in the emotional and analytical responses of participant summaries. The results showed a strong correlation between ChatGPT's sentiment scores, human ratings, and traditional sentiment analysis. ChatGPT was also capable of producing meaningful meta-summaries from interviews. Furthermore, ChatGPT was shown to be able to highlight subtle differences between the think-aloud transcripts of two control groups in an experiment. The results were consistent with a keyword analysis and also highlighted the capabilities of such tools in contrast to a human experimenter operating alone.

Despite the promising results, the chapter also highlighted a number of challenges which need to be taken into consideration. While ChatGPT has shown promise in various applications, it still lacks deep reasoning abilities and introspective capabilities, meaning that it cannot self-correct or perform sanity checks on its outputs. The chapter closes with a number of mitigating solutions, with the key message being that with appropriate precaution, ChatGPT can serve as a useful analysis tool in human factors and HCI research.

Chapter 6: Immersive Insights - Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment

In Chapter 6, results from a pedestrian simulator experiment were presented. The interfaces were evaluated by 30 participants in a Cave automatic virtual environment (CAVE). One of the aims of the experiment was to examine whether the prior subjective measures from the online evaluation of the AR interfaces presented in Chapter 4 could be replicated in an immersive virtual environment. A further aim was to assess whether the AR design effectiveness depends on pedestrian attention allocation. In contrast to the online questionnaire, which only measured subjective measures from respondents, the CAVE experiment also measured objective behaviour. The experiment consisted of 10 blocks (9 interfaces, and 1 baseline condition with no interface), with each block containing several trails which either displayed the non-yielding or yielding states of the interface as a vehicle approached from

the right onto a one-way road environment.

To emulate visual distraction, and accordingly test how distraction interacts with the AR interface's effectiveness, participants had to look into an attention-attractor circle that disappeared 1 second after the AR interface appeared. Participants were invited to cross the environment if they felt it was safe to do so. Following each trial, the interfaces were rated on their intuitiveness, while additional open-ended qualitative data was captured through voice recording after each block. These recordings were then transcribed using an artificial intelligence tool, and subsequently analysed by ChatGPT in order to generate key strengths and weaknesses of each interface as described by all 30 participants.

The results from the experiment indicated a strong correlation between the online study ratings, and the CAVE study ($r \approx 0.90$, for the 9 interfaces). Moreover, it was noted that similar to the online questionnaire study, the head-locked interfaces and familiar designs yielded higher intuitiveness scores, and in the case of the CAVE study, these interfaces elicited quicker crossing initiation times (CITs) than vehicle-locked interfaces. It was also noted that vehicle-locked interfaces were less effective when the attention-attractor was on the environment's opposite side, while head-locked interfaces were unaffected. The results highlight the importance of interface placement in relation to the user's gaze direction and implications for placing eHMIs on the vehicle.

CHAPTER 7: AUGMENTED REALITY FOR SUPPORTING THE INTERACTION BETWEEN PEDESTRIANS AND AUTOMATED VEHICLES: AN EXPERIMENTAL OUTDOOR STUDY

Following a number of user evaluations using methods with lower ecological validity, Chapter 7 presented a study that utilised an AR HMD in an outdoor environment. The study focused on evaluating the effectiveness of four AR interfaces that were chosen amongst the nine that were developed and evaluated in previous chapters. The user evaluation was conducted with 28 participants and included a baseline condition without an AR interface. In contrast with user evaluations presented in Chapters 4 and 6, which used AR within a VR setting, this study utilised AR in an outdoor environment. During the experiment, participants were required to indicate when they felt safe to cross a street by pressing a button, in scenarios involving an augmented AV. This method allowed for consistent vehicle movement and AR activation timings across all trials.

Amongst the key findings was that the ratings for the AR interfaces differed from those in previous studies. Noteworthily, the *Planes on vehicle* interface received a high rating, around 6 out of 7, which differs from its lower ratings in earlier research. In contrast, the *Fixed pedestrian lights* interface scored lower, which again differed from previous studies. However, despite the differences in ratings, a number of observations about the AR interfaces corresponded with prior research. For example, the study reaffirmed issues related to the *Virtual fence* interface, such as the confusion caused by a red-coloured crosswalk, and the risk of occluding other road users with virtual walls, or HUDs. Furthermore, issues related to egocentric vs. exocentric confusion when a vehicle emits a red signal were also observed (similar to comments made in Chapters 4 and 6), as were issues with world-locked interfaces such as the *Fixed pedestrian lights*, which caused divided attention. The latter issue was also observed in findings from Chapter 6, where the moment of first glance for the *Fixed pedestrian lights* was the highest when the pedestrian was initially looking to the left or right.

In this chapter, the issues related to divided attention were further reinforced by a number of practical limitations, such as the elements in the HUD interface being too small.

The study faced some experimental challenges such as HUDs being slightly off-centred, but also other challenges posed by the cluttered real-world setting, requiring further divided attention demands from the participants for interfaces like the *Fixed pedestrian lights*. Participants struggled with tasks like localising attention circles and remembering to press response buttons, suggesting higher cognitive load in real-world AR compared to virtual settings. Other highlighted limitations included the necessity for different tracking technologies for real-world applications, limited field of view in current AR devices, and the limits of using simulated AVs. The single AV setup was also a constraint, highlighting the need for complex scenarios in future research.

Despite these challenges, the study shed light on AR interfaces' effectiveness in real-world contexts for pedestrian safety with AVs. Future research should aim for AR interfaces requiring less direct focus and investigate AR in realistic, multi-vehicle settings.

Chapter 8: Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios

Lastly, Chapter 8 presented a study which explored the effectiveness of AR solutions in preventing collisions between pedestrians and cars due to poor visibility, specifically when a parked vehicle obstructs the view. In contrast to previous chapters, in this chapter, the concept of diminished reality (DR), whereby information is removed from the world, was also used.

A VR experiment involving 28 participants using head-mounted displays was conducted to analyse 6 different AR and DR designs. The AR solutions involved displaying live video of the obscured area directly on the HMD (*VideoHead*) or showing the video feed on the opposite side of the street (*VideoStreet*). On the other hand, the two conformal DR designs included making the parked vehicle semi-translucent (*TransparentVehicle*) or projecting the view from the front of the parked vehicle on the rear windshield (*VideoSeeThrough*). The study found that interface designs with high local embeddedness, which means that they can be seamlessly blended with the real environment, were preferred by the participants. In fact DR solutions which were highly conformal to the environment were rated highly as these respected the proximity compatibility interface-design principle, placing task elements which need to be processed concurrently together (similar to findings from Chapters 6 and 7). It was however noted that head-locked solutions, such as *VideoHead*, caused discomfort and dissatisfaction. Interestingly, adding guiding arrows to the *VideoHead* and *VideoStreet* conditions did not show significant benefits.

The study suggests that while AR and DR designs with high local embeddedness are effective for addressing visibility issues, simpler head-up-display solutions (such as just presenting light in the peripheral vision) should not be disregarded, as they are more practical for tasks requiring clear warnings or instructions.

9.2 REFLECTIONS

9.2.1 WHAT IS THE POTENTIAL OF AUGMENTED REALITY IN FACILITATING PEDESTRIAN-VEHICLE INTERACTIONS?

Augmented Reality has the potential to change the way we interact with the surrounding environment and also with one another. As noted in earlier chapters, AR has already been applied to various domains which recognise its potential and benefit. One of the questions posed in this thesis was whether AR would also offer any solutions for pedestrian-vehicle interactions.

Chapter 2 highlighted expert opinions on the eHMIs, some of their limitations, and the use of AR in the domain of pedestrian-vehicle interactions. Limitations pertaining to eHMIs which were mentioned included the use of text and the time it takes for pedestrians to read them, while detached eHMIs were also discouraged because they may be placed far away from the vehicle, requiring divided attention by the pedestrian. The research presented in the thesis now confirms that AR does indeed offer a solution to these issues. AR makes it possible to present text directly in the field-of-view (FOV) of the pedestrian, and also in their preferred language (as demonstrated by the willingness of participants to customise the interfaces in Chapters 4, and 6). Moreover, the best most intuitive interfaces had a high level of stimulus-response compatibility, considering that the interface elements were close to the task at hand, either because they followed the participant's gaze (HUDs), had a high level of embeddedness (eg. *Augmented zebra crossing*), or because they were conspicuous and made use of the person's peripheral vision (eg. *Virtual fence*). These findings indicate that AR interfaces which are detached from the main interaction partner (in this case the AV) are promising, as they resolve the spatial disconnect that may occur with static interfaces.

A further advantage over eHMIs is that since AR is predominantly software-based, standardisation across different vehicle manufacturers is more achievable. Moreover, one of the standout capabilities of AR, explored in Chapter 8, is its ability to filter out distractions and clarify occluded scenarios, something that no current eHMI can achieve. The perceived safety among participants increased when visual obstructions were eliminated, which is a demonstration of the transformative impact AR could have on the safety and transparency of pedestrian-vehicle interactions.

Despite these advantages, privacy and invasiveness are legitimate concerns in the face of mass proliferation of AR wearable devices, as evidenced by the failure of projects like Google Glass. Companies are now exploring solutions, such as having spatial features computed locally on an on-board chip, and having only simple non-geotagged features (such as lines) transmitted to servers. In this way, on-board cameras would not capture and send sensitive information to offsite data servers. Therefore, it is my belief that mass adoption of AR wearables will not be hampered by privacy concerns once companies clearly demonstrate how their customer privacy is being actively safeguarded in their AR product offerings.

A further issue raised in Chapter 2 was the limited access to AR technology due to economic disparity. As the research in this thesis has shown, the AR interfaces serve as a form of redundancy gain and confirmation cue to the implicit communication of the vehicle. Therefore, someone without an AR wearable can still make use of the vehicle's implicit cues, or make use of traditional infrastructure instead of crossing at an unsignalised area. Furthermore, it is unrealistic to expect universal ownership of AR wearables, similar to the

case with smartphones or smartwatches. However, continuous technological improvements will mitigate these technological barriers, ensuring that these solutions are more widely accessible over time.

Initially, the effectiveness of AR for short-term tasks was questioned in the expert perspectives, with the assumption that it might not be feasible or useful for such purposes. However, the behavioural findings from Chapters 6 and 7 demonstrated that AR could indeed facilitate quicker pedestrian crossing initiation time (CIT) and achieve higher crossing decisions when compared to baseline conditions without explicit AR information. Furthermore, the potential of AR to reduce VRU workload and confusion was noted by experts (Chapter 2), which led us to investigate the intuitiveness, among other measures, of these AR interfaces in our experiments (Chapters 4, 6, 7, 8). Intuitiveness was generally higher in situations where an interface was present, as opposed to when it was absent. These results, coupled with a reported diminished sense of safety when interfaces were not present (Chapter 8), and the willingness of participants to utilise AR in traffic (Chapter 4) affirms the utility of AR in future traffic communication and its potential in building trust. The universality of the AR interfaces was further supported by their general robustness across different cultural contexts, as evidenced by the results achieved in Chapters 4 and 6, with the latter featuring a more diverse participant pool. These results suggest that the technology has a broad application potential.

In summary, the results achieved in this thesis, affirm and demonstrate the potential of AR as a technology which can be integrated as part of future V2X solutions. The technology's provided advantage over eHMIs was demonstrated, as were the behavioural effects elicited when AR interfaces were presented to participants. A number of key takeaways were extracted from these investigations, and are presented in subsequent sections. In general, the progression from theoretical concerns to empirical validations with increasing levels of ecological validity marks a step towards realising an environment where vehicles and humans share and navigate spaces safely and intuitively.

9.2.2 How Should AR be Designed, Integrated, and Anchored onto the Real Environment ?

The experiments conducted throughout this thesis have provided a number of insights which can serve as guidelines on how information pertaining to pedestrian-automated vehicle interactions should be presented in AR. Firstly, the studies highlighted the critical role of redundancy gain in AR interface design. In fact, findings in Chapters 3, 4, and 6 underlined the importance of not relying on a single cue, as interfaces lacking redundancy, such as the *Field of safe travel*, and the *Phantom car*, which both displayed only a coloured element, exhibited poor performance. It is therefore recommended that *multiple cues are utilised in order to ensure intuitiveness*. Moreover, the importance of *applying top-down processing principles* was also highlighted as AR interfaces which incorporated traditional traffic design elements were found to be more intuitive.

Head-locked interfaces, such as the HUDs scored highly in preference throughout the experiments. Furthermore, behavioural findings in Chapter 6 show that *HUDs are less affected by attention-attractors*, since the interface is in the pedestrian's FOV independent of where they are looking at. Hence, the *HUDs are suitable for complex environments such as interactions in traffic*, since in a real traffic situation, the pedestrian is not expected to look

at the incoming vehicle's direction right away. Moreover, HUDs are suitable interfaces for interaction scenarios which involve multiple vehicles communicating to the same pedestrian, since HUDs do not require constant monitoring of multiple elements. It is recommended that when designing HUD interfaces, the elements *should be placed in areas that are not used for constant monitoring of the situation*, such as the top area of the user's FOV. Moreover, in this context, the importance of offering *customisation and language adaptability* to the user should also not be overlooked. Furthermore, for HUDs and any head-locked interface, it is important to consider issues related to *vergence-accommodation conflicts*, and ascertain that the design elements can be clearly seen at all times without extra effort by the user (as highlighted by the decreased performance of HUD interfaces in Chapter 8 when compared to Chapter 7, when the distance of the interface from the participant's foveal vision was increased).

Another important factor to take into account would be to *present task-relevant elements together in order to facilitate user interaction*. In the thesis, we have seen how the proximity compatibility principle played a vital role in how experiment participants rated the interfaces. Placing interface elements close to the automated vehicle would ensure that there would be a reduction in the number of head turns in cases where the pedestrian is distracted and looking elsewhere. Furthermore, such conformal AR interfaces (mixed reality in the xReality framework), where interface elements are blended into the real environment, exhibit higher local presence. This was evidenced by the high performance of the two diminished reality solutions (*VideoSeeThrough* and *TransparentVehicle*) in Chapter 8.

Saliency (conspicuousness), and the user's *peripheral vision* should also be taken into account when designing and anchoring AR interfaces. When these principles were adhered to, the participant ratings for the interfaces also increased. For example, the *Virtual fence* achieved higher ratings in the CAVE (Chapter 6) when compared to the online questionnaire (Chapter 4) since in the latter experiment, the peripheral vision could not be taken advantage of with videos to facilitate information access effort.

On the other hand, there was a decrease in preference for the *Fixed pedestrian light* in the AR HMD study (Chapter 7) when contrasted to previous chapters, due to its smaller size, further distance, and lower saliency. Adherence to these principles would also serve the purpose of enabling a quicker response time for the user through rapid direction of foveal vision. If we take the two previous suggestions (on presenting task relevant elements together, and information access effort) into account, it is recommended to *avoid vehicle-locked interfaces* since these are harder to detect with peripheral vision, and mainly prove to be advantageous in CITs when the pedestrian is already looking towards the vehicle at the onset. The interface may also obscure the 'drivenger' inside the vehicle (as was the case with the *Planes on vehicles*, and *Conspicuous looming planes*), making the situation even more ambiguous.

If we take the two previous suggestions (on presenting task relevant elements together, and information access effort) into account, it is recommended to *avoid vehicle-locked interfaces* since these are harder to detect with peripheral vision, and mainly prove to be advantageous in CITs when the pedestrian is already looking towards the vehicle at onset. The interface may also obscure the 'drivenger' inside the vehicle (as was the case with the *Planes on vehicles*, and *Conspicuous looming planes*), making the situation even more ambiguous.

	ROAD-MAPPED World-locked	VEHICLE-MAPPED World-locked	HEAD-UP DISPLAY Head-locked
			
Embeddedness (local presence) ⑥ ⑦	High	High	Low
Proximity-compatibility ③ ⑥ ⑦	Medium	High	Medium
Required head-movement ⑥ ⑦ ⑧	Medium	High	Low
Information access effort ⑥	Medium	Low	High
Saliency ④ ⑥ ⑦ ⑧	Medium	Medium	Low
Comfort ⑥ ⑦ ⑧	Medium	High	Medium*
Sensor requirements ③	Medium	High	Low

Figure 9.1: In this figure, the differently mapped interfaces' pros and cons in terms of capabilities, and sensor requirements are presented. Encircled numbers refer to the chapters which discuss the concepts listed on the left. *Comfort is medium for HUDs if necessary precautions (ie. proper anchoring, sizing, and blending) are taken into account.

The above considerations are supported by the behavioural data collected in Chapters 6 and 7, which indicate faster CITs and a higher crossing willingness for conspicuous interfaces (eg. *Virtual fence*), and head-locked interfaces.

Finally, it is worth pointing out that despite the trust shown in the interfaces in contrast to having none, the risks of over-reliance and over-trust of such interfaces, as reported in the CAVE results from Chapter 6, raise concerns about users placing excessive trust in such technological guidance. It is therefore recommended to design interfaces which *focus more on communicating dangerous states* (eg. non-yielding states) than the equivalent safe states; in fact, in the online questionnaire, non-yielding states were noted to be clearer at times in contrast to their yielding equivalents (Chapter 4).

Figure 9.1 presents a key summary of how each interface map category adheres to the key properties discussed above.

Key findings (as presented in Figure 9.1):

- Road-mapped and vehicle-mapped interfaces have high embeddedness since elements can be seamlessly blended onto real-world objects. On the other hand, HUDs have lower embeddedness since they overlay the environment.

- Proximity-compatibility is highest for the vehicle-mapped interfaces, since these map exactly onto the interacting agent, in contrast to the other mapped categories (Chapter 3), hence also making them the most comfortable in terms of. User head movement would be higher when interfaces are mapped to the vehicle since the pedestrian would have to turn their heads towards oncoming traffic (Chapter 6, and 7), but road-mapped interfaces are less susceptible to head-turns since these are mapped in the centre and would typically require half a head turn; while there is no head-turning required for head-locked interfaces since these are perpetual in the user's FOV.

- Consequently information access effort (aided with peripheral vision) is higher for head-locked interfaces, than road-mapped or vehicle-mapped interfaces (Chapter 6) where peripheral vision is concerned, however saliency and clarity risk being lower in head-locked interfaces if elements are not properly sized and placed in the environment, causing vergence-accommodation conflicts (Chapters 7 and 8). Therefore, comfort (in terms of vergence conflict and visual vestibular conflict) is lower for HUDs, in contrast to other mapped categories.

- Vehicle-mapped interfaces have higher sensor requirements than road-mapped interfaces since in the former case persistent AR would have to be used, thereby requiring tracking of the vehicle's geographical position, and orientation (Chapter 3). Furthermore, in instances where multiple vehicles are presented, care would need to be taken to ensure that the sensors keep track of each vehicle's state. On the other hand, road-mapped interfaces require knowledge of the terrain surface, and vehicle location to locate the appropriate location to project the interface. Since there is no mapping on a moving agent, the requirements are lower, as attested by the more straightforward implementations of road-mapped interfaces in Chapters 3 and 7, in contrast to vehicle-mapped ones. Head-locked interfaces have even lower sensor requirements in addition to the sensors needed for proper AR projection on the HMD.

9.2.3 REFLECTIONS ON EVALUATION

In the thesis, we have also shown that researchers may opt for more accessible and less resource-intensive methods without significantly compromising the quality of their findings. In contrasting the subjective measures rating the interfaces that were captured in the CAVE experiment with the ones from the online questionnaire study, we revealed high similarities (and correlations). This result begets the notion that high fidelity and expensive setups which promote higher ecological validity are not necessary if the goal is just to rank interface designs based on just aesthetics, clarity, and understandability.

The utilisation of a diverse pool of evaluation methods also enabled the investigators to unearth a number of flaws in the design of the interfaces. Unintended effects could be revealed as the interfaces were subjected to increased levels of ecological validity, from participants exploring them from a computer screen to then experiencing the interfaces in a CAVE environment, and ultimately in an AR HMD. The mixture of testing methods allowed for a critical reflection on experience based ('genius') design methodology in contrast to involving the target users from the onset (User Centred Design), which was difficult due to COVID-19 pandemic during the initial design phase in early 2021. Utilising a mixed evaluation approach not only improves the design quality but also ensures that the end product resonates with the user's needs and preferences.

9.2.4 LARGE LANGUAGE MODELS AS A TOOL FOR HUMAN FACTORS RESEARCH

During the course of writing this thesis, LLMs made a significant breakthrough when GPT-3 was released in the summer of 2022, and subsequently made available to the general public in the form of ChatGPT in the autumn of that year. As the technology was gaining traction in possible applications, it was decided to incorporate it into the main research line of

the thesis to assess whether it would be a useful tool in the evaluation stack. Therefore, ChatGPT (versions 3.5 and 4) was used to evaluate the qualitative data that was captured in the pedestrian CAVE simulator experiment presented in Chapter 6, following further experiments that validated the method (presented in Chapter 5).

In the experiments presented in Chapter 5, ChatGPT was found to be proficient in assessing sentiment; this proficiency also extends to its ability to serve as a valid tool for summarising interview data from human subjects, adeptly handling inputs that may contain gibberish while still being capable of highlighting key takeaways. In other experimental contexts, the LLM was also found to be valid in highlighting subtle differences between two task conditions, but more importantly, it was noted that this tool could be applied to bypass a limitation of human cognition, specifically confirmation bias, enabling a more objective analysis.

However, these models are not without their limitations. For instance, due to the autoregressive nature of these models, during a content analysis process, the models were unable to count how many times a particular comment was made. Moreover, there is also the issue of other biases, which are still introduced during the model's training. An example of this is that a summary generated by ChatGPT may highlight comments that do not feature predominantly in the captured qualitative data but are then featured by ChatGPT as major findings due to their alignment with human values or political thought. Hence, as recommended in Chapter 5, it remains important for the time being to involve a human researcher in the pipeline when using ChatGPT. Such involvement ascertains that the outputs produced are balanced, being neither too generic nor excessively detailed, thereby maintaining the integrity of the research process.

The findings presented in this thesis demonstrated that LLMs are a valid tool that can be used to support the researcher during data analysis, with appropriate usage guidelines presented in Chapters 5 and 6.

9.2.5 SIGNIFICANCE OF THIS THESIS

The work in this thesis presents a number of contributions in the domain of pedestrian-automated vehicle interactions, and the utilisation of augmented reality technology in such contexts. By investigating novel AR interfaces for interaction, the research in this thesis offers a novel approach for increasing pedestrian safety in the future urban environment. The outcomes are not just theoretical proposals, but tangible augmented reality interfaces which have been rigorously tested in multiple user evaluations.

A core contribution of the thesis, targeting Objective 4 of the thesis, was a number of expert perspectives (Chapter 2) which not only discussed the potential of AR in the field but also the realities of the current time. Hence, the perspective also presents an expert outlook on the future urban environment, and the effects of further technological advances in the area, such as the continued introduction of AVs in urban areas. The work presented in Chapter 2 has since offered a solid base from which other researchers' work has drawn upon to set the stage before introducing their respective contributions.

Further to expert perspectives, the work in this thesis also gauged (in Chapter 3) the readiness of respondents and participants in their willingness to embrace AR technology for pedestrian-vehicle interactions (Objectives 1 and 4). Moreover, the designed interfaces were not merely speculative but analysed in-depth through several evaluation methods (targeting

Objective 2) which investigated separate elements and measures (Chapters 3, 4, 6, and 7). The approach taken in this thesis ensures that the development of such AR interfaces is not only technologically sound but also corresponds with the expectations and behaviours of the end-users.

Throughout the course of the thesis, new tools and technological developments allowed for explorations that serve as an added contribution to their novelty. For example, Chapter 3 was the first academic work to make use of Unity MARS technology, and therefore served as a demonstration to the Unity team of how their XR stack could be used for research. Moreover, the work in this thesis also harnessed the potential of new (at the time of writing) technologies such as LLMs in the evaluation process.

A novel aspect of the research conducted is the use of eye-tracking technology within a CAVE pedestrian-simulator environment, specifically the first use of an eye-tracker in the HIKER lab at the University of Leeds. The technology enabled a more in-depth investigation into how people will interact with AVs through AR, and more specifically enabling a further contribution of this thesis. Thanks to the pedestrian simulator experiment, the work in this thesis addressed the question of where should the AR elements be displayed and anchored within the user's environment (Objective 3). The insights derived from the simulator experiment (Chapter 6), and further investigations in subsequent work (Chapters 7 and 8) not only inform the placement of AR elements for pedestrian-vehicle interactions but may also serve to inform other AR use cases across various domains, impacting how information is integrated into our physical surroundings. In the same vein, the research not only considered the design and impact of new communication methods but also the critical challenge of distractors in a busy environment such as the one in which these AR interfaces would operate. By investigating how these distractions impact the effectiveness of AR interfaces, the research addresses a vital concern for the design of these systems, which is to ensure that the interfaces capture the attention of users without overwhelming their sensory inputs, as they are met with numerous distractions in real traffic situations.

The initial exploration with experts, and the subsequent design of the interfaces, and their evaluation, leads to perhaps the most significant outcomes of this thesis. That is a set of guidelines for how AR interfaces which support pedestrian-vehicle interactions should be designed and presented. These guidelines, presented throughout the chapters of this thesis, and further encapsulated in the concluding sections, offer insights into how these interfaces could support other cues in the future, and which elements and anchoring locations are best suited to support intuitiveness and trust in interacting with AVs. Furthermore, these findings offer guidance on how new AR interfaces can be designed and presented in future traffic scenarios.

It is envisioned that the learnings from the research presented in this thesis offer a roadmap for future development in the field. They serve as a foundation upon which to build AR interfaces which support safe, transparent, and trustworthy interactions between pedestrians and automated vehicles. The Open Science nature of the research presented also ensures that the appropriate tools and frameworks necessary for others to build upon have been made available, thereby ensuring that other researchers may build upon the work presented here.

9.3 RECOMMENDATIONS AND FUTURE WORK

THE presented thesis has explored the potential of augmented reality technology as a tool for pedestrian-automated vehicle interactions in urban environments. As the objectives presented in the introductory chapter of the thesis were pursued, a number of findings were made which allowed for a number of recommendations to be presented.

1. Insights from Chapter 4 showed that respondents found the non-yielding state of several interfaces clearer than the yielding state. To avoid unnecessary ambiguity, it is recommended to *place the major focus on the dangerous state*, and display only interfaces which communicate danger (such as non-yielding states). Such an approach aims to eliminate confusion and ensure immediate and correct user response in critical moments, whereby safe states are confirmed through implicit cues from the vehicle. It would also scale when multiple vehicles are present, as the situation is either safe or not.

2. It is recommended to initially *display information as head-locked HUDs*, due to their tendency to be less affected by distractors, and lower sensor requirements. Unlike the HUD designs presented in this thesis, the interface should be presented in the peripheral vision to limit any discomfort related to visual conflicts or visual-induced motion sickness (VIMS), and any risks that pedestrians would neglect looking at the rest of the environment and over-rely on the signal. *Road or vehicle-mapped interfaces should then be used once the pedestrian is looking in the correct direction.*

3. The *use of the principle of redundancy gain* is critical Interfaces which had little redundancy, such as the *Field of safe travel* and the *Phantom car*, scored low intuitiveness scores in the user evaluation as they lacked clarity (in terms of top-down cues). Avoid unintended effects by providing additional design elements that support the message to be communicated.

4. It is also recommended to *use universal and user-centric symbols* for redundancy gain. Despite their use in conventional traffic, symbols like the outstretched palm on the *Planes on vehicle*, and *Conspicuous looming plane*s, were reported to cause confusion in the results presented in Chapters 4, 6, and 7. It was unclear to whom (vehicle state or pedestrian) the message was referring to. Therefore, certain symbols which traditionally hold in traditional traffic situations may not indeed do so in AR. The use of reinforcing symbols such as a zebra crossing in a non-yielding state should also be avoided (e.g. the zebra in the *Virtual fence*), as these may diminish the sense of danger.

5. The proximity-compatibility principle should also be adhered to so that *task relevant elements should be presented together to facilitate user interaction*. Hence, the vehicle, and the interface should always remain close to each other in the pedestrian's FOV, requiring minimal head turns.

6. Aim for *higher local embeddedness if the aim is to assist pedestrians in navigating occluded scenarios* (interfaces with higher local presence performed best in Chapter 8). However, for crossing safety tasks, *assisted reality HUD solutions with lower*

local presence would be preferable over mixed reality solutions, to reduce ambiguity if the design is not universally recognised, or visibility issues (as highlighted in the qualitative data captures of Chapters 4 and 6).

7. HUDs are also ideal for scaling up to multi-actor traffic situations that involve multiple pedestrians and vehicles. Information access effort and sensor requirements would be lower for such head-locked interfaces.

8. Keep *the interface design as simple as possible*. Takeaways from Chapter 7 revealed poor interface performance when participants needed to keep track of a lot of activity which was happening around them (since it was a real-world study). The interfaces should aim to maintain good postural stability for the user and ensure safety in the physical world.

9. The *utilisation of arrows to nudge the attention of users is discouraged* in instances where interfaces are not head-locked since it introduces the need for divided attention between a fixed interface and oncoming traffic. Instead of arrows, a signifier in the form of a light (kind of a halo effect) can be introduced to the user's peripheral vision.

10. Diminished reality can be used to remove occluding objects, or even highlight unsafe vehicles by slightly 'diminishing' the safe vehicles.

It is hoped that the presented recommendations will inform future work that will build upon this thesis. However, despite the findings presented in the previous chapters, there are still a number of avenues for potential future development.

Firstly, since the aim of the thesis was to investigate the potential of a speculative AR wearable in future traffic scenarios, the experiments presented only considered one pedestrian interacting with one vehicle approaching from the same direction in a unidirectional road. The next step would be to explore the multi-actor problem in more detail, by investigating how one or more vehicles would interact with one or multiple pedestrians. Furthermore, the ecological validity could be further increased, and real vehicles could be used in a test-track study since subjective safety reported in simulation studies does not necessarily imply objective safety.

The results presented in Chapters 4 and 6 showed that there was not much divergence between gender, and nationality, thereby suggesting that these interfaces scale well in diverse locations. However, the constrained diversity of the online questionnaire study still leaves an open avenue to further investigate such interfaces in other regions of the world (despite the simulator study in Chapter 6 having a more diverse participant pool, but limited sample size).

It is also still unclear which universal traffic symbols will scale well in an AR interface setting. It is therefore recommended that future work should evaluate which symbols clearly communicate their intended message to the target audience without any ambiguity. Furthermore, there are grounds for formal design guidelines related to such AR interfaces to be created.

Another element that was explored in this thesis was the utilisation of Large Language Models (LLMs) for the analysis of qualitative data that was captured as part of the conducted experiments. A further potential utilisation of the technology would be to embed it into the

AR interfaces themselves and utilise the LLM to build context and response to a situation. For example, the environment around the user is monitored through cameras and LiDARs, and custom AR interfaces would be built dependent on the situation, or the user's input (e.g. voice input). Of course, at the time of writing, the technology is not mature enough to be used at a consumer level for such an application, however potential for experimental testing exists. Furthermore, as an addendum to the use of LLMs, consumer-level AR HMDs could also be tested in outdoor scenarios in order to investigate their potential for road use.

9.4 Future Vision and General Conclusion

Self-driving taxis are already ferrying passengers across cities like Las Vegas and San Francisco. Automated vehicles in the urban environment are no longer futuristic concepts but are already to be found around us. The current proliferation of these vehicles and the upcoming developments in spatial computing are testament to the relevance of the research presented in this thesis; for the issues and solutions discussed are no longer mere speculations, but imminent realities which will shape our environment in the years to come.

The future urban environment will consist of automated vehicles co-existing with micro-mobility options, and pedestrians. In contrast to the current taxis, which are not equipped with eHMIs, further integration of AVs in daily life, and less uncontrolled urban environments, necessitates the need for AVs to convey intentions clearly, as echoed by experts earlier in this thesis. Such communication would not only be about conveying messages, but also about resolving uncertainties and addressing liability concerns. Vulnerable road users would still wait for a vehicle to stop before crossing the road or engaging in any other behaviour, but they would await explicit confirmation from the vehicle, which would be presented in the form of interfaces.

Just like automation has allowed the 'drivengers' in a vehicle to engage in non-driving tasks, so too will automation allow other road users to engage in other tasks such as consuming media, while only receiving traffic information when needed. Augmented reality technology can accomplish this; just like how drivers are already making use of AR inside the vehicle for driving tasks, so would other road users benefit from the technology in the coming years. Through AR, they can seamlessly receive traffic information, only when necessary, without disrupting other tasks. This approach reflects a broader trend of AR's integration into everyday life, not only inside traffic but also in other settings like warehouses, airports, factories, museums, educational settings, and also in the private context, such as the home, as noted in recent consumer releases of more capable HMDs.

We are likely to witness a number of AR interfaces that would be introduced to support pedestrian-vehicle interactions. Perhaps, these would be different from those speculated in this thesis, yet the exploratory and speculative designs that have been discussed and evaluated in this thesis would offer a foundation of guidelines that will be instrumental in building optimal solutions. These interfaces would offer a number of customisation options to the user, just like current vehicle interiors are customisable through software. Moreover, through wearables, pedestrians would also be able to navigate areas which present occluded scenarios thanks to diminished reality that will visualise occluded objects, and reduce user-defined visual clutter (such as advertisements). DR which elicits high local presence would offer alternate realities indistinguishable from reality, in contrast to solutions that just overlay information. As sensor technology improves, so too will the capabilities of these interfaces,

creating possibilities that are so far unimaginable.

The met objectives of this thesis offer a deeper understanding of how AR can shape the future of urban traffic interactions. It highlights key considerations for effective user interaction and buildup of trust in automation; while making recommendations to open avenues for further future research. The journey from speculative to practical, from experimental to everyday life, is fraught with challenges and opportunities.

In conclusion, this thesis does not just provide a snapshot of where we are; it offers a roadmap for where we are headed, in a world increasingly defined by intelligent machines and augmented realities, that will drive us into the future.

Acknowledgments

It is said that no man is an island, and what truer words have been uttered. Slightly over four years ago, on a misty November morning, as I sat in one of my promotor's offices, I was told that I was accepted for a Marie Curie PhD program at TU Delft. Thus began a long-awaited and much anticipated academic journey that was to shape the next four years of my life. A period of growth, reflection, exploration, and learning. The period between 2020 - 2024 was rather turbulent in the grand scheme of things, and keeping the ship on a steady course was no easy task; for a captain cannot do so alone. Which is why I am thankful for all the people that remained or came into my life during this time, and impacted my journey for the better.

Firstly, I want to express my sincere gratefulness and appreciation to my promotors, **Prof. Joost C. F. de Winter** and **Prof. Riender Happee**. Joost, as my advisor, you were a steady presence throughout the course of my time in Delft, always offering academic advice, and sharp insights about my work. Your great keen eye for detail was of great help in sharpening things when they were unclear. I appreciate the autonomy you afforded me throughout the past years, where I was free to take the project into the direction I wanted. Riender, thank you for the advice you provided on numerous things, and for helping me be more assertive and 'Dutch' when tackling problems. Our conversations, with some banter here and there, were always quite enjoyable. Despite both of you having busy schedules, you always made time for me, and at points were available daily to chat. All in all you have made me a better scientist, with the capacity of being more critical in general. Thank you for everything. I want to also thank all my previous supervisors, and other academics who had encouraged me to follow my wish to persue a PhD. Your push led me here.

Throughout the course of the PhD, I had the pleasure of collaborating with over 44 individuals whom I am proud to count as my co-authors on a number of journal articles, conference papers, and project deliverables. The list is too long to mention here, yet some feature later in the acknowledgements. In general, I am thankful for the opportunity to work with you all, and for giving me the opportunity to grow and learn alongside you.

I have only the deepest of gratitude to my dear societas, **Alberto Bertipaglia**, and **Hidde Boekema** (my paranymphs and office partners in crime), and **Vishrut Jan** (awesome housemate). Alberto, and Hidde, I will cherish all the memories we have built together during my time in Delft. From operas in Amsterdam, to our adventures in Gouda, Maastricht, and Groningen, then all the way to Belgium, Ireland, UK, Italy, France, and Malta. The PhD would not have been the same without you and all the memorable experiences we shared together. Thanks for all the lovely philosophical discussions and debates during IO lunches, all the films we watched, museums we visited, chairs we discussed, tunes we danced to, heritage sites we visited, and of course the nice times at ID Cafe! Vishrut, thanks for being a cheerful presence in most places that we have been together, for the discussions after work, and the great company throughout the PhD. In thea veritas.

I want to also thank my fellow TU Delft SHAPE-IT early stage researchers (ESRs). **Siri Hegna Berge**, thank you for listening and discussing, especially when we both needed

to vent, also cheers for your travel companionship (and all the anecdotes we collected throughout the years), and for all the spontaneous days spent shopping in Den Haag or Rotterdam. It was always a great pleasure to hang, to listen, and to share. **Xiaolin He**, thank you for being a great office mate, and travel companion throughout the project. It was lovely to get to know you better during the SHAPE-IT events, and I have learned a lot about Chinese culture thanks to you.

Belonging to the Human-Robot Interaction department section, yet sitting in Intelligent Vehicles, meant that I got to feel like I belong to two families. Thanks to my IV colleagues, **Varun Kotian**, **Chrysovalante Messiou**, **Ted de Vries**, **Shiming Wang**, **Zimin Xia**, **Mubariz Zaffar**, **Dr. Yancong Lin**, **Dr. Marko Cvetkovic**, and **Gabriele Rini** (visitor) who were fantastic fellow E-wingers; I wish you all the best in your future careers. Thank you **Dr. Barys Shyrokau** for always willing to provide advice throughout the years. From the old guard, **Dr. Pavlo Bazilinskyy**, for being a role model in how a MCA student should be like, **Dr. Tuğrul Irmak**, a man of impeccable determination, and fellow thesis into business course taker (yet a business still needs to come out of it), and **Dr. András Pálffy**, fellow desk mate, fan of Norman doors, and someone who I wanted to thank for his initial assistance as I settled in during those lovely 2020 times. From the Learning and Autonomous Control, and the Robot Dynamics sections, a thank you goes to **Dr. Corrado Pezzato**, and **Elia Trevisan**, for including me in their events in the past few years.

From the HRI side, I want to thank **Prof. David A. Abbink** and **Dr. Laura Marchal Crespo** for their leadership, and **Dr. Arkady Zgonnikov** for his support in the early days, and whom I got to know more when we bumped into each other in Malta. I am also grateful that I got to know **Vishal Onkhar**, and am thankful to him for always taking up the opportunity to geek out on something together, and discuss lore, which is great fun; and also **Tom Driessen**, and **Olger Siebinga** whom I will have fond memories of traversing all across New York with. Tom, I am sure you are still happy to this day that the Cursed Child on Broadway wasn't a musical.

From BioMechanical Engineering, I want to thank **Dr. Dimitra Dudou**, who I had the pleasure of collaborating with on a number of works throughout the project. Some fond memories that come to mind include the two days we spent in the Rijksmuseum during COVID times, locked in with masterpieces of Rembrandt and Vermeer. Now that was something. And of course the time I had to cycle around the Civil Engineering building in freezing weather, for the love of science and data. That was actually a nice experiment. Further to our working together, you were always available to talk, to give advice, and to provide opinions on other issues that were not necessarily about the work at hand. For this, I am grateful. Also, thanks to the secretariat of the department for always being of assistance with any issues that arise.

I want to also thank a number of friends and colleagues from the Faculty of Industrial Design Engineering. **Dr. Maria Luce Lupetti**, for allowing me to build a bridge between my department and the department of human-centred design, and for the lovely discussions we had at CHI over dinner. **Soyeon Kim**, for being such a great honorary IO guest at the ME events, and for the lovely moments at IO cafe. **Willem van der Maden**, whom I had the pleasure of getting to know at CHI, and subsequently during some interesting discussions at IO symposia. Finally, **Caiseal Beardow**, for being 'the design' friend during the last two years of my PhD. It was definitely a pleasure to attend so many art events and museums

across the Netherlands together, and to discuss the matter at hand. You added a touch of colour to some grey walls, and we proudly became honorary KABKers together.

The project I was part of, the SHAPE-IT Marie Curie ITN, would have not have been successful without the great contributions of my fellow ESRs: **Nikol Figalová**, **Naomi Mbelekani**, **Chi Zhang**, **Yue Yang**, **Chen Peng**, **Yuan-Cheng Liu**, **Amna Pir Muhammed**, **Siri Hegna Berge**, **Xiaolin He**, **Sarang Jokhio**, **Dr. Amir Hossein Kalantari**, **Ali Mohammadi**, and **Xiaomi Yang**. Together we managed to rise above the challenges we faced during those dark initial months of the pandemic. Despite meeting in person for the first time more than a year and a half then originally planned, we managed to create a wonderful group that managed to have fun during all the training events. I have fond memories of treasure hunts, escape rooms and karaoke in Ulm, high tea in Delft, fashion competitions around the professors' table in Munich, a wonderful time touring Nordic islands by boat and visiting the Liseberg theme park in Gothenburg during Midsummer, and of course a grand Lord of the Rings-esque grand hike around Malham in North Yorkshire. All of this would not have been possible without the great support of the principal investigator **Dr. Jonas Bärgman**, and project administrator extraordinaire **Jacqueline Plette** who made sure that everything ran smoothly and always made themselves available to support us during the difficult times we individually faced. A further thank you goes to the supervisory team **Prof. Natasha Merat**, **Prof. Martin Baumann**, **Prof. Marjan Hagenzieker**, **Prof. Christian Berger**, **Dr. Eric Knauss**, **Prof. Gustav Markkula**, **Prof. Klaus Bengler**, **Prof. Marco Dozza**, **Prof. Olga Pollatos**, **Dr. Tobias Hecht**, **Prof. Meng Wang**, and **Dr. Yee Mun Lee**, for steering the project, and hosting us in their respective universities (**Delft University of Technology**, **University of Leeds**, **Chalmers University of Technology**, **Ulm University**, **Technical University of Munich**, and the **University of Gothenburg**) for training and dissemination events. Thanks also to **Prof. John D. Lee**, and **Prof. Linda Boyle** who oversaw the project from a scientific standpoint. It was a great pleasure to have you visiting us from the USA to deliver targeted Human Factors training sessions for us. I also cherish fond memories of some lengthy hike discussions on shared interests ranging from pens to Harry Potter.

As part of my Marie Curie secondment commitments, between April and July 2022, I was hosted by the Human Factors and Safety Group at the Institute of Transport Studies, University of Leeds, where I conducted the CAVE simulator experiment. I want to firstly thank **Prof. Natasha Merat** for her hospitality and for facilitating my visit, **Dr. Yee Mun Lee** for her continuous academic support and insights during the experiment, and of course fellow ESRs **Chen Peng** for the support and fun events, and **Yue Yang** for all the help during those long hours in the HIKER simulator. A further thanks to **Jorge García** for all the technical development, assistance, and great talks on digital heritage and culture. In general I want to thank the group for their hospitality, and for the great walks around Yorkshire, from Knaresborough, to Harrogate, and York. My time in Yorkshire would not have been as fun without the kind contributions of **Dr. Vishnu Radhakrishnan**, **Dr. İbrahim Öztürk**, **Dr. Albert Solernou**, **Ehsan Sadraei**, **Dr. Courtney Goodridge**, **Dr. Rafael Goncalves**, **Thanos Tzigieras**, **Yang Li**, and of course **Max Theisen** who was a fellow visitor. Cheers to some great times at the nearby pubs. The Fenton lives on, and Leeds will retain a special place in my heart.

I want to also take this opportunity to also thank my friends from the USA, **Paul Schmitt** and **Avram Block**. Despite not getting the opportunity to work together face to face in

Boston during my planned second secondment, we still managed to meet up at MassRobotics events, and collaborate together on two papers during the last year of my PhD.

During my time at the TU Delft, I had the absolute pleasure of mentoring over 22 MSc students, and supervising 8 BSc, and 4 MSc students. A special shout-out to the (awesome) MSc students which I supervised: **Max Oudshoorn** (2020), **Lokkeshver Thirunavukkarasu Kumaaravelu** (2022), **Joris Peereboom** (2022), and **Thomas Aleva** (2023). I want to thank them all for all the lessons I learned from them, and for collectively helping me in my continuous development.

Moreover, I want to thank all the 1,609 experiment participants that I had throughout the four years at the University. Your contribution to my, and my student's experiments, made the work worthwhile and meaningful.

While at the University, I also got involved with art-related activities which were of a great benefit to my growth as a creative. I want to thank **Dr. Eduardo Mendes** for facilitating the wonderful art-science course back in 2022, which further deepened my interest in new media art. I subsequently got involved in a team effort to design a permanent art sculpture on the TU Delft campus. Firstly thanks to the main artists, **Niels Albers**, and **Jan Koen Lomans** for involving me and entrusting to be part of their team; and of course my fellow student team members **Rita Gorrìz**, **Olaf Kamperman**, **Pablo Yániz González**, **Joshua Gallegos**, **Jens van Heijningen**, and **Tieme van Wijnen**, who were great fun to work with. I will fondly remember all our cultural trips, design sessions in the botanical gardens, and the creative journey we had.

In 2022 I went on what felt to be a grand tour across the world to attend conferences and project events, since travel restrictions had finally eased. I am deeply appreciative to all the lovely people I met and had insightful discussions with in New York, Leiden, Gothenburg, Singapore, Hamburg, and Copenhagen. In these places I not only appreciated the interaction with attendees, but also enjoyed visiting a multitude of museums which offered a fingerprint of the place. In a way, I want to thank the curators of all the (around 100) Dutch and international cultural institutions and museums I visited over the course of two years and a half for teaching me a lot, and offering places of respite and reflection throughout my PhD.

Separate from University life, I want to also thank **H.E. Mark Pace**, the Ambassador of Malta to the Netherlands, for creating a home far away from home in the Hague, and for fostering community building for us Maltese living in the Netherlands. I will cherish the Malta day events we had at the Embassy, and also the opportunity of being a Maltese flag bearer on a canal boat during a very rainy Pride parade in Amsterdam. What a memorable and emotional day that was.

I also want to thank a number of friends in the Netherlands. **Sem van Royen**, I am deeply grateful for our friendship since the Groningen days. Your constant and reassuring presence in the Netherlands offered a continuation from those great times spent in the north. I appreciate all the events we organised together, your trips to Malta, our co-working sessions, and of course the advice we always offered each other. **Ricardo Brouwer**, I am also happy that we kept in touch after Groningen. I am particularly thankful for the soft landing in 2020, and also fond of the time we spent exploring around Arnhem, and grateful for the care you afforded me in that 2021 incident. I wish you well in your Primeministerial journey.

Despite being away from the rock, I am thankful for a number of friends from Malta and Gozo, which kept in touch during the past four years. Collectively, I am grateful to the many

people which have encouraged me during the past years, most especially in the first two years when I was trapped home because of the pandemic. I am still deeply touched by all those local friends which helped out in 2021 when I was facing difficulties. I am deeply grateful to **Arch. Adrian Mamo**, **Andrew Spiteri**, **Melvin Zammit**, **Jon Aedyn King**, and **Simon Godfrey**. Your friendship, discussions, dinners, lunches, and your encouragement, meant a lot to me. You always had time to meet up when I was around, and it was always a great joy to be around you. **Eman Borg**, **Antonella Bugeja**, **Antoine Spiteri**, **Cat Camilleri**, **Luke Camilleri**, and **Reb Xiberras**; it was my honour to still work alongside you as part of the LGBTI+ Gozo executive, also during my time in Delft. But more importantly I am grateful for your friendship, and the great times we spent together - most especially during EuroPride 2023, which gave us some amazing memories that we will cherish for years to come. I also want to thank the choir I form part of, **Cappella Sanctae Catharinae** for retaining an open door for me, but for also offering me a space to let go of all stress and lose myself in great music during the pandemic. Dear cousin **Sachelle Schembri**, thank you for your support throughout life and the great memories we share. **Nicholas Abela**, **Lucas Farrugia**, and **Reynold Sacco**; thank you for your long years of friendship, for always having time to meet up every time I visited the rock, for the great events we attended together during this time, the festas, and Lucas for visiting me. But what I would most cherish during the past four years are the long walks we went on during the pandemic, where we could vent our frustrations to one another in a safe space and offer advice to each other. That was indeed a special time, and I am thankful for it.

Last but not least, I want to deeply thank my bedrock and support system. My grandparents **Saviour Tabone**†, **Josephine Tabone**†, **Victor Sacco**, and **Mary Grace Sacco**, for always allowing me to explore my creativity in childhood, and for being ever supportive during the period I was away. I am also deeply grateful to my dearest parents, **Jesmond Tabone** and **Carmen Tabone**, my sister **Lenice Cassar Tabone**, brother in law **Gabriel Cassar Tabone**, and of course my dear partner **Kurt Meli**. Dear family, not only were you my companions throughout life, but also my lab mates and fellow experimentalists in the initial work from home period in the first two years of my PhD. Your help in my studies and outreach were much appreciated. You, and your sacrifices and love are the reason I am. It is your unrelenting love and constant support that nourished and pushed me, as I steered the ship through stormy waters in the past few years. This thesis is dedicated to you.

Nixtieq ukoll nirringrazzja lill-membri tal-familja kollha li inkoraġġewni matul dan il-vjaġġ, lin-nanniet, iz-zijiet u l-kuġini, u nirringrazzja wkoll lil Daniela, Leli, u Brian tal-appoġġ kollu tagħhom. Għeżież ġenituri, grazzi ta' mħabbitkom u talli dejjem kontu hemmhekk għalija, u talli pprovdejtuni b'dak kollu li kelli bżonn sabiex nasal hawn, minkejja li xi kultant dan kien jirrikjedi s-sagrifiċċji. Ta' dan grat ħafna u kuntent li dak li għentu tibnu ssarraf s'hawn. Kburi li ġej minn ġenerazzjonijiet tal-ħaddiem, li stinkaw qatigħ biex jibnu pedament sod għal dawk li ġew warajhom u jarawhom f'post aħjar milli sabu ta' qabilhom. Grat ukoll li kontu hemm għalija fi żminijiet iebsa li għaddejt minnhom fl-ewwel sentejn tad-dottorat minħabba l-effetti tal-pandemija. Minkejja li forsi kien diffiċli li tifħmu eżatt minn xiex kont għaddej dak iż-żmien, kontu dejjem ta' għajnuna — mhux l-inqas li saħansitra għentu fl-esperimenti li kont qed nagħmel. Lenice, minn dak l-ewwel filmat li ġbidtli għall-proġett lura f'Mejju 2020, sa issa, għaddew bosta avvenimenti u ġrajjiet li sawru iżjed l-għaqda ta' bejnietna. Napprezza li inti u Gabriel kontu ta' wens waqt dan

il-vjaġġ twil tul l-aħħar erba' snin, grat tal-inkoraġġiment u l-imħabba tagħkom dejjem. Fl-aħħar nixtieq nirringrazzja lilek għażiż Kurt ta' dak kollu li għamilt miegħi, tal-wens u s-sabar, tal-paċenzja u fl-aħħar mill-aħħar ta' mħabbtek. Inti għext dan il-vjaġġ miegħi, fit-tajjeb u l-ħażin, u l-preżenza tiegħek kemm Malta u l-Olanda, inħasset u swiet ħafna. Ta' dan grat immens.

Fit-tlajja' u l-inżul kontu kollha hemm miegħi. Grazzi mill-qalb.

Wilbert Tabone
Delft, November 2023

Curriculum Vitæ



Wilbert Tabone

May 19, 1994	Born in Pietà, Malta.

Education

2020-2024	PhD. in Cognitive Human-Robot Interaction as a Marie Skłodowska-Curie Early Stage Researcher at Cognitive Robotics, **Delft University of Technology**. *Visitor*: University of Leeds (UK). *Thesis*: Bridging Worlds: Augmented Reality for Pedestrian-Automated Vehicle Interactions.
2015-2017	MSc. in Artificial Intelligence, **University of Malta**. *Visitor*: University of Groningen (NL).
2016-2016	Dip. in European and International Studies, **University of Vienna**.
2012-2015	BSc. (Hons.) in Creative Computing, **Goldsmiths, University of London**. *Graduated with first class honours.*

PROFESSIONAL EXPERIENCE

2022-PRESENT	Member of Government Commission THE NATIONAL COMMISSION FOR THE PROMOTION OF EQUALITY, MALTA.
2020-2024	Marie Skłodowska-Curie Early Stage Researcher DELFT UNIVERSITY OF TECHNOLOGY, THE NETHERLANDS.
2019-2020	Creative Computing Executive, and Data Protection Officer HERITAGE MALTA, MALTA.
2018-2019	Member, and Secretary of Government Board MALTA NATIONAL TASK FORCE ON ARTIFICIAL INTELLIGENCE, MALTA.
2017-2019	Project, Communications, and Information Technology Consultant MALTA NATIONAL MUSEUM OF ART, MALTA.
2017-2017	Communications, and Digital Technology Officer MINISTRY FOR TOURISM, MALTA.
2015-PRESENT	External Quality Assurance Auditor and Consultant MALTA FURTHER & HIGHER EDUCATION AUTHORITY, MALTA.
2015-2016	Junior Software Developer IXARIS SYTEMS LTD., MALTA.
2014-2014	Summer Student Software Developer GO PLC., MALTA.
2012-2013	Intern Quality Assurance Technician BIGPOINT INTERNATIONAL SERVICES LTD., MALTA, AND GERMANY.

ORGANISATIONAL EXPERIENCE

2023-PRESENT	Member of TU Delft Artistic Team EXECUTIVE BOARD APPOINTED PERMANENT SCULPTURE PROJECT, THE NETHERLANDS.
2023-PRESENT	Advisory Board Member LGBTI+ GOZO., MALTA.
2022-PRESENT	Consortium Member SOCIALLY AWARE AUTONOMOUS MOBILITY (SAAM) CONSORITUM., THE UNITED STATES.

2019-2020	Committee Member MALTA SOCIETY OF ARTS., MALTA.
2018-2022	Subcommittee Member - Europe and Africa Region COMMONWEALTH YOUTH COUNCIL., THE UNITED KINGDOM.
2018-2023	Executive Member LGBTI+ GOZO., MALTA.
2018-2018	Secretary General WEARE - YOUTH & STUDENT LGBT ORGANISATION., MALTA.
2016-2017	International Subcommittee Member GANYMEDES LGBT STUDENTENVERENIGING., THE NETHERLANDS.
2015-2016	Public Relations Officer GOOGLE DEVELOPERS GROUP., MALTA.
2015-2016	Media and International Subcommittee JEF, MALTA.
2015-2015	Artist and Floor Management COMMONWEALTH HEADS OF GOVERNMENT MEETING (CHOGM)., MALTA.
2014-2016	Royal Navy Petty Officer, and Instructor MALTA SEA CADET CORPS (SCC)., MALTA.
2012-2016	Board Member INTERNATIONAL GAME DEVELOPERS ASSOCIATION (IGDA)., MALTA.

List of Outreach Activities

1. **Tabone, W.** (2018). Artificial Intelligence is upon us - Wilbert Tabone. https://timesofmalta.com/articles/view/artificial-intelligence-is-upon-us-wilbert-tabone

2. Andrew, D. (2019). Interview with Wilbert Tabone, MUŻA, Malta. https://uk.icom.museum/interview-with-wilbert-tabone-muza-malta/

3. Costa, M. (2019). [WATCH] Wilbert Tabone: 'AI will empower humans, not suppress them'. https://www.businesstoday.com.mt/people/people/499/watch_wilbert_tabone_ai_will_empower_humans_not_suppress_them

4. **Tabone, W.** (2021). Vodcast on automated vehicles. https://www.facebook.com/100067547438124/videos/139781804908165

5. Gent, E. (2022). Making driverless cars more expressive: Exaggerated stopping movements help pedestrians read autonomous cars' minds. https://spectrum.ieee.org/self-driving-car-safety-features

6. **Tabone, W.** (2023). Self-driving cars: will they really happen? https://medium.com/@wiltabone/self-driving-cars-will-they-really-happen-f768ac21119

7. **Tabone, W.** (2023). Case Study: AR interfaces for automated vehicle-pedestrian interactions. https://medium.com/@wiltabone/self-driving-cars-will-they-really-happen-f768ac21119

8. **Tabone, W.** (2023). Public address on digitisation and AI to the Maltese cabinet of ministers and chiefs of the public service. https://bit.ly/3RbAEjp

9. **Tabone, W.** (2023). Speech on AI and automation delivered during a hearing on 'young people and mental health'. European Economic and Social Committee and the European Parliament. Brussels. https://www.eesc.europa.eu/en/agenda/our-events/events/young-people-and-mental-health

DEMO VIDEOS

1. **Tabone, W.**, Lee, Y. M., Merat, N., Happee, R., & De Winter, J. C. F. (2021). *Video Demo: To-wards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes* (YouTube; Version V1) [Video]. YouTube. https://youtu.be/tpWv0cW2vJQ

2. **Tabone, W.**, Happee, R., García, J., Lee, Y. M., Lupetti, M. C., Merat, N., & De Winter, J. C. F. (2023). *Video Demo: Augmented reality interfaces for pedestrian-vehicle interactions: An online study* (YouTube; Version V1) [Video]. YouTube. https://youtu.be/t9tuTA3vgn4

List of Open Data Sets

1. **Tabone, W.**, Lee, Y. M., Merat, N., Happee, R., & De Winter, J. C. F. (2021). *Supplementary materials for the article: Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes* (4TU. ResearchData; Version V1) [Data set and software]. 4TU. ResearchData. https://doi.org/10.4121/14933082.v1

2. **Tabone, W.**, Dodou., D., & De Winter, J. C. F. (2022). *Supplementary data for the paper 'How do people distribute their attention while observing The Night Watch?* (4TU. ResearchData; Version V3) [Data set]. 4TU. ResearchData. https://doi.org/10.4121/20496411.v3

3. **Tabone, W.**, Happee, R., García, J., Lee, Y. M., Lupetti, M. C., Merat, N., & De Winter, J. C. F. (2023). *Supplementary materials for the article: Augmented reality interfaces for pedestrian-vehicle interactions: An online study* (4TU. ResearchData; Version V1) [Data set]. 4TU. ResearchData. https://doi.org/10.4121/21603678.v1

4. **Tabone, W.**, & De Winter, J. C. F. (2023). *Supplementary materials for the article: Using ChatGPT for Human Computer Interaction Research: A Primer* (4TU. ResearchData; Version V1) [Data set]. 4TU. ResearchData. https://doi.org/10.4121/21916017.v1

List of Publications

1. **Tabone, W.**, & Seychell, D. (2015). Recognising Familiar Facial Features In Paintings Belonging to Separate Domains. Proceedings of the 16th International Conference on Computer Analysis of Images and Patterns (pp. 125-136).

2. **Tabone, W.**, Wilkinson, M.H.F., Van Gaalen, A.E.J.V., Georgiadis, J., & Azzopardi, G. (2019). Alpha-Tree Segmentation of Human Anatomical Photographic Imagery. Proceedings of the 2nd International Conference on Applications of Intelligent Systems (pp. 1-6).

3. **Tabone, W.** (2019). Technology for Cultural Heritage. Tesserae Spring Edition. Heritage Malta.

4. **Tabone, W.** (2020). The Effectiveness of an Augmented Reality Guiding System in an Art Museum. Rediscovering Heritage Through Technology (pp. 197-214). Springer, Cham.

5. **Tabone, W.**, De Winter, J. C. F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock. P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B.,& Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transportation Research Interdisciplinary Perspectives , 9, 100293.

6. **Tabone, W.**, Lee, Y.M., Merat, N., Happee, R., & De Winter, J.C.F. (2021). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 209–218).

7. Figalová, N., Mbelekani, N.Y., Zhang, C., Yang, Y., Peng, C., Nasser, M., Yuan-Cheng, L., Pir Muhammad, A., **Tabone, W.**, Berge, S. H., Jokhio, S., He, X., Hossein Kalantari, A., Mohammadi, A., & Yang, X. (2021). SHAPE-IT Deliverable 1.1: Methodological framework for modelling and empirical approaches.

8. Merat, N., Yang, Y., Lee, Y. M., Berge, S. H., Figalová, N., Jokhio, S., Peng, C., Mbelekani, N.Y., Nasser, M., Pir Muhammad, A., **Tabone, W.**, Yuan-Cheng, L., & Bärgman, J. (2021). SHAPE-IT Deliverable 2.2: An overview of interfaces for automated vehicles (inside/outside).

9. De Winter, J. C. F., Dodou, D., & **Tabone, W.** (2022). How do people distribute their attention while observing The Night Watch? Perception, 51, 763–788.

10. **Tabone, W.**, Happee, R., García, J., Lee, Y. M., Lupetti, M. C., Merat, N., & De Winter, J. C. F. (2023). Augmented reality interfaces for pedestrian-vehicle interactions: An online study. Transportation Research Part F: Traffic Psychology and Behaviour, 94, 170–189.

11. Joshi, S., Block, A., **Tabone, W.**, Pandya A., & Schmitt, P. (2023). Advancing the State of AV-Vulnerable Road User Interaction: Challenges and Opportunities. Proceedings of the AAAI 2023 Spring Symposium. Palo Alto, California, United States.

12. **Tabone, W.**, & De Winter J. C. F. (2023). Using ChatGPT for Human–Computer Interaction Research: A Primer. Royal Society Open Science, 10, 231053.

13. **Tabone, W.**, Happee, R., Yang, Y., Sadraei, E., García, J., Lee, Y. M., Merat, N., & De Winter J. C. F. (2023). Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment. Under review.

14. Peereboom, J*., **Tabone, W*.**, Dodou, D., & De Winter J. C. F. (2023). Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios. Under review.

15. Block, A., Joshi, S., **Tabone, W.**, Pandya A., Lee, S., Patil, V., Britten, N., & Schmitt, P. (2023). The Road Ahead: Advancing Interactions between Autonomous Vehicles, Pedestrians, and Other Road Users. Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). Busan, South Korea.

17. Figalová, N., Mbelekani, N.Y., Yang, Y., Yuan-Cheng, L., **Tabone, W.**, Hossein Kalantari, A., Mohammadi, A., & Yang, X. (2023). SHAPE-IT Deliverable 1.4: An extension of the human-factors methodological toolbox for human-AV interaction design research.

18. De Winter, J. C. F., Berge, S. H., **Tabone, W.**, Yang, Y., Pir Muhammad, A., Jokhio, S., & Hagenzieker, M. P. (2023). SHAPE-IT Deliverable 2.5: Design strategies and prototype HMI designs for pedestrians, cyclists, and non-automated cars.

16. Aleva, T*., **Tabone, W*.**, Dudou, D., & De Winter J.C.F. (2024). Augmented reality for supporting the interaction between pedestrians and automated vehicles: An experimental outdoor study. Frontiers in Robotics and AI, 11, 1324060.

Included in this thesis.
* Equal contribution (joint first authors).

Propositions

accompanying the dissertation

BRIDGING WORLDS: AUGMENTED REALITY FOR PEDESTRIAN-AUTOMATED VEHICLE INTERACTIONS

by

Wilbert TABONE

1. It is unlikely that eHMIs will make it to the market.
   This proposition pertains to Chapter 2 and Chapter 4.
2. It is futile to introduce automated vehicles in locations predominated by car culture.
   This proposition pertains to Chapter 2.
3. Experience-based or 'genius' design should be avoided when designing for augmented reality.
   This proposition pertains to Chapters 3, 4, 6, 7 and Chapter 8.
4. Concerns about Large Language Models are the direct result of poor scientific communication.
5. It is correct for European research projects to have European associated partners exclusively.
6. The COVID-19 pandemic had a long-lasting impact on the well-being of PhD candidates.
7. Marie Curie Initial Training Networks should serve as a model for how other PhDs should operate.
8. There is a positive correlation between productivity and loosely defined working hours.
9. Art courses should be part of scientific education, and vice versa.
10. With VR headsets offering immersive experiences from the sofa, there is no need to go outside with AR headsets.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotors prof.dr.ir. Joost C.F. de Winter, and prof.dr.ir. Riender Happee.

