

---

# DESIGNING GAZE-BASED INTERACTION FOR PERVASIVE PUBLIC DISPLAYS

---

## DISSERTATION

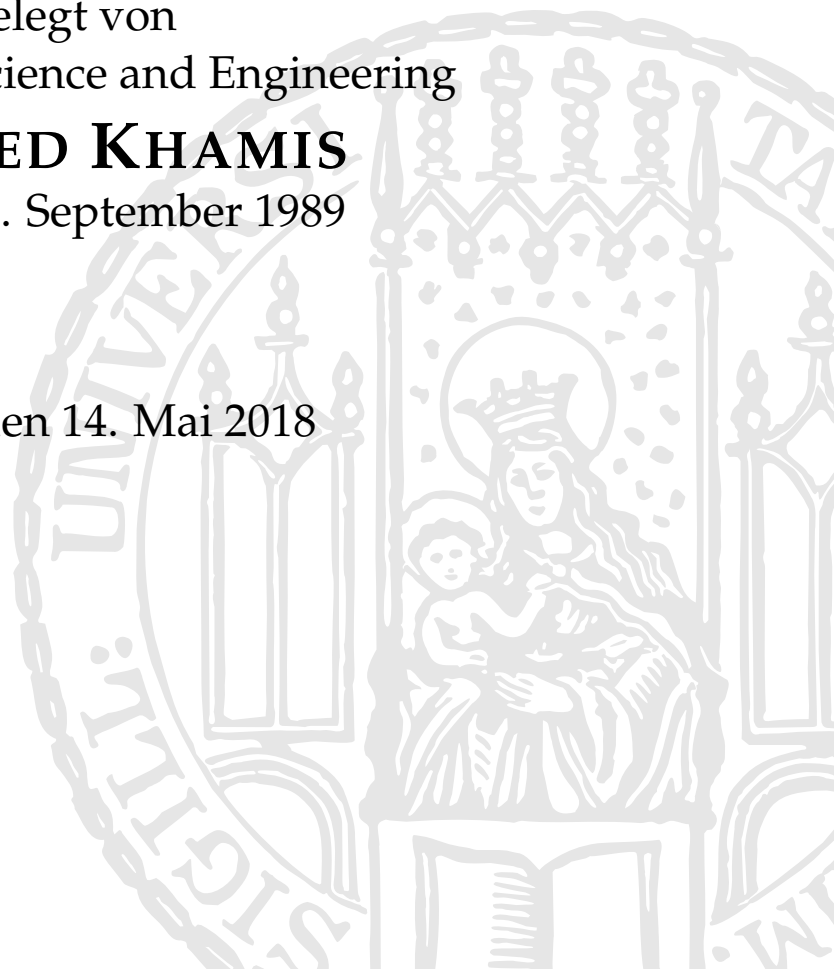
an der Fakultät für Mathematik, Informatik und Statistik  
der Ludwig-Maximilians-Universität München

vorgelegt von  
M.Sc. Computer Science and Engineering

**MOHAMED KHAMIS**

geboren am 4. September 1989

München, den 14. Mai 2018





---

Erstgutachter: Prof. Dr. Florian Alt  
Zweitgutachter: Prof. Dr. Andreas Bulling  
Drittgutachter: Prof. Dr. Marc Langheinrich

Tag der mündlichen Prüfung: 7<sup>th</sup> of August, 2018





## ABSTRACT

The last decade witnessed an increasing adoption of public interactive displays. Displays can now be seen in many public areas, such as shopping malls, and train stations. There is also a growing trend towards using large public displays especially in airports, urban areas, universities and libraries.

Meanwhile, advances in eye tracking and visual computing promise straightforward integration of eye tracking on these displays for both: 1) monitoring the user's visual behavior to evaluate different aspects of the display, such as measuring the visual attention of passersby, and for 2) interaction purposes, such as allowing users to provide input, retrieve content, or transfer data using their eye movements. Gaze is particularly useful for pervasive public displays. In addition to being natural and intuitive, eye gaze can be detected from a distance, bringing interactivity to displays that are physically unreachable. Gaze reflects the user's intention and visual interests, and its subtle nature makes it well-suited for public interactions where social embarrassment and privacy concerns might hinder the experience.

On the downside, eye tracking technologies have traditionally been developed for desktop settings, where a user interacts from a stationary position and for a relatively long period of time. Interaction with public displays is fundamentally different and hence poses unique challenges when employing eye tracking. First, users of public displays are dynamic; users could approach the display from different directions, and interact from different positions or even while moving. This means that gaze-enabled displays should not expect users to be stationary at a specific position, but instead adapt to users' ever-changing position in front of the display. Second, users of public displays typically interact for short durations, often for a few seconds only. This means that contrary to desktop settings, public displays cannot afford requiring users to perform time-consuming calibration prior to interaction.

In this publications-based dissertation, we first report on a review of challenges of interactive public displays, and discuss the potential of gaze in addressing these challenges. We then showcase the implementation and in-depth evaluation of two applications where gaze is leveraged to address core problems in today's public displays. The first presents an eye-based solution, EyePACT, that tackles the parallax effect which is often experienced on today's touch-based public displays. We found that EyePACT significantly improves accuracy even with varying degrees of parallax. The second is a novel multimodal system, GTmoPass, that combines gaze and touch input for secure user authentication on public displays. GTmoPass was found to be highly resilient to shoulder surfing, thermal attacks and smudge attacks, thereby offering a secure solution to an important problem on public displays.

The second part of the dissertation explores specific challenges of gaze-based interaction with public displays. First, we address the user positioning problem by means of active eye tracking. More specifically, we built a novel prototype, Eye-

---

Scout, that dynamically moves the eye tracker based on the user's position without augmenting the user. This, in turn, allowed us to study and understand gaze-based interaction with public displays while walking, and when approaching the display from different positions. An evaluation revealed that EyeScout is well perceived by users, and improves the time needed to initiate gaze interaction by 62% compared to state-of-the-art. Second, we propose a system, Read2Calibrate, for calibrating eye trackers implicitly while users read text on displays. We found that although text-based calibration is less accurate than traditional methods, it integrates smoothly while reading and thereby more suitable for public displays. Finally, through our prototype system, EyeVote, we show how to allow users to select textual options on public displays via gaze without calibration. In a field deployment of EyeVote, we studied the trade-off between accuracy and selection speed when using calibration-free selection techniques. We found that users of public displays value faster interactions over accurate ones, and are willing to correct system errors in case of inaccuracies.

We conclude by discussing the implications of our findings on the design of gaze-based interaction for public displays, and how our work can be adapted for other domains apart from public displays, such as on handheld mobile devices.

## ZUSAMMENFASSUNG

In den letzten zehn Jahren wurden vermehrt interaktive Displays in öffentlichen Bereichen wie Einkaufszentren, Flughäfen und Bahnhöfen eingesetzt. Große öffentliche Displays finden sich zunehmend in städtischen Gebieten, beispielsweise in Universitäten und Bibliotheken. Fortschritte in der Eye-Tracking-Technologie und der Bildverarbeitung versprechen eine einfache Integration von Eye-Tracking auf diesen Displays. So kann zum einen das visuelle Verhalten der Benutzer verfolgt und damit ein Display nach verschiedenen Aspekten evaluiert werden. Zum anderen eröffnet Eye-Tracking auf öffentlichen Displays neue Interaktionsmöglichkeiten. Blickbasierte Interaktion ist besonders nützlich für Bildschirme im allgegenwärtigen öffentlichen Raum. Der Blick bietet mehr als eine natürliche und intuitive Interaktionsmethode: Blicke können aus der Ferne erkannt und somit für Interaktion mit sonst unerreichbaren Displays genutzt werden. Aus der Interaktion mit dem Blick (Gaze) lassen sich Absichten und visuelle Interessen der Benutzer ableiten. Dadurch eignet es sich besonders für den öffentlichen Raum, wo Nutzer möglicherweise Datenschutzbedenken haben könnten oder sich bei herkömmlichen Methoden gehemmt fühlen würden in der Öffentlichkeit mit den Displays zu interagieren. Dadurch wird ein uneingeschränktes Nutzererlebnis ermöglicht.

Eye-Tracking-Technologien sind jedoch in erster Linie für Desktop-Szenarien entwickelt worden, bei denen ein Benutzer für eine relativ lange Zeitspanne in einer stationären Position mit dem System interagiert. Die Interaktion mit öffentlichen Displays ist jedoch grundlegend anders. Daher gilt es völlig neuartige Herausforderungen zu bewältigen, wenn Eye-Tracking eingesetzt wird. Da sich Nutzer von öffentlichen Displays bewegen, können sie sich dem Display aus verschiedenen Richtungen nähern und sogar währenddessen mit dem Display interagieren. Folglich sollten "Gaze-enabled Displays" nicht davon ausgehen, dass Nutzer sich stets an einer bestimmten Position befinden, sondern sollten sich an die ständig wechselnde Position des Nutzers anpassen können. Zum anderen interagieren Nutzer von öffentlichen Displays üblicherweise nur für eine kurze Zeitspanne von ein paar Sekunden. Eine zeitaufwändige Kalibrierung durch den Nutzer vor der eigentlichen Interaktion ist hier im Gegensatz zu Desktop-Szenarien also nicht adäquat.

Diese kumulative Dissertation überprüft zunächst die Herausforderungen interaktiver öffentlicher Displays und diskutiert das Potenzial von blickbasierter Interaktion zu deren Bewältigung. Anschließend wird die Implementierung und eingehende Evaluierung von zwei beispielhaften Anwendungen vorgestellt, bei denen Nutzer durch den Blick mit öffentlichen Displays interagieren. Daraus ergeben sich weitere greifbare Vorteile der blickbasierten Interaktion für öffentliche Display-Kontexte. Bei der ersten Anwendung, EyePACT, steht der Parallaxeneffekt im Fokus, der heutzutage häufig ein Problem auf öffentlichen Displays darstellt, die über Berührung (Touch) gesteuert werden. Die zweite Anwendung ist ein neuartiges multimodales System, GTmoPass, das Gaze- und Touch-Eingabe zur sicheren Benutzerauthentifi-

---

zierung auf öffentlichen Displays kombiniert. GTmoPass ist sehr widerstandsfähig sowohl gegenüber unerwünschten fremden Blicken als auch gegenüber sogenannten thermischen Angriffen und Schmierangriffen. Es bietet damit eine sichere Lösung für ein wichtiges Sicherheits- und Datenschutzproblem auf öffentlichen Displays.

Der zweite Teil der Dissertation befasst sich mit spezifischen Herausforderungen der Gaze-Interaktion mit öffentlichen Displays. Zuerst wird der Aspekt der Benutzerpositionierung durch aktives Eye-Tracking adressiert. Der neuartige Prototyp EyeScout bewegt den Eye-Tracker passend zur Position des Nutzers, ohne dass dieser dafür mit weiteren Geräten oder Sensoren ausgestattet werden muss. Dies ermöglicht blickbasierte Interaktion mit öffentlichen Displays auch in jenen Situationen zu untersuchen und zu verstehen, in denen Nutzer in Bewegung sind und sich dem Display von verschiedenen Positionen aus nähern. Zweitens wird das System Read2Calibrate präsentiert, das Eye-Tracker implizit kalibriert, während Nutzer Texte auf Displays lesen. Der Prototyp EyeVote zeigt, wie man die Auswahl von Textantworten auf öffentlichen Displays per Blick ohne Kalibrierung ermöglichen kann. In einer Feldstudie mit EyeVote wird der Kompromiss zwischen Genauigkeit und Auswahlgeschwindigkeit unter der Verwendung kalibrierungsfreier Auswahltechniken untersucht. Die Implikationen der Ergebnisse für das Design von blickbasierter Interaktion öffentlicher Displays werden diskutiert. Abschließend wird erörtert wie die verwendete Methodik auf andere Bereiche, z.B. auf mobile Geräte, angewendet werden kann.



## ACKNOWLEDGMENTS

There are many people that I would like to thank for their support in the last 4 years.

I would like to thank my supervisors **Florian Alt**, and **Andreas Bulling**, whom I was very fortunate to work under their supervision. I consider myself very lucky for having had the chance to work with both of you. Although each of you has a unique approach in research, both of you are exemplary researchers, mentors and supervisors – you both complemented each other and this made the co-supervision flawless. I would like to thank you for your continuous support on personal and professional levels. You did not only support, guide, and refine my work, but you also significantly helped in shaping me as an academic and helped me make the most out of those 4 years. The three of us make a great team! I would also like to thank **Marc Langheinrich** for responding immediately and positively to my request for being my external examiner, and traveling all the way for my defense.

I am very grateful for the help I got from all the professors in the Chair for Media Informatics. Thank you **Heinrich Hussmann**, **Andreas Butz**, and **Albrecht Schmidt** for your continuous support ever since I have known you. You always provided constructive and helpful feedback in the doctoral colloquiums, and helped me in a lot of situations where I needed expert opinions from senior academics.

I would like to thank **Rainer Fink**, who ensures that all technicalities run smoothly, and **Franziska Schwamb** for her help with paperwork and bureaucratic hurdles.

I have made a lot of friends during my time at the LMU. Thank you **Malin Eiband** for being such a nice, truthful and supporting friend. Our “shoulder surfing” project remains one of my most pleasant collaboration experiences. Thank you for introducing me to my *favorite bottle*. Thank you **Mariam Hassib** for being such a thoughtful friend and helping out in all sorts of matters, ranging from introducing me to Florian 4 years ago, to offering me to stay over until I find an apartment in Munich. I have always enjoyed our collaborations. I wish you, Amr and your little one all the best, and I am happy to say “dude, we are no longer screwed”. Thank you **Nada Terzimehić** for being a great friend. It was always very nice to chat, enjoying Sababa’s Falafel together, and it was a pleasure to build Lunchocracy together. Thanks for all the coffee and food you offered me! may you have a lot of “Yummy!” moments. Thanks **Sarah Theres Völkel** for all the inspiration you spread with your activeness and engagement. I am sure you will do very well in your PhD – you are and you will continue to be a very successful person. I would like to thank **Danniel Buschek** for all the interesting discussions and successful collaborations. It is always nice to work with you. Thank you **Christina Schneegass** and **Pino Colada** for spreading happiness in the chair. And thank you Christina for organizing the first German HCI booklet and the awesome CHI party. Thanks **Sarah Aragon Bartsch** for all the interesting discussions about languages, and for bravely embracing ERASMUS. Thank you **Ceenu George** for the nice collaborations, it was

---

a pleasure to work with you. Let's hope we do make it to Kazakhstan one day. Thanks **Christian Mai** for the many recent collaborations and taking care of all the VR stuff. Thank you **Renate Häuslschmid** for being an awesome office mate and for the many moments of laughter watching funny videos. Thank you **Hanna Schneider** for being a cool office mate, and an awesome travel mate in Heidelberg. Thank you **Tobias Seitz** for being such a cool workmate, and for the opportunity to work on PasswordBlur together. Thank you **Axel Hösl** for your great help with EyeScout, it was nice to work with you. Thank you **Thomas Kosch, Jakob Karolus, Matthias Hoppe**, and **Pascal Knierim** for all the nerf gun fights and the cool moments, as well as the interesting crazy researchie discussions. Thank you **Sylvia Rothe** for the collaborations, it was a pleasure to work with you. Thanks **Maria Fysaraki** for bringing all the Greek food and sweets. Thank you **Emanuel von Zezschwitz** for the fruitful discussions and successful collaborations. It is always great to collaborate with you and I have learned a lot from you. Thank you **Bastian Pfleging** for all the help with finding the best deals and offers. Thank you **Daniel Ullrich** for being my movie nerd companion, and the awesome ice coffee. Thanks **Alexander Wiethoff** for the helpful feedback on my first IDC. Thank you **Heiko Drewes** for the nice collaborations and interesting discussions, it was a pleasure to work with you. Thank you to all the external PhD students: **Christian Lachner, Veronika Gamber, David Englmeier, Gesa Wiegand**, and **Michael Braun**. Thank you and best of luck to the new PhD students, **Kai Holländer, Sarah Prange, Lukas Mecke, Beat Rossmly**. Finally I'd like to thank them all again for the amazing graduation hat!

I extend my thanks to all external collaborators **Stefan Schneegass, Niels Henze, Sven Mayer, Huy Viet Li, Yomna Abdelrahman, Menna Essam, Christian Sturm, Ville Mäkelä, Diako Mardanbegi, Vito Gentile, Salvatore Sorce**, and **Hans Gellersen**. I also thank my students whom I had the privilege to work with.

I would like to thank my friends who were very close to me and supportive during the last 4 years. Thank you to the "Gang": **Mostafa Ismail, Mohamed Anwar, Lojayna Sherif**. Thank you **Passant El Agroudy**, and all my friends in Kaiserslautern especially: **Yasmina Adams, Mohamed Selim**, and **Bahaa Beih**.

The guidance and support of my academic mentors **Slim Abdennadher, Christian Sturm**, and **Fatma Meawad** extends beyond these 4 years. Thank you.

I cannot find enough words to thank my mother **Eiman Hafez**, sisters **Engy** and **Lamiaa**, and father **Mohsen Khamis**. Thank you all for being the best family ever, for your love, and unconditional support throughout my life, as well as encouraging and motivating me to strive for the best.

Last but most definitely not least, I would never be able to thank my wife, **Maha Badreldin**, enough for her support, understanding, and encouragement. Thank you for your support, love, patience, and encouragement, without which I would have neither got this degree, nor the lectureship position in the University of Glasgow. Thank you for standing by me.

# TABLE OF CONTENTS

<b>List of Figures</b>	<b>xv</b>
<b>Publications and Declaration of Co-Authorship</b>	<b>1</b>
<b>I INTRODUCTION AND BACKGROUND</b>	<b>5</b>
<b>1 Introduction</b>	<b>7</b>
1.1 Research Objectives . . . . .	9
1.2 Research Questions . . . . .	9
1.3 Summary of Contributions . . . . .	10
1.3.1 Practical Contributions . . . . .	10
1.3.2 Theoretical Contributions . . . . .	11
1.3.3 Methodological Contributions . . . . .	11
1.4 Dissertation Overview . . . . .	12
1.5 Full Publications List . . . . .	14
<b>2 Background and Related Work</b>	<b>21</b>
2.1 The Advent of Pervasive Public Displays . . . . .	22
2.1.1 Uniqueness of Public Displays . . . . .	22
2.1.2 Interacting with Public Displays . . . . .	23
2.1.2.1 Touch . . . . .	23
2.1.2.2 Mid-air Gestures . . . . .	23
2.1.2.3 Mobile Device Interaction . . . . .	24
2.1.2.4 Gaze . . . . .	24
2.2 The Eyes in Human-Computer Interaction . . . . .	24
2.2.1 Opportunities of Eye Tracking . . . . .	25
2.2.2 General Limitations of Eye Tracking . . . . .	25
2.3 Previous Work on Gaze-enabled Displays . . . . .	26
2.3.1 Attention Monitoring on Public Displays . . . . .	26

---

2.3.1.1	Methods to Monitor Attention . . . . .	26
2.3.1.2	Attention Monitoring in Practice . . . . .	27
2.3.2	Implicit Gaze Interaction with Public Displays . . . . .	27
2.3.2.1	Eye-Contact Aware Implicit Interaction . . . . .	28
2.3.2.2	Gaze-Behavior Aware Implicit Interaction . . . . .	28
2.3.3	Explicit Gaze Interaction with Public Displays . . . . .	28
2.3.3.1	Gaze-only Interaction . . . . .	29
2.3.3.2	Gaze-supported (Multimodal) Interaction . . . . .	29
2.4	Trends and Limitations of Previous Work . . . . .	29
<b>II</b>	<b>SUMMARY OF PEER-REVIEWED PUBLICATIONS</b>	<b>31</b>
<b>3</b>	<b>Opportunities of Gaze on Public Displays</b>	<b>33</b>
3.1	EyePACT: Implicit Eye-based Parallax Correction . . . . .	34
3.2	GTmoPass: Explicit Gaze-supported Authentication . . . . .	35
<b>4</b>	<b>Challenges of Gaze-enabled Public Displays</b>	<b>37</b>
4.1	EyeScout: Movement and Position Independent Gaze Interaction . . .	38
4.2	TextPursuits: Implicit Calibration and Calibration-free Gaze Interaction	39
4.3	EyeVote in the Wild: Understanding the Trade-off between Accuracy and User Experience . . . . .	40
<b>III</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>43</b>
<b>5</b>	<b>Closing Remarks</b>	<b>45</b>
5.1	Design Recommendations . . . . .	46
5.2	Future Work . . . . .	46
5.2.1	Our Methodology Beyond Public Displays . . . . .	47
5.2.2	There is More to Gaze Interaction than Accuracy . . . . .	47
5.2.3	Privacy Implications of Gaze-enabled Public Displays . . . . .	48

<b>IV</b>	<b>OPPORTUNITIES OF GAZE-ENABLED PUBLIC DISPLAYS</b>	<b>49</b>
<b>6</b>	<b>Tackling Challenges of Interactive Public Displays using Gaze</b>	<b>51</b>
<b>7</b>	<b>EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays.</b>	<b>57</b>
<b>8</b>	<b>GTmoPass: Two-factor Authentication on Public Displays Using Gaze-Touch passwords and Personal Mobile Devices.</b>	<b>77</b>
<b>V</b>	<b>CHALLENGES OF GAZE-ENABLED PUBLIC DISPLAYS</b>	<b>87</b>
<b>9</b>	<b>EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.</b>	<b>89</b>
<b>10</b>	<b>TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.</b>	<b>103</b>
<b>11</b>	<b>EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays?</b>	<b>117</b>
	<b>Bibliography</b>	<b>125</b>
	<b>Curriculum Vitae</b>	<b>143</b>



# LIST OF FIGURES

1.1	Our Methodology . . . . .	12
3.1	EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays . . . . .	34
3.2	GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices. . . . .	35
4.1	EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. . . . .	38
4.2	TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays . . . . .	39
4.3	EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays? . . . . .	41





# PUBLICATIONS AND DECLARATION OF CO-AUTHORSHIP

All the following papers were peer-reviewed.

**Chapter 6** corresponds to the following workshop paper:

Mohamed Khamis, Andreas Bulling, and Florian Alt. 2015. **Tackling challenges of interactive public displays using gaze**. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15 Adjunct). Pages 763-766.

This paper is based on my initial PhD research proposal. The work was discussed with my supervisors: Florian Alt and Andreas Bulling. However both the initial and camera ready versions of the paper were written by me.

**Chapter 7** corresponds to the following journal paper:

Mohamed Khamis, Daniel Buschek, Tobias Thieron, Florian Alt, Andreas Bulling. **EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays**. In Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (IMWUT) 1, 4, Article 146 (January 2018), 18 pages.

I conceived the idea and the original research contributions of EyePACT after noticing the issue on a public display in Münchner Freiheit, in Munich, Germany. The concept was discussed with my supervisors: Florian Alt and Andreas Bulling. Daniel Buschek assisted with the regression analysis reported in section 5 of the paper ( $\approx$  half a page). Tobias Thieron assisted in the implementation reported in the second paragraph of section 3 ( $\approx$  5 lines), which was done as part of his master thesis under my supervision. He conducted the study, which I designed and supervised. I conducted the analysis, reporting and interpretation of the results. I wrote the initial draft of the article and did the subsequent corrections after feedback from my supervisors: Florian Alt and Andreas Bulling.

**Chapter 8** corresponds to the following conference paper:

Mohamed Khamis, Regina Hasholzner, Andreas Bulling, Florian Alt. **GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices**. In Proceedings of the 6th International Symposium on Pervasive Displays (PerDis 2017).

---

I received the Google IoT research award based on the idea of GTmoPass. I conceived the idea as a followup project on my Late-Breaking-Work article at CHI 2016. I carried out the full implementation of GTmoPass. Regina Hasholzner conducted the user study using my implementation and reported on the initial analysis as part of her internship at the chair of media informatics. I developed the study's design and procedure, which Regina carried out under my close supervision. I significantly extended the analysis of the results, and performed the interpretation and reporting of the results. The work was discussed with my supervisors: Florian Alt and Andreas Bulling.

**Chapter 9** corresponds to the following conference paper:

Mohamed Khamis, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, Andreas Bulling. **EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.** In Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology (UIST 2017).

I conceived the idea and the original research contributions of EyeScout. Axel Hösl and I co-supervised the master theses of Martin Reiss and Alexander Klimczak. They both built the initial prototype under my close supervision. They also conducted the study, whose design and procedure was developed by me. I wrote the initial draft and performed the analysis, interpretation and reporting of the results. The work was regularly discussed with Florian Alt, Andreas Bulling, and Axel Hösl. The three revised the first version of the article, and Axel produced figure 1. I did the subsequent corrections.

**Chapter 10** corresponds to the following conference paper:

Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, Florian Alt. **TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.** In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016).

I conceived the idea and original research contributions of TextPursuits, which constitutes two systems: EyeVote and Read2Calibrate. The idea of EyeVote was described in my initial PhD proposal. The implementation was partially realized by Katharina Stolz as part of her bachelor thesis. I conducted the analysis, interpretation and reporting of the results. The second system, Read2Calibrate, was implemented by Ozan Saltuk as part of his master thesis, who also wrote part of the implementation subsection ( $\approx 1/6$  of page 6). I closely supervised both theses. Alina Hang co-supervised Ozan's thesis, and suggested improvements to the first draft of the paper. I wrote the first draft, and made the majority of the subsequent corrections. My supervisors, Florian Alt and Andreas Bulling, helped me in improving the draft.

**Chapter 11** corresponds to the following conference paper:

Mohamed Khamis, Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, Florian Alt. **EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays?** In Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM 2016).

I conceived the idea and original research contributions of this work, which was a follow up on the TextPursuits paper (Chapter 10). I developed the study's design and procedure. The study was then conducted by Ludwig Trotter, Markus Tessmann, and Christina Dannhart in the context of the "Advanced Topics in HCI" module<sup>1</sup>. The students customized the existing implementation of EyeVote for this work's user study. I closely supervised the deployment of the field study. The work was discussed with my supervisors: Florian Alt and Andreas Bulling. Ludwig Trotter wrote the first draft of the implementation and results sections. I significantly revised the sections, wrote the first draft of the entire paper, and made the vast majority of the subsequent corrections. Florian Alt assisted me in improving the framing of the story.

Munich, 14.05.2018

---

Mohamed Khamis

---

<sup>1</sup> <http://www.medien.ifi.lmu.de/lehre/ss16/ath/>



# I

## INTRODUCTION AND BACKGROUND



# 1

## Introduction

---

Many aspects of Mark Weiser’s vision for computers in the 21st century [124] are already part of our daily lives. Computing devices are seamlessly integrated into environments around us. These devices come in many forms, including pervasive public displays. In fact, our present day public displays resemble what Weiser and his colleagues referred to as boards – one of the key components of ubiquitous computing. Public interactive displays, and in particular large ones, are increasingly becoming ubiquitous. They can now be found deployed in public areas, such as train stations, airports, malls, supermarkets, and universities. They bring a lot of tangible benefits to passersby. For example, they are used to deliver information (e.g., train schedules and news) and personalized content [21], allow purchase of tickets and services, encourage civic engagement [114], and nudge behavior change [94]. These applications are largely driven by advances in sensing technologies, which in turn enable a myriad of modalities to empower public displays. One particularly promising modality that is increasingly gaining attention is eye gaze.

Being a subtle, natural and intuitive modality, the use of eye gaze for interaction aligns nicely with the Mark Weiser’s vision of pervasive technologies that can weave themselves into everyday life. Humans use their eyes intuitively to express visual interests, reveal their intentions, and regulate conversations. In addition to being valuable to human-computer interaction in general, eye gaze is exceptionally attractive for pervasive public displays. Being fast [100], gaze has a lot of potential in delivering “immediate usability” – a requirement of pervasive public displays [81]. Gaze can be used from a distance, hence supplying users with a contact-free and hygienic method for interacting in public, even if the display is physically unreachable (e.g., behind glass window or above user height [20]), or if the user is wearing gloves and hence cannot interact via touch.

Acknowledging the potential of gaze on public displays, researchers proposed novel methods that leverage gaze to improve the user’s experience [104, 105, 118, 128, 131]. While these works presented a leap towards gaze-enabled public displays, there are several gaps in previous work in this area. First, prior work mainly focused on introducing gaze interaction techniques, or improving gaze estimation accuracy. That being said, one gap is that there is a lack of gaze-based solutions to problems that are specific to public displays. A second gap is that to date, eye tracking technologies and techniques that are intended for desktop settings are still being used for public displays. As a result, there is a need to adapt eye tracking to the many unique aspects of public displays, which are under-investigated in the context of gaze-based interaction. Examples of said aspects are the facts that opposed to desktop settings, users of public displays could be interacting from different positions and even while moving, and that public displays cannot afford requiring time-consuming eye tracker calibration. There is also a lack of understanding of the needs of users of gaze-enabled public displays; e.g., how far are users striving for quick interactions? and how can we design systems that optimize interaction speed at the expense of accuracy, yet still provide a pleasant user experience?



## 1.1 Research Objectives

To address these gaps, this dissertation reports on multiple novel concepts (2 methods, and 3 systems), and results from a series of lab studies (N=133) and field studies (N=49–106). More specifically, this dissertation closes the aforementioned gaps by addressing the following research objectives:

1. Understanding the potential of eye gaze in the context of pervasive public displays (Chapter 6).
2. Designing, implementing, and evaluating gaze-based interfaces that address key challenges of pervasive public displays (Chapters 7 and 8).
3. Exploring and addressing the unique aspects and challenges arising from integrating gaze with pervasive public displays (Chapters 9, 10, and 11).

To fulfill Objective 1, we first report on a review of challenges of pervasive public displays that can be addressed by leveraging the passerby's eye gaze (Chapter 6).

As for Objective 2, we report on the design, implementation, and evaluation of two systems that employ gaze to address core challenges on pervasive public displays. Namely, in Chapter 7 we introduce the concept, implementation and evaluation of EyePACT, an effective approach for mitigating the parallax effect on public displays, which is a fundamental problem with many of today's public displays. Second, in Chapter 8 we introduce the concept, implementation and evaluation of a novel approach for secure user authentication on public displays.

Finally, we address Objective 3 by studying two unique properties of gaze interfaces on pervasive public displays: eye tracking on the move, and calibration-free gaze interaction. We do so by introducing novel systems, and evaluating them in lab and field studies. More specifically, we introduce EyeScout in Chapter 9, the first system to leverage active eye tracking to expand the areas from which gaze interaction with large displays is possible, while significantly decreasing the time required to initiate gaze interaction. Second, in Chapter 10, we show how reading text on public displays can be used to enable calibration-free gaze interaction, or to calibrate eye trackers implicitly. Finally in Chapter 11, we report on a field study that shows that passersby would rather use gaze interaction techniques that are fast even if they were inaccurate, as long as usable input correction mechanisms are provided.

## 1.2 Research Questions

By addressing the aforementioned objectives, we answer the following research questions:

Research Question	Chapter
<b>I. Opportunities of Gaze-enabled Public Displays</b>	
What are the problems of today's pervasive public displays that eye gaze can solve?	Chapter 6
How can eye gaze be leveraged for parallax correction on public displays?	Chapter 7
How can eye gaze improve the security of public displays?	Chapter 8
<b>II. Challenges of Gaze-enabled Public Displays</b>	
What are challenges of integrating gaze-based interaction into public displays?	Chapters 9, 10, & 11
How can we design gaze-enabled displays that adapt to the dynamic behavior of passersby?	Chapter 9
How can we design accurate and immediately usable gaze-based interaction on public displays despite the need for calibration?	Chapter 10
How can we balance the trade-off between speed and accuracy of gaze-based interaction on public displays?	Chapter 11

**Table 1.1:** Summary of Research Questions

## 1.3 Summary of Contributions

This dissertation makes practical, methodological, and theoretical contributions.

### 1.3.1 Practical Contributions

This dissertation provides a number of practical contributions for improving the design of gaze-enabled public displays. In this work, we have designed various approaches that leverage gaze to address challenges on today's public displays, such as parallax-free touch interaction and secure authentication. We have also presented the design, implementation and evaluation of forward-looking systems that facilitate gaze interaction on the move, implicit eye tracker calibration, and calibration-free interaction. The proposed concepts are novel yet applicable on today's public displays, and can be key to enabling wide-scale adoption of gaze on public displays. We conclude with recommendations for designing gaze-based interaction for public displays (Section 5.1)

### 1.3.2 Theoretical Contributions

This dissertation contributes a body of knowledge that augments our understanding of gaze-based interaction with public displays. First, our work has identified **opportunities** brought forth by integrating gaze into pervasive public displays. Second, we explored core **challenges** that are unique to gaze-enabled public display settings. This dissertation proposed systems that leverage said opportunities, and proposed systems and studies that address and deepen our understanding of **challenges** of gaze-enabled public displays.

At the same time, this work lays a foundation on which future studies and systems can build on: 1) many of the identified **opportunities** are waiting to be explored and leveraged by researchers and practitioners. 2) Our exploration of unique aspects of public displays has set the scene for the upcoming research **challenges** for gaze-based interaction on public displays.

### 1.3.3 Methodological Contributions

In addition, this dissertation also makes research methodology contributions. Namely, we propose a methodology for studying gaze-based interaction while moving (e.g., walking across large displays, or approaching them from random positions). Work in gaze-based interaction has mostly considered a stationary user – there had been little to no experimental research on gaze interaction while on the move. Our methodology contribution is demonstrated in Chapter 9, where we designed the first user study in which users perform gaze-based selections while walking. This study design can be used as a basis for follow up work on gaze-based interaction while moving. This is not only relevant to public displays, but also for gaze interaction with mobile, wearable, and head mounted devices. For example, our methodology was adopted in follow up work on gaze-based interaction while walking in virtual environments [56].

Another core contribution is the overarching methodology adopted in this work to explore and understand the use of eye gaze in the domain of pervasive public displays (Figure 1.1). The methodology adopted in this dissertation is as follows: The first step is to identify challenges users face in this domain – in Chapter 6, we do that for pervasive public displays. The second step is to study how gaze can address core challenges in this domain – in Chapters 7 and 8, we report on methods for addressing core problems on public displays using gaze. The third step is to understand the novel challenges arising from integrating gaze in this domain. The fourth step is to design, implement, and evaluate solutions to address these challenges. We identify a novel challenge of using gaze in this domain in Chapters 9, 10 and 11, and then we design, implement, and evaluate solutions to address them.



**Figure 1.1:** Our methodology is as follows: First, we identify the challenges in the field of pervasive displays. Second, we study how gaze can address these challenges. Third, we investigate the novel challenges arising from the use of gaze in this domain. Finally, we design, implement, and evaluate solutions to address these challenges. This methodology can be adopted for exploring gaze-based interaction on other platforms, such as handheld mobile devices.

In section 5.2.1, we discuss how this methodology can be transferred to other domains, such as mobile devices, mixed reality, and internet of things.

## 1.4 Dissertation Overview

This cumulative dissertation comprises 6 peer-reviewed scientific publications, and is complemented with the following chapters:

- **Chapter 2: Background and Related Work** This chapter summarizes the relevant background in both pervasive public displays, and eye tracking. It also discussed previous work in gaze-enabled public displays, and highlights its trends and limitations.
- **Chapter 3: Opportunities of Gaze on Public Displays** This chapter summarizes the peer-reviewed publications presented in Chapters 6, 7, and 8.
- **Chapter 4: Challenges of Gaze-enabled Public Displays** This chapter summarizes the peer-reviewed publications presented in Chapters 9, 10, and 11.
- **Chapter 5: Closing Remarks:** This chapter concludes and summarizes the outcomes of this work, presents design recommendations for gaze-enabled public displays, and highlights directions for future research.
- **Chapters 6–11:** each of those chapters corresponds to the a peer-reviewed publication as shown in Table 1.2 on the following page.

Chapter	Publication
6	Tackling Challenges of Interactive Public Displays using Gaze. <b>Mohamed Khamis</b> , Andreas Bulling, and Florian Alt In Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing ( <b>UbiComp 2015</b> ).
7	EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays. <b>Mohamed Khamis</b> , Daniel Buschek, Tobias Thieron, Florian Alt, and Andreas Bulling In Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. ( <b>IMWUT</b> ) 1, 4, Article 146 (December 2017), 18 pages.
8	GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices. <b>Mohamed Khamis</b> , Regina Hasholzner, Andreas Bulling, and Florian Alt. In Proceedings of the 6th International Symposium on Pervasive Displays ( <b>PerDis 2017</b> ).
9	EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. <b>Mohamed Khamis</b> , Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology ( <b>UIST 2017</b> ).
10	TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. <b>Mohamed Khamis</b> , Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt In Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing ( <b>UbiComp 2016</b> ).
11	EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays? <b>Mohamed Khamis</b> , Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, and Florian Alt. In Proc. of the 15th International Conference on Mobile and Ubiquitous Multimedia ( <b>MUM 2016</b> ).

**Table 1.2:** The dissertation comprises 6-peer reviewed papers published as a journal paper at IMWUT, conference papers at UIST 2017, PerDis 2017, UbiComp 2016, MUM 2016, and a workshop paper at UbiComp 2015.

---

## 1.5 Full Publications List

Apart from the papers included in this dissertation, below is a full list of all 38 publications published in my academic career.

### Journal papers (4)

41. **Mohamed Khamis**, Malin Eiband, Martin Zürn, Heinrich Hussmann. EyeSpot: Leveraging Gaze to Protect Private Text Content on Mobile Devices from Shoulder Surfing. *Journal of Multimodal Technologies and Interaction* 2018, 2(3), 45 DOI: <https://doi.org/10.3390/mti2030045>
40. **Mohamed Khamis**, Daniel Buschek, Tobias Thieron, Florian Alt, Andreas Bulling. EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **IMWUT 1, 4**, (December 2017). DOI: <https://doi.org/10.1145/3161168>
39. **Mohamed Khamis**, J. Henderson, Guiying Du. PerDis 2017. IEEE Pervasive Computing (Volume: 16, Issue: 4, October-December 2017). DOI: <https://doi.org/10.1109/MPRV.2017.3971126>
38. Dominic Gorecky, **Mohamed Khamis**, Katharina Mura. Introduction and establishment of virtual training in the factory of the future. *International Journal of Computer Integrated Manufacturing.* 30, 1, (May 2015), 182 – 190. DOI: <http://dx.doi.org/10.1080/0951192X.2015.1067918>

### Conference papers (21)

2018

37. **Mohamed Khamis**, Florian Alt, Andreas Bulling. The Past, Present, and Future of Gaze-enabled Handheld Mobile Devices: Survey and Lessons Learned. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (**MobileHCI 2018**). ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3229434.3229452>
36. Thomas Mattusch, Mahsa Mirzamohammad, **Mohamed Khamis**, Andreas Bulling, Florian Alt. Hidden Pursuits: Evaluating Gaze-selection via Pursuits when the Stimulus Trajectory is Partially Hidden. In Proceedings of the ACM International Symposium on Eye Tracking Research & Applications (**ETRA 2018**). ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3204493.3204569>
35. Christian Mai, **Mohamed Khamis**. Public HMDs: Modeling and Understanding User Behavior around Public Head-Mounted Displays. In

- Proceedings of the 7th International Symposium on Pervasive Displays (**PerDis 2018**). ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3205873.3205879>
34. **Mohamed Khamis**, Carl Oechsner, Florian Alt, Andreas Bulling. 2018. VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '18)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3206505.3206522>
  33. **Mohamed Khamis**, Christian Becker, Andreas Bulling, Florian Alt. 2018. Which one is me? Identifying Oneself on Public Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA. **HONORABLE MENTION AWARD** DOI: <https://doi.org/10.1145/3173574.3173861>
  32. **Mohamed Khamis**, Anita Baier, Niels Henze, Florian Alt, Andreas Bulling. Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones used in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3173574.3173854>
  31. Ville Mäkelä, **Mohamed Khamis**, Lukas Mecke, Jobin James, Markku Turunen, Florian Alt. Pocket Transfers: Interaction Techniques for Transferring Content from Situated Displays to Mobile Devices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA. DOI: <http://dx.doi.org/10.1145/3173574.3173709>
- 2017
30. **Mohamed Khamis**, Linda Bandelow, Stina Schick, Dario Casadevall, Andreas Bulling and Florian Alt. 2017. They are all after you: Investigating the Viability of a Threat Model that involves Multiple Shoulder Surfers. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3152832.3152851> **HONORABLE MENTION AWARD**
  29. Mariam Hassib, **Mohamed Khamis**, Susanne Friedl, Stefan Schneegass and Florian Alt. 2017. BrainAtWork: Logging Cognitive Engagement and Tasks in the Workplace using Electroencephalography. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3152832.3152865>
  28. **Mohamed Khamis**, Mariam Hassib, Emanuel von Zezschwitz, Andreas Bulling, Florian Alt. 2017. GazeTouchPIN: Protecting Sensitive Data

- 
- on Mobile Devices using Secure Multimodal Authentication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3136755.3136809>
27. **Mohamed Khamis**, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, Andreas Bulling. 2017. EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *Proceedings of the 30th Annual Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, USA. DOI: <https://doi.org/10.1145/3126594.3126630>
  26. **Mohamed Khamis**, Regina Hasholzner, Andreas Bulling, and Florian Alt. 2017. GTmoPass: two-factor authentication on public displays using gaze-touch passwords and personal mobile devices. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*. ACM, New York, NY, USA, Article 8. DOI: <https://doi.org/10.1145/3078810.3078815>
  25. Vito Gentile, **Mohamed Khamis**, Salvatore Sorce, and Florian Alt. 2017. They are looking at me!: understanding how audience presence impacts on public display users. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*. ACM, New York, NY, USA, Article 11. DOI: <https://doi.org/10.1145/3078810.3078822>
  24. Yomna Abdelrahman, **Mohamed Khamis**, Stefan Schneeeggass, and Florian Alt. 2017. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3751-3763. DOI: <https://doi.org/10.1145/3025453.3025461>  
**HONORABLE MENTION AWARD**
  23. Malin Eiband, **Mohamed Khamis**, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4254-4265. DOI: <https://doi.org/10.1145/3025453.3025636>
  22. Ceenu George, **Mohamed Khamis**, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt Florian Alt, Heinrich Hussmann. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *USEC conference in conjunction with (NDSS '17)*. 2016
  20. **Mohamed Khamis**, Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, and Florian Alt. 2016. EyeVote in the wild:



do users bother correcting system errors on public displays?. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 57-62. DOI: <https://doi.org/10.1145/3012709.3012743>

19. **Mohamed Khamis**, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: using text for pursuits-based interaction and calibration on public displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 274-285. DOI: <https://doi.org/10.1145/2971648.2971679>

2015

18. Omar Mowafi, **Mohamed Khamis** and Wael Abouelsaadat. AirDisplay: Experimenting with Air Flow as a Communication Medium. 2015. In *15th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT '15)*. Springer International Publishing, 316-323. DOI: [https://doi.org/10.1007/978-3-319-22701-6\\_23](https://doi.org/10.1007/978-3-319-22701-6_23)

2013

17. Jörn Hees, **Mohamed Khamis**, Ralf Biedert, Slim Abdennadher and Andreas Dengel. Collecting Links between Entities Ranked by Human Association Strengths. In *The 10<sup>th</sup> Extended Semantic Web Conference (ESWC '13)*. Springer Berlin Heidelberg, 517-531. DOI: [https://doi.org/10.1007/978-3-642-38288-8\\_35](https://doi.org/10.1007/978-3-642-38288-8_35)

Technical reports, adjunct papers (16)

16. Heiko Drewes, **Mohamed Khamis**, Florian Alt. DialPlate: Enhancing the Detection of Smooth Pursuits Eye Movements Using Linear Regression. Technical report at arXiv: 1807.03713, 2018. DOI: <https://arxiv.org/pdf/1807.03713.pdf>
15. Ebtisam Alabdulqader, Shaimaa Lazem, **Mohamed Khamis**, Susan Dray. Exploring Participatory Design Methods to Engage with Arab Communities. In *Proc. of the 36th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2018)*, Montréal, QC, Canada 2018.
14. Mennatallah Saleh, **Mohamed Khamis**, Christian Sturm. Privacy Invasion Experiences and Perceptions: A comparison between Germany and the Arab World. In *Proc. of the 36th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2018)*, Montréal, QC, Canada 2018.

- 
13. Nađa Terzimehić, Mohamed Khamis, Florian Bemmman, Heinrich Hussmann. Lunchocracy: Improving Eating Dynamics in the Workplace Using a Bot-Based Anonymous Voting System. In *Proc. of the 36th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2018)*, Montréal, QC, Canada 2018.
  12. Christian Mai, Lukas Rambold, **Mohamed Khamis**. TransparentHMD: Revealing the HMD User's Face to Bystanders. In *Adj. Proc. of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM 2017)*, Stuttgart, Germany. 2017.
  11. Mohammed Al Sada, **Mohamed Khamis**, Akira Kato, Shigeki Sugano, Tatsuo Nakajima, Florian Alt. Challenges and Opportunities of Supernumerary Robotic Limbs. In *Proc. of the CHI 2017 Workshop on Amplification and Augmentation of Human Perception (Amplify 2017)*.
  10. Oliver Duerr, **Mohamed Khamis**, Daniel Buschek, Florian Alt. HelpMe: Assisting Older Adults in Performing Tasks on Mobile Devices. In *Proc. of the CHI 2017 Workshop on Designing Mobile Interactions for the Ageing Populations*.
  9. **Mohamed Khamis**, Florian Alt, Andreas Bulling. Challenges and Design Space of Gaze-enabled Public Displays. In *Adj. Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016)*, Heidelberg, Germany 2016.
  8. Diako Mardanbegi, Shahram Jalaliniya, **Mohamed Khamis**, Päivi Majaranta. The 6th International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction. In *Adj. Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016)*, Heidelberg, Germany 2016.
  7. **Mohamed Khamis**, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, Andreas Bulling. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proc. of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI '16)*, San Jose, CA, USA 2016.
  6. Mariam Hassib, **Mohamed Khamis**, Stefan Schneegass, Alireza Sahami Shirazi, Florian Alt. Investigating User Needs for Bio-sensing and Affective Wearables. In *Proc. of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI '16)*, San Jose, CA, USA 2016.
  5. **Mohamed Khamis**, Andreas Bulling, Florian Alt. TA Field Study on Spontaneous Gaze-based Interaction with a Public Display using Pursuits. In *In Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 865-874, 2015. (**UbiComp '15**), Osaka, Japan 2015.

4. **Mohamed Khamis**, Andreas Bulling, Florian Alt. Tackling Challenges of Interactive Public Displays using Gaze. In *In Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 763-766, 2015. (**UbiComp '15**), Osaka, Japan 2015.
3. Menna Bakry, **Mohamed Khamis** and Slim Abdennadher. Are-CAPTCHA: Outsourcing Arabic Text Digitization to Native Speakers. In *The 11<sup>th</sup> IAPR International Workshop on Document Analysis Systems (DAS '14)*, Tours, France, 2014.
2. Sara Nassar, Nada Sharaf, **Mohamed Khamis**, Slim Abdennadher and Caroline Sabty. Collecting Arabic Dialect Variations using Games With A Purpose: A Case Study Targeting the Egyptian Dialect. In *The 2<sup>nd</sup> Workshop on Games and NLP (GAMNLP '13)*, Istanbul, Turkey, 2013.
1. Christian Schneiker, **Mohamed Khamis**, Dietmar Seipel: Prolog Server Faces - A Declarative Framework for Dynamic Web Pages. In *The 24<sup>th</sup> Workshop on Constraint Logic Programming (WLP'10)*, Cairo, Egypt, 2010.



# 2

## Background and Related Work

Work on gaze-enabled public displays builds on two strands of work: (1) Pervasive Public Displays, and (2) Eye Tracking in HCI. While there are in depth reviews of prior work in public displays [20], and eye tracking [26, 27, 69], in this chapter we focus on the developments that led to the emergence of gaze-enabled public displays. We first provide an overview on the history of displays, and how they developed to become the public, pervasive, and interactive displays that we encounter nowadays. Second, we discuss how the tracking of eyes developed from a method to study and understand humans, to an effective modality that comes with challenges and opportunities for 1) attention monitoring, 2) implicit interaction, and 3) explicit interaction. Finally, we focus exclusively on previous work in gaze-enabled public displays, and conclude with trends and limitations in previous work.

---

## 2.1 The Advent of Pervasive Public Displays

For thousands of years, humans utilized different types of media to publicly display content. Drawings were painted on cave walls in prehistoric ages. Mesopotamians, Pharaohs, and Mesoamericans engraved text and drawings on walls, pillars, and temples. Ancient Egyptians made wall posters out of papyrus [12], and advertising plates that date back to ancient times were found in China [66]. Public displays in the form of signs were also used in the middle ages for advertising and making announcements. Public displays thrived in the 1800s, when they appeared in the form of painted billboards to advertise circuses and theatres [109].

Fast-forward to today, and we will find that digital displays have become ubiquitous in almost all public spaces. Industries saw the benefits of distributed networks of interactive displays, and rushed to deploy interactive displays as ticket vending machines in train and bus stations, interactive information boards in airports [8], and self check out terminals in supermarkets [77]. Universities and governments started to deploy displays in city centers [85].

HCI researchers took it upon themselves to explore novel uses of public displays. As early as 1980, Galloway and Rabinowitz introduced their prototype “Hole in Space” [32]. Hole in space was presented as an art installment, with the core idea of linking passersby around a display in New York with others around a display in Los Angeles. The opportunities arising from connecting distant places using interactive public displays is still pursued up until this decade [36]. Researchers explored other uses of public displays. For example, they were used for collecting votes in public with the aim to encourage civic discourse [103, 114]. Other displays aimed at promoting community interaction and place awareness [76], nudging behavior change [94], and visualizing urban information [17, 58]. Public displays were also found to be effective platforms for crowdsourcing [34] and serious games [48].

### 2.1.1 Uniqueness of Public Displays

Today’s interactive public displays correspond to what Weiser [124] described as boards. Indeed, the unique aspects of public displays justify the recognition of “boards” as a distinct computing device. Although public displays share some similar aspects with mobile devices and desktop computers, they are profoundly different in many ways. Researchers have studied the distinct audience behavior around public displays. For example, several works reported on the honeypot effect [14, 20], which refers to the influence that users, who are engaged with the display, induce on surrounding passersby. Another unique behavior is the landing effect [83], which describes cases in which a passerby realizes that a display is interactive after walking by the display, and then returns back to examine it.

When it comes to interaction, public displays are again unique compared to other platforms. For example, unlike desktop computers, users spend much less time interacting with public displays; average interaction durations with public displays were reported to be between 23 and 26 seconds in several deployments [33, 83]. Consequently, public displays need to be “immediately usable” [81]. Another distinction is that unlike desktop settings where users are mostly stationary and interact from the same position every time, users approach public displays from different positions, and could perceive content or even interact while they are moving [82].

For these reasons, we argue that the unique aspects of pervasive public displays warrant an in-depth investigation of how gaze-based interaction can be leveraged and enabled in this distinct domain.

## **2.1.2 Interacting with Public Displays**

Present day’s deployments of public displays feature interactivity in different forms. Some displays are implicitly interactive, that is, they react to the user’s natural behavior, for example, when they approach the display. Previous deployments displayed silhouettes that mimic users’ movements in an effort to make the display noticeable and fight interaction blindness [55, 83]. Other displays allowed explicit input through modalities beyond keys and buttons, such as touch [87], mid-air gestures [72], mobile devices [84], and recently also eye gaze [130].

### **2.1.2.1 Touch**

Touch interfaces were a significant improvement over physical hardware (e.g., keypads, buttons and joysticks) since they expand the entropy of interaction possibilities, and allow faster software-based updates of the user interface. However touch comes with its short-comings. Namely, touch interfaces have to be physically reachable, while public displays are often mounted above user’s height for visibility, or placed behind shop windows. Touch interactions also come with hygienic concerns, and could leave traces that can be maliciously exploited to extract the user’s interactions [9, 79]. This means that deploying touch-based displays requires making them reachable and thereby protecting them from 1) theft, 2) unauthorized access control, 3) intentional or unintentional damage [20].

### **2.1.2.2 Mid-air Gestures**

Interacting via mid-air gestures was proposed as an alternative modality to touch. It allows at-a-distance interaction, and can extend interaction durations due to their playful nature [3]. On the downside, mid-air gestures require teaching passersby how to perform them [2, 121], and can be socially embarrassing in public [80].

---

### 2.1.2.3 Mobile Device Interaction

Several works proposed using mobile devices (e.g., smartphones or smartwatches) as gateways to public displays. They are promising platforms for carrying profiles of their owners, hence paving the way for straightforward personalization of public displays [21, 59, 75]. They can also be used for gesture-based interaction with remote displays [23, 93]. A disadvantage of mobile device interaction is that it often requires configuring the personal device (e.g., installing an app, or connecting to a server) prior to interaction – hence the user needs to be “prepared” for interaction. Users might be reluctant to do that if they will only occasionally interact with the display, or if they do not see a clear benefit of the display yet. In case of smartphones, users might be unwilling to take it out of their pocket or bag [71].

### 2.1.2.4 Gaze

Recently, eye gaze has been gaining attention as an alternative and promising modality for interaction with public displays. Unlike touch, gaze can be used for at-a-distance interaction. Unlike mid-air gestures, gaze is subtle and less embarrassing to use in public [53]. Unlike mobile device interaction which often require installing app or taking devices out of one’s pockets, gaze is fast and intuitive [100]. Gaze provides a lot of opportunities in addressing many environmental and technical challenges on public displays, such as the observability of user’s input, and the parallax effect – we summarize some of these opportunities in chapter 3, and discuss them in details in chapters 6, 7 and 8.

We discuss the advantages and limitations of eye tracking in the following section.

## 2.2 The Eyes in Human-Computer Interaction

Since the 19<sup>th</sup> century, eye-gaze tracking has been employed in numerous fields, such as optics, medicine, as well as behavioral and psychological research. The first studies were done using direct observations [41]. For example, Javal identified saccades for the first time by studying eye movements while subjects were reading using a mirror [45]. In 1978, Rosen et al. published a report on an eye-based system for non-vocal communication for the disabled [95]. The earliest published research about the use of eyes for interaction with computers was by Bolt [13] in 1981. Bolt introduced the “World of Windows”, in which users could interact by gazing at one of multiple displays to zoom into its content. Multiple follow up works about eye tracking in HCI contexts were published afterwards [40, 65, 123]. Jacob then explored eye gaze in further details for human-computer interaction [42, 43], and identified the so-called “Midas Touch” problem on gaze-based interfaces, which refers to the problem of distinguishing whether a user is looking at something in order to perceive it, or in order to control it.



### 2.2.1 Opportunities of Eye Tracking

When tracked, the eye movements can be leveraged for understanding users. Eye movements can reflect many aspects that go beyond visual interests, such as cognitive processes, thoughts, activities and intentions. For example, eye behavior can communicate how tired, or how much cognitive demand the user is under [90]. Eye movements can give hints on a user's language proficiency [50]. Last but not least, gaze monitoring is widely used for usability evaluations [91].

In addition to tracking, eye movements can also be leveraged for interaction. Two types of gaze-based interaction are distinguished: explicit gaze-based interaction, and implicit gaze-based interaction. Explicit gaze-based interaction refers to intentionally moving the eyes to signal commands. One example of explicit gaze-based interaction is eye typing [70], where users dwell at letters to type them. While the majority of work on gaze for explicit interaction was intended for desktop settings, researchers explored other contexts including mobile devices [126], smart homes [115], and smart watches [31]. On the other hand, implicit gaze-based interaction refers to cases where the system reacts to the user's natural eye movements, i.e., the user does not control the eyes intentionally. For example, Smith et al. proposed implicit interaction by turning screens on when the user gazes at them [102], and Mariakakis et al. proposed SwitchBack, which tells the user which line they were reading before looking away from the display [73].

Researchers have already summarized eye tracking applications and opportunities [27, 69]. In our work we focus on eye tracking on public displays (see Section 2.3).

### 2.2.2 General Limitations of Eye Tracking

Eye tracking technologies do not come without limitations. As mentioned earlier, one of the most prominent problems of gaze-based interfaces is the Midas touch [42]; the eyes are perceptual organs, hence systems need to distinguish whether the user is looking at content to control it, or to merely perceive it. To counter that problem, researchers employed different methods, such as interaction by dwell time—gazing at the target for some time before activating it. Others proposed leveraging special eye movements, such as smooth pursuit [118], optokinetic nystagmus [44], and gaze gestures [25]. A second problem is eye tracking accuracy. Usually accuracy is improved if the user goes through a calibration process. However calibration is often tedious and tiring to perform [89]. Furthermore, even with perfect calibration, it is not feasible to distinguish which part of the user's  $2^\circ$  of visual view is being attended to [69]. The reason is that humans switch attention within these two degrees without moving their eyes. Another issue is that today's eye tracking technologies are influenced by light conditions – sunlight disturbs the infrared sensors in commercial eye trackers, and video-based eye tracking is not reliable in the dark.

---

## 2.3 Previous Work on Gaze-enabled Displays

Gaze has been utilized in numerous ways in HCI. Previous works classified gaze-based systems into different clusters. For example, Duchowski classified them into interactive and diagnostic applications [27], while Majaranta and Bulling classified them into explicit gaze-based systems, attentive user interfaces, gaze-based user modeling, and passive eye monitoring [69]. Similarly, eye tracking can be employed in different ways for public displays [53]. We classify the uses of gaze on public displays into three categories: (1) Attention Monitoring, (2) Implicit Gaze-based Interaction, and (3) Explicit Gaze-based Interaction.

### 2.3.1 Attention Monitoring on Public Displays

The first category is *attention monitoring*: which refers to systems that monitor the gaze behavior of passersby to quantify attention [5, 19, 106]. These systems are built with the aim of understanding where passersby look. This means that in contrast to the other categories, they do not react to the user’s gaze, but rather silently monitor visual attention for post-hoc analysis and diagnostic purposes.

Many of the models that describe the behavior of the audience around public displays distinguish a phase where users start to become aware of the presence of a display [14, 78, 81]. Knowledge about when, how, and why users transition to this phase can help public display owners make key design decisions to optimize for the display’s use case. For example, if a certain type of animation was found to attract the passerby’s attention, designers could program the displays to show these animations more often or when someone approaches. Eye tracking is hence valuable on public displays since it can provide a reliable measurement of visual attention. This, in turn, can be used to assess the methods for attracting attention, and the display’s overall effectiveness [81]. Furthermore, public displays have long been missing an equivalent for a user “clicking through” content [20]. Eye tracking can provide an alternative metric by quantifying visual attention to elements on the display.

There are two lines of previous work in this area: (1) works that proposed *technical* methods to quantify attention by leveraging eye gaze, and (2) works where gaze behavior detection was *applied* and used to understand passerby’s attention.

#### 2.3.1.1 Methods to Monitor Attention

Previous works tried to estimate attention to the display using the head-pose [10, 101, 120]. More recently, some works incorporated the gaze direction to estimate attention to public displays [5, 106]. Alt et al. used a geometric-based gaze estimation method [5]. Geometric-based approaches first detect facial and eye features such as the pupil’s center and eye corners, and then try to fit a geometric

model of the eye onto the user's eye. A gaze vector is then extended from the center of the estimated eye ball to the pupil's center, to eventually intersect the display at a certain point. This point is then deemed to be the gaze point. Geometric approaches are sometimes referred to as feature-based approaches if they rely solely on the features without model fitting. While geometric approaches can often yield highly accurate gaze estimates, they are susceptible to changing lighting conditions.

On the other hand, works such as AggreGaze [106] employ appearance-based gaze estimation approaches. Appearance-based approaches directly map images of the user's eyes and/or face to gaze estimates based on training datasets and machine learning, e.g., convolutional neural networks [130]. Appearance-based approaches are often more robust against different lighting conditions, and can even be optimized by artificially changing the illumination levels of the images in the training dataset [125]. Their downside is that they require training data.

### 2.3.1.2 Attention Monitoring in Practice

Apart from technical works, in this section we discuss findings from actual deployments that employed eye tracking to remotely monitor the audience's attention. For example, Ravník and Solina [92] mounted a camera on a public display and measured attention time to find that passersby gazed at the display for an average of 0.7 seconds. They found that men and children are more likely to gaze longer at the display compared to women and adults respectively. They also found that showing dynamic content results in higher attention time. Using face detection and machine learning, ReflectiveSigns [82] scheduled content to be displayed based on previous experiences of which content attracted passersby attention the most.

Other works investigated the use of mobile eye tracking while recruited participants walked in front of displays. For example, researchers have run studies to quantify attention to public displays in transportation and shopping streets [39, 97]. Dalton et al. [19] recruited 24 participants, and asked them to walk through a shopping mall while wearing mobile eye trackers. They found that users gazed at displays for brief durations of approximately 0.3 seconds on average. Their results also suggest that the architecture of the building in which the displays are deployed influences the gaze behavior.

## 2.3.2 Implicit Gaze Interaction with Public Displays

Second, rather than requiring users to intentionally control their eyes, *implicit gaze interaction* refers to adapting the experience based on the user's natural eye behavior. This area is sometimes referred to as attentive user interfaces [69, 98, 116].

Implicit gaze interaction can bring a myriad of benefits to interactive public displays. The user's gaze could reflect interests [44, 68], and intentions [61]. The ad-

---

vantage of implicit interaction is that it does not require training the user to interact via gaze, since no explicit eye-based actions are needed.

This type of gaze-based interaction can be classified to two categories: (1) systems that react once the user establishes *eye contact* with the system, and (2) systems that *continuously react* to the estimated gaze point based on the user's eye movements.

### **2.3.2.1 Eye-Contact Aware Implicit Interaction**

A straight forward way to exploit implicit gaze-based interaction, is to program displays to be attentive to the user's gaze, i.e., react once users look at the display [117]. Relatively less research has been done in this area. Early works detected eye contact to mobile devices [22, 117], and home appliances [99]. Similarly, Smith et al. [102] and Zhang et al. [127] proposed methods for detecting eye contact on public displays, and accordingly trigger actions, such as turning the display on.

### **2.3.2.2 Gaze-Behavior Aware Implicit Interaction**

While the aforementioned works focused on eye contact detection, other works estimated which on-screen objects the user is looking at and adapted the experience accordingly. For example, PeepList [49] is a pervasive display that estimates the perceived importance of on-screen information through the user's gaze. The system dynamically generates a list of content sorted by importance. Mubin et al. [80] developed an intelligent shopping window where the system responded to user's gaze towards products, which was determined via head tracking. Brudy et al. [15] and Eaddy et al. [28] exploited eye tracking to protect public display users from shoulder surfing by hiding private content that passersby gaze at. Schmidt et al. [96] displayed text that follows the user's eye gaze as they walk past large public displays. Lander et al. developed the collaborative newspaper [63], a system that allows multiple users to read different parts of the same article simultaneously. Karolus et al. [50] detected language proficiency on public information displays through eye gaze, and accordingly showed translations of the on-screen text.

## **2.3.3 Explicit Gaze Interaction with Public Displays**

The final category is *explicit gaze-based interaction*, which refers to explicitly employing eye movements to provide input on public displays.

Explicit gaze interaction is attractive for public displays because it is fast [100], and natural in most of the times [118]. Furthermore, gaze-based input can be provided at a distance, which is particularly useful for public displays that are physically unreachable; users can still interact by gaze or a combination of gaze and mid-air gestures even if the display is placed behind glass windows or above user's height. Finally, gaze-based input has several advantages over existing input methods. For

example, input via touch is associated with hygienic concerns, while mid-air gestures are sometimes embarrassing to perform in public [72].

Explicit Gaze interaction can be further dissected into (1) explicit gaze-only interaction, where gaze is the sole input method, and (2) explicit gaze-supported interaction, where gaze is used alongside another input modality to improve it.

### **2.3.3.1 Gaze-only Interaction**

Gaze has also been used as the sole input modality for interaction with public displays. Sippl et al. [101] estimated the user's horizontal and vertical gaze, to determine which of four quadrants the user is looking at. Zhang et al. [128, 130, 131] estimated the horizontal gaze direction to browse content, such as news and albums. Vidal et al. [118] introduced the use of smooth pursuit eye movements for calibration-free gaze interaction with public displays. They referred to their method as "Pursuits", and since then it has been used in multiple applications [18, 67]. Jalaliniya and Mardanbegi proposed using Optokinetic Nystagmus for interaction, i.e., an eye movement that features multiple saccades followed by an extended smooth pursuit movement after finding the target of interest [44]. While they do not evaluate their system on a public display, they present public displays as one of the main domains where their method is promising. In GazeProjector [62], users interacted via eye gaze with surrounding displays by wearing a mobile eye tracker. Gaze was estimated on displays with the help of an edge detection algorithm that identifies borders of surrounding displays.

### **2.3.3.2 Gaze-supported (Multimodal) Interaction**

Gaze can be effective when used alongside other input modalities. For example, Stellmach and Dachsel [104, 105], as well as Turner et al. [110] combined gaze with touch to facilitate the manipulation of targets on large public displays. Gaze has been combined with other modalities for transferring content from public displays to personal devices. For example, Mäkela et al. combined mid-air gestures with gaze to retrieve content from public displays [72], Turner et al. used multimodal touch and gaze input to transfer content on public displays [111, 112, 113]. Gaze was even combined with feet for interaction with situated displays in operation rooms [37].

## **2.4 Trends and Limitations of Previous Work**

The early days of gaze-enabled displays focused on the technical realization of eye tracking and gaze-based interaction. With the advancements in hardware and visual computing, researchers and practitioners moved from using the head-pose and face orientation [80, 96, 101, 120], to using the actual gaze point estimated based on images of the user's eyes [5, 128].

---

Previous work focused on specific aspects of gaze-enabled displays, such as how to estimate gaze based on face orientation and head-pose [80, 96, 101, 120], methods for calibration-free gaze interaction [25, 44, 118, 129], using mobile eye trackers for interaction with public displays [62, 64, 111], measuring the audience attention [5, 19, 39, 92, 97, 106], and inventing novel concepts for gaze-supported interaction [72, 88, 104, 105, 110]. However, existing work is lacking in several directions.

First, there is a lack of interactive systems that leverage gaze to address “practical issues” that are specific to public displays. In our work, we propose both implicit and explicit gaze-based approaches to address important problems on public displays, such as the parallax effect (Chapter 7) and secure authentication (Chapter 8).

Second, there is a lack of work in enabling remote gaze-based interaction for users of large public displays who are, in contrast to desktop settings, typically approaching the display from different directions, and even interacting while moving. We explore this unique aspect of public displays in our implementation and evaluation of EyeScout (Chapter 9), which enables gaze-based interaction from spontaneous positions in front of the display, as well as while moving past the display.

Third, with the exception of the work of Pfeuffer et al. [89], there is a lack of work in intuitive calibration methods that are suitable for public display deployments. We propose a novel calibration method that calibrates eye trackers while users read text on public displays (Chapter 10), which is one of the most common types of content on public displays [6, 38, 57, 86, 107, 114, 119].

Fourth, when designing gaze-based interaction, it is often the case that the configuration that results in the highest accuracy is not the most usable one (e.g., in terms of interaction speed). We make this observation in our TextPursuits project (Chapter 10), and we find this trade-off as well when choosing between highly accurate calibration methods that take time and faster ones that are inaccurate (cf., 9-point calibration vs 5-point calibration). It is not clear if public display users would prefer highly accurate methods, or would rather tolerate the inaccuracies of more usable and faster methods. We investigate this issue in our in-the-wild field study of a gaze-based voting system (Chapter 11).

# II

## SUMMARY OF PEER-REVIEWED PUBLICATIONS





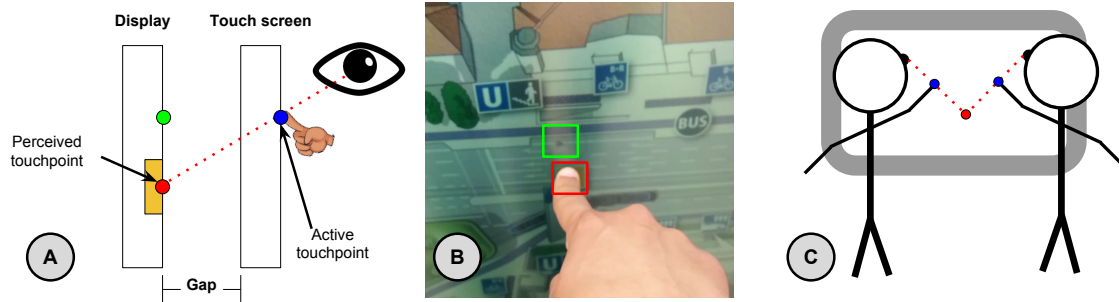
# 3

## Opportunities of Gaze on Public Displays

Eye tracking can be leveraged in many ways to bring tangible benefits to public displays. In Chapter 6, we provide a review of challenges of pervasive public displays, and discuss how gaze shows promise in addressing many of them. For example, detecting the passerby's attention is a challenge at the outset of interaction with displays [47, 83]; gaze can be an indicator of the user's attention towards the display [19, 102]. Passersby are trained to ignore displays in public and often do not know that they are interactive [60, 74, 83], but gaze can be leveraged to communicate that the display is interactive by showing call-to-action labels [20] right where the user is looking, hence making it unlikely the user will miss it. In many cases displays are behind glass windows for security, or above the user's height for visibility; gaze can be used at-a-distance and hence allows interaction with displays that are physically unreachable. Gaze is generally subtle and hence less embarrassing to use in public compared to other modalities, such as mid-air gestures [14, 72]. Finally, interactive public displays require "immediate usability" [20, 81]; gaze is fast (e.g., faster than pointing [100]) and hence promising for enabling immediate usability.

In this work we focus on gaze-based *interaction*, that is, allowing users to intentionally or unintentionally provide input that is either supported by eye gaze or entirely done using the eyes. We address issues that are particular to interacting with displays in public space. Namely, the parallax effect, and the observability of the sometimes sensitive user input. In the following, we summarize our contributions in implicit and explicit gaze-based interaction with public displays.

- Section 3.1 describes EyePACT, an implicit gaze-based method that overcomes the parallax effect, which is a recurrent problem on today's touch-based public displays.
- Section 3.2 describes GTmoPass, a secure multimodal approach that combines gaze and touch input for explicit user authentication on public displays.



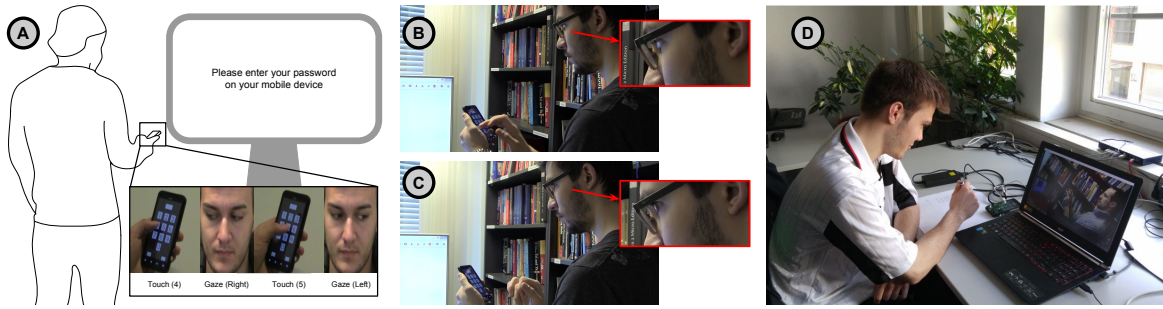
**Figure 3.1:** Public displays are often augmented with thick layers of glass to protect them against vandalism and harsh weather. This results in a gap between the touch-screen and the display, which in turn introduces the parallax effect illustrated in (A). The parallax effect refers to the displacement between the perceived touchpoint and the active one, and can be observed on today’s public displays. (B) shows a situation where parallax was observed on a display in Münchner Freiheit station in Munich, Germany. EyePACT solves this problem by leveraging the eye’s position.

### 3.1 EyePACT: Implicit Eye-based Parallax Correction

*Problem:* Public displays that employ touch interaction utilize touch sensitive (capacitive) surfaces at the outset of the display. A problem arises when displays are deployed in public space; public displays are often augmented with thick layers of glass to protect them against vandalism and harsh weather conditions. These protective layers are placed in between the touchscreen and the display, resulting in a relatively large distance between the surface being touched and the display elements. As shown in Figure 3.1A, the distance between the display and touchscreen causes a displacement between the perceived touch point (red area in Figure 3.1) and the touch point detected by the system (green area in Figure 3.1). This effect is referred to as the Parallax effect.

*Opportunity:* As described in section 2.3.2 knowledge about the user’s gaze direction can implicitly complement interactions when using other modalities, such as touch. We exploited implicit gaze-based interaction to address the parallax effect, a prominent problem on today’s interactive displays.

*EyePACT:* EyePACT is an Eye-based Parallax Correction technique for Touch-enabled displays. EyePACT is a simple, yet effective, method to correct for parallax by estimating a gaze vector that starts from the user’s eyes, intersecting the user’s touch point, and then finally intersecting the display at a certain point. The touch event is then initiated at the point the user is intending to touch. This work is the first to systematically evaluate the concept of eye-based parallax correction; we built an apparatus to create and control the parallax effect in our lab.



**Figure 3.2:** GTmoPass employs explicit gaze-based interaction to allow users to securely authenticate on public displays using multimodal passwords consisting of gaze and touch input, and a personal mobile device.

*Evaluation:* We carried out multiple studies (N=38) to evaluate the method’s effectiveness and how well it is perceived by users in case of single and multiple users, different target sizes, and different levels of parallax. When evaluating EyePACT for multiple users, we also found that it prevents interfering fingers when multiple users interact with the same on-screen target via touch (see Figure 3.1C). This presents an opportunity that could enable novel multiuser interaction concepts.

*Findings:* In summary, we found that EyePACT significantly improves the accuracy of touch-enabled displays even with varying gap distances between the touch surface and the display, that it adapts to different levels of parallax, and that it maintains a significantly large distance between multiple users’ fingers when interacting with the same target. This project is discussed in more details in **Chapter 7**.

## 3.2 GTmoPass: Explicit Gaze-supported Authentication

*Problem:* There are many situations in which users need to securely authenticate on public displays. In addition to ATMs, users could authenticate at public terminals when buying tickets for museums, buses, etc. Furthermore, the increasing demand for personalized and context-related experiences on public displays [21] underlines the need for secure user authentication to ensure that only the legitimate user accesses their potentially sensitive data. However, many of today’s concepts for authentication are vulnerable to various types of side-channel attacks, such as shoulder surfing [29], smudge attacks [9], and thermal attacks [1].

*Opportunity:* As explained in section 2.3.3, in addition to the implicit use of gaze, gaze can also be employed explicitly to provide input. An advantage stemming from the subtleness of eye movements, is that it can be employed for inconspicuous authentication in public space. This benefit is amplified when gaze is combined

---

with another modality, since each additional modality translates to an additional channel of input that attackers need to observe. In our work, we leveraged explicit gaze-based interaction for secure authentication on public displays.

*GTmoPass:* Users authenticate using GTmoPass by providing multimodal passwords on their mobile devices. The password consists of gaze and touch input (see Figure 3.2A). The password is then securely transmitted to a server through Bluetooth beacons. The proposed system is a two-factor authentication scheme where the multimodal password is the knowledge factor (i.e., something the user knows) and the mobile device is the possession factor (i.e., something the user has). Gaze input is detected through the front facing camera of an off-the-shelf mobile device, and touch input is directly provided on the touchscreen. This way, in order for an adversary to attack the user, the adversary would need to (1) observe the user's eyes to find the gaze input, (2) observe the phone's screen to find the touch input, (3) combine the observations to generate the password, and (4) acquire the mobile device (e.g., by theft).

*Evaluation:* We evaluated the usability and security of the concept (N=32). Usability of GTmoPass was evaluated in the lab by inviting 16 participants to enter Gaze-Touch passwords on their mobile devices (Figure 3.2B and 3.2C). Afterwards we conducted a security study with 16 different participants, who took the roles of attackers and tried to observe passwords by watching videos of users authenticating using GTmoPass (Figure 3.2D).

*Findings:* We found that although authentication using GTmoPass is slightly slower than traditional methods, it is highly resilient against the aforementioned threats. This project is discussed in more details in **Chapter 8**.

*Acknowledgement:* This project was supported by Google IoT research award received in March 2016.

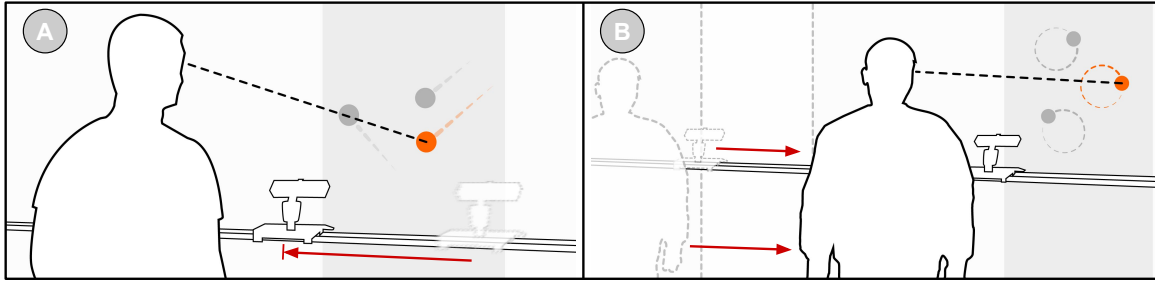
## Challenges of Gaze-enabled Public Displays

Although research in gaze detection and gaze-based interaction is already well established for desktop settings, the domain of gaze-enabled public displays is unique and imposes novel challenges that are under-investigated. In our work, we identified three main unique aspects of gaze-enabled public displays that require further investigation. Namely, in contrast to desktop settings where the user is (1) positioned in front of the eye tracker, (2) mostly stationary, and (3) interacting for a long period of time, public displays (1) expect users to interact from different positions, (2) expect users to move while interacting, and (3) cannot afford requiring time-consuming eye tracker calibration.

We focus on the use of remote eye trackers; that is, we investigate gaze-based interaction with public displays without augmenting the user. While head-mounted eye trackers are becoming affordable [51], and show promise in tackling the aforementioned challenges [62], they are still special-purpose equipment that require augmenting individual users [64]. Indeed there is a vision of having eye trackers already-integrated into daily Eyewear [16], and also the vision of having pervasive interconnected display networks in the future [20]. Accordingly, it is possible that in the future passersby would be wearing eye trackers that are able to communicate the gaze behavior to surrounding displays. However, a pervasive integration on such a big scale would require taking concepts from lab settings to the field, which is currently challenging to investigate using mobile eye trackers unless participants are explicitly hired [19]. Until wearing mobile eye trackers becomes a norm, there is a need to study the natural user behavior on gaze-enabled public displays using remote eye tracking.

In the following we discuss how we addressed these unique aspects of gaze-enabled public displays in our research.

1. Section 4.1 describes our prototype system, EyeScout, that allows movement and position independent gaze-based interaction via active eye tracking.
2. Section 4.2 summarizes our investigation of implicit calibration (Read2Calibrate) and explicit interaction (EyeVote) with textual content on public displays via gaze.
3. Section 4.3 summarizes an in the wild deployment of EyeVote, that allowed us to investigate the trade off between accurate and fast gaze interaction techniques, and the implications on willingness of passersby to correct system errors resulting from inaccuracy on gaze-enabled displays.



**Figure 4.1:** EyeScout is an active eye tracking system that allows users to interact via gaze from random positions in front of large public displays (A), or while moving past the display (B).

## 4.1 EyeScout: Movement and Position Independent Gaze Interaction

*Problem:* Public displays expect users to interact from different locations, distances, and orientations relative to the display [83, 122]. Large public displays in particular allow passersby to interact while on the move [96]. On the other hand, most existing remote eye tracking approaches require users to keep their head facing the tracker in a relatively confined tracking box.

*EyeScout:* In our work, we explored the use of “active eye tracking” for the first time for public displays to address these issues. Active eye tracking refers to the automatic adaptation of the eye tracking hardware to adapt to the user’s state rather than restricting the user’s movements. We do so by presenting EyeScout, a novel active eye tracking system that enables gaze interaction on large public displays from different lateral positions in front of the display and while on the move (see Figure 4.1). EyeScout consists of two main modules: a body tracking module, and an eye tracking module that is mounted on a rail system. The system detects the user’s position, and then dynamically adapts the eye tracker’s position by moving it to face and following the user. This way, EyeScout (1) allows gaze-based interaction with large public displays for users interacting from *different positions* within a certain distance in front of the display (Figure 4.1A), and (2) allows users to interact via eye gaze *while moving* in front of the display (Figure 4.1B). We evaluated EyeScout for scenarios in which users “Walk *then* Interact” (to test for position independence), as well as “Walk *and* Interact” (to test for movement independence).

*Findings:* We found that EyeScout is well-suited for both interaction modes, it is well perceived, and it reduces the interaction initiation time by 62% compared to state of the art systems [7, 130]. Based on these positive results, we proposed several contexts in which EyeScout can be particularly useful, such as large displays along escalators and moving pathways, as well as large cylindrical displays. This project is discussed in more details in **Chapter 9**.



**Figure 4.2:** In our project TextPursuits, we implemented and evaluated two systems: (A) Read2Calibrate allows users to implicitly calibrate the eye tracker while reading text on public displays, and (B) EyeVote allows users to vote on public displays by picking one of several text-based options by following them with their eyes. Through two user studies, we identified several factors that influence the accuracy and user perception, as well as a trade-off between accuracy and perceived user experience.

## 4.2 TextPursuits: Implicit Calibration and Calibration-free Gaze Interaction

*Problem:* Immediate usability is one of the main requirements of interactive public displays [81]. Users interact with public displays for very short durations [33, 81, 83], hence it is very important to make sure that the time needed to kick off interaction on public displays is as fast as possible. On the other hand, eye trackers typically need to be calibrated for each user [69]. Calibration is a procedure in which the user is asked to gaze at several fixed points on the screen to allow the eye tracker to collect eye images, and establish a mapping of eye movements to gaze points on the screen. While calibration results in more accurate gaze estimation, the process is tedious and time-consuming to the user [89]. Investing time to calibrate an eye tracker in a desktop setting might be acceptable because the user is going to interact for a relatively long period of time. In contrast, spending a big portion of the interaction duration solely to calibrate the eye tracker is unacceptable in public settings [53]. In the following, we discuss two prototypes: *Read2Calibrate* and *EyeVote* that demonstrate two approaches to address this problem.

*Read2Calibrate:* One way to reduce the overhead caused by calibration is to make the process easier by interleaving it into the user’s usual interactions. Read2Calibrate is an implicit calibration method, with which users calibrate eye trackers simply by reading text on a public display (Figure 4.2B). The idea is motivated by the abundance of text-based content on public displays that the users read anyway. Read2Calibrate shows animated text on the screen and feeds the gaze data to an algorithm that gradually enhances the calibration of the gaze tracking. The concept builds over previous work in Pursuits, in which moving targets were shown to calibrate as users follow the moving targets with their eyes [89].

---

*Findings:* In a user study (N=18), we found that text-based calibration is not as precise as state-of-the-art calibration procedures, but it is sufficient to identify the gaze area within  $6^\circ$  of visual angle. The results show that the inclination of the shown text and the speed at which it is gradually revealed have a strong effect on accuracy. Most interestingly, the findings indicate that there is a trade-off between the configuration leading to the most accurate results and the users' preferred configuration.

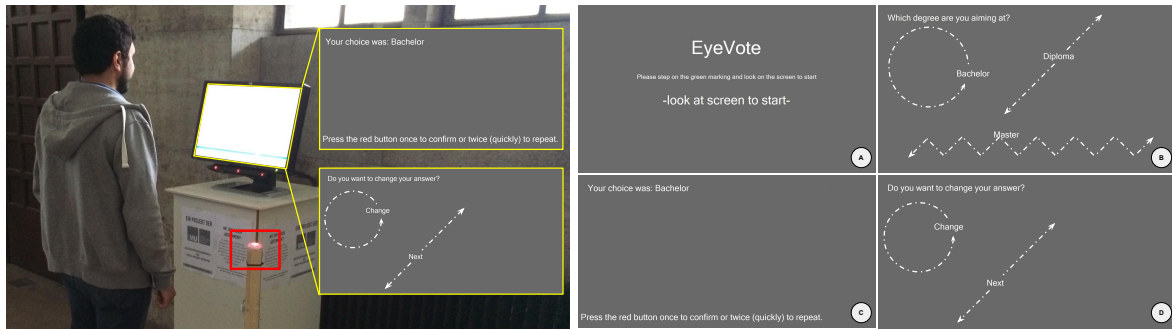
*EyeVote:* Another approach to reduce the overhead caused by calibration is to completely eliminate it and use calibration-free gaze techniques, such as Pursuits [118]. While Pursuits has been used for many use cases, such as games [52, 118], authentication [18], text entry [67], and smart rooms [115], it had never been explored with text prior to our work. Pursuits is not necessarily straight forward to employ with textual stimuli: reading is typically associated with saccadic eye movements rather than smooth pursuit. Hence, eye movements performed while reading a piece of text that is moving could overlay the smooth pursuit movement, which could in turn result in difficulty in correlating movements of the target with that of the eyes. Also, due to the Midas effect, gaze-based systems need to distinguish users reading textual content from interacting with it. To better understand this, we implemented EyeVote (Figure 4.2A), a survey application that allows users to cast their votes on public displays by selecting one of several text-based options.

*Findings:* Results from a user study (N=19) revealed that representing text in short concise form makes it suitable for Pursuits-based interaction. We also found that moving text-based targets in circular motion results in the highest accuracy, but worst user experience since circular eye movements are demanding and tiring. These projects are discussed in more details in **Chapter 10**.

### **4.3 EyeVote in the Wild: Understanding the Trade-off between Accuracy and User Experience**

*Problem:* An outcome of both systems that were discussed in Section 4.2 is that the configuration that results in the highest accuracy is not necessarily the one that results in the best perceived user experience. In fact, there seems to be a trade-off between accuracy and user experience. For example, in *Read2Calibrate*, users preferred to read flat text because it is more natural to read, but text inclined at  $45^\circ$  yielded the highest calibration accuracy because it spans across both the x- and y-axes. Similarly, in *EyeVote*, users found targets moving in circular motion to be the most demanding to select. Yet the fact that circular eye movements is not common also meant that circular trajectories result in the least false positives. This presents designers with a challenge: should we design interfaces that are highly accurate even though it might be ill perceived? or should we use the less accurate methods and allow users to correct or confirm their input? While similar questions





**Figure 4.3:** We deployed EyeVote in a field study to understand the willingness of public display users to correct system errors. EyeVote was configured to occasionally show intentionally falsified feedback and prompt users to correct their vote. Results from a field study show that most participants do correct system errors when the correction method is straight forward.

were investigated in several domains, public display systems pose very unique design challenges in terms of error recovery. Public display users interact for very short amounts of time and are believed to abandon the display when interrupted or forced to deviate from the main task [35, 121].

*EyeVote in the Wild:* To answer these questions, we conducted an in-the-wild field study of EyeVote where we deployed the system in a university campus. To experiment with the users' willingness to correct system errors, we programmed EyeVote to intentionally evoke errors to see if users correct them. We did that by showing falsified feedback; this was done by showing an answer in the recap view that was not among the options the user had available to choose from.

*Findings:* We found that public display users are willing to correct system errors provided that the correction is fast and straightforward. For example, we found that users are less likely to correct errors if the correction mechanism requires a change of interaction modality. This project is discussed in more details in **Chapter 11**.

---

# III

## CONCLUSION AND FUTURE WORK



# 5

## Closing Remarks

This dissertation investigated the design of gaze-based interaction for pervasive public displays. This was done through two lines of work. First, based on an in-depth review of challenges of pervasive public displays, we identified a manifold of *opportunities* eye gaze could bring to this setting. We then proposed the concept, implementation, and in-depth evaluation of two systems that address profound problems on today's pervasive public displays, namely, these systems offer accurate parallax correction (Chapter 7), and secure user authentication (Chapter 8) on pervasive public displays. Second, we identified core challenges of gaze-enabled public displays that are unique to this setup and are very different compared to previously investigated challenges. Specifically, we identified 1) the problem of supporting interaction from different user positions, including cases where the user is moving (Chapter 9), and 2) the issues pertaining to calibration, where we investigated implicit calibration (Chapter 10), calibration-free interaction (Chapters 10 and 11), and input correction in case of fast but inaccurate gaze-based selections (Chapter 11).

---

## 5.1 Design Recommendations

In addition to the individual recommendations in each of Chapters 9, 10, and 11, we highlight the following recommendations for designing gaze-based interaction on pervasive public displays:

1. **Employ Active Eye Tracking for Gaze Interaction in Public Space:** Gaze-enabled displays need to adapt to dynamic users who could approach the display from different directions or interact from different positions and while moving. In EyeScout (Chapter 9), we employed active eye tracking, i.e., dynamically adapting the tracker to the user's position, with success. Active eye tracking can be achieved in many other ways, such as by using drones, or using Pan-Tilt-Zoom cameras for eye tracking.
2. **Use Implicit Calibration, or Calibration-free Gaze Interaction Techniques:** Calibration is tedious and unacceptable for public interactions, especially if the interaction duration is expected to be short. Employing implicit calibration by integrating it into interaction reduces the perceived calibration overhead. This can be done by integrating calibration into reading text as we did in Read2Calibrate (Chapter 10). Alternatively, gaze interaction techniques that do not require calibration, such as Pursuits [118] and gaze gestures [25], should be used.
3. **Optimize for Fast Gaze Interaction, Even if it is at the Expense of Accuracy:** In many of our projects, we found a trade-off between gaze interaction speed and gaze interaction accuracy. We found that passersby are willing to correct system errors on public displays as long as correction methods are fast and straightforward (Chapter 11). This means that using faster interactions techniques despite their low accuracy, and allowing users to correct incorrect detections would yield better results in public space.

## 5.2 Future Work

Apart from public displays, the methodology followed in this dissertation can be adopted for other pervasive domains, such as mobile devices. Furthermore, while this dissertation closes multiple gaps, it highlights opportunities for future work relevant to researchers and practitioners working in eye tracking, public displays, and gaze-enabled displays.

### 5.2.1 Our Methodology Beyond Public Displays

Advances in ubiquitous computing brought forth a plethora of novel devices. At the same time, eye tracking technology is continuously advancing, hence nudging researchers to explore the opportunities it brings to new domains in which eye tracking is relatively less explored. For example, researchers have started studying eye tracking in domains such as handheld mobile devices [54], on smart watches [30], internet of things [115], as well as AR [31] and VR [56].

As briefly mentioned in section 1.3.3, the methodology we adopted in this work can be adapted for studying the use of eye gaze for other domains. Our methodology is as follows: 1) we identify challenges in the field of pervasive displays, 2) study how gaze can address them, 3) investigate novel challenges arising from the use of gaze in this domain, and then 4) design, implement and evaluate solutions to address these challenges. This methodology can be adapted for other domains for which eye tracking is being explored. For example, researchers interested in studying eye tracking and gaze-based interaction for handheld mobile devices can follow a similar approach: The first step is to identify the challenges users currently face on mobile devices. An example of such challenge is the difficulty of reaching and tapping on targets that are at the top of the mobile device's screen. The second step is to experiment how eye gaze can help solve these problems. Following the same example, a solution could be to employ multimodal interaction where the user gazes at an unreachable target, and then taps on a specific area that is easily reachable. Finally, challenges that are particular to gaze-enabled handheld mobile devices need to be investigated. For example, unique challenges in this domain include the shaky environments in which mobile devices are used, the susceptibility to motion blur, and the fact that the user's holding posture influences the visibility of their eyes in the front-facing camera's view. In fact, we started employing this methodology for gaze-enabled handheld mobile devices in a number of side projects, which has resulted in a better understanding of the opportunities and challenges of eye gaze in this domain [54].

### 5.2.2 There is More to Gaze Interaction than Accuracy

Research in eye tracking has traditionally focused on improving the accuracy of gaze estimation. While improving accuracy is important, our findings suggest that there are aspects that could sometimes be more important than accuracy.

Many of our studies shed light on a trade-off between gaze interaction accuracy and gaze interaction speed – the setup that results in the highest accuracy is usually the slowest. In the particular, we found that compromising accuracy for the sake of faster interactions is more suitable on public displays; users are willing to correct

---

system errors of a fast inaccurate system rather than interacting with a slow accurate one. We discuss this in details in Chapters 10 and 11.

In this work, we have extensively used techniques that leverage gaze behavior (e.g., gaze gestures [25] and Pursuits [118]). These techniques are fundamentally different compared to the traditionally used gaze fixations, since they offer accurate means for interaction even though they rely on inaccurate gaze estimates. This makes Pursuits and gestures not only suitable for spontaneous calibration-free interaction, but also for interaction in dynamic setups, such as when walking along displays, where gaze estimates are likely to be inaccurate. We discuss this in more details in Chapter 9. This demonstrates another example where gaze estimates can be useful despite their inaccuracy.

While building highly accurate gaze estimation methods is important, future work in gaze-enabled displays should accept this compromise. Instead of focusing only on improving accuracy, we should additionally build methods that embrace inaccurate gaze estimates to enable highly usable interaction.

### 5.2.3 Privacy Implications of Gaze-enabled Public Displays

In this dissertation, we leveraged the user’s eyes to build systems and methods that improve interaction with public displays. However, being a rich source of information about the user, tracking the user’s eyes in public space comes with implications on privacy. Namely, eye movements do not only reflect the user’s visual attention, but also potentially sensitive information that the user might not wish to share, such as the mental disorders [46], psychiatric diseases [108], and could even reflect political temperaments [24]. Additionally, since eye trackers are basically cameras, users might be concerned about the misuse of the video feed. For example, users could think that the space owner shares their behavioral patterns with third parties to generate targeted advertisements [55].

This is a challenge for future gaze-enabled public displays. One direction to address this issue is to enforce a policy, similar to that of certificate authorities used for HTTPS, that requires space owners to *certify* gaze-enabled public displays that do not violate the user’s privacy. Another approach would be to communicate to the user how the video feed is being processed, or what is being done with the gaze data. To this end, gaze-enabled public displays can borrow concepts used for mobile devices; there has been work on nudging users when private data or permissions (e.g., the user’s location) is shared with a mobile application [4, 11].



# IV

## OPPORTUNITIES OF GAZE-ENABLED PUBLIC DISPLAYS



# 6

## **Tackling Challenges of Interactive Public Displays using Gaze**

---

# Tackling Challenges of Interactive Public Displays using Gaze

**Mohamed Khamis**

Media Informatics Group  
University of Munich  
Munich, Germany  
mohamed.khamis@ifi.lmu.de

**Andreas Bulling**

Perceptual User Interfaces  
Group  
Max Planck Institute for  
Informatics  
Saarbrücken, Germany  
bulling@mpi-inf.mpg.de

**Florian Alt**

Media Informatics Group  
University of Munich  
Munich, Germany  
florian.alt@ifi.lmu.de

**Abstract**

Falling hardware prices led to a widespread use of public displays. Common interaction techniques for such displays currently include touch, mid-air, or smartphone-based interaction. While these techniques are well understood from a technical perspective, several remaining challenges hinder the uptake of interactive displays among passersby. In this paper we propose addressing major public display challenges through gaze as a novel interaction modality. We discuss why gaze-based interaction can tackle these challenges effectively and discuss how solutions can be technically realized. Furthermore, we summarize state-of-the-art eye tracking techniques that show particular promise in the area of public displays.

**Author Keywords**

Gaze; gaze-based interaction; public displays; pervasive displays; digital signage

**Introduction**

Public displays have become ubiquitous in public spaces, such as shopping malls or transit areas in airports and train stations. Equipped with an increasing number of sensors, their interactive capabilities promise informative, entertaining and engaging applications that provide tangible benefits to users. Such sensors include, for example, touch screens, cameras, and depth sensors, thereby enabling interaction

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
*UbiComp/ISWC '15* Adjunct, September 07-11, 2015, Osaka, Japan.  
© 2015 ACM. ISBN 978-1-4503-3575-1/15/09...\$15.00  
DOI: <http://dx.doi.org/10.1145/2800635.2807951>

based on touch, mid-air gestures, smartphones, and recently also gaze. Nevertheless, the uptake of interactive public displays has been slowed due to many challenges.

This work suggests using gaze to tackle core challenges of display interaction. Gaze-based interaction has numerous advantages: gaze is intuitive [21], natural to use [22], indicates visual attention and usually precedes action [13].

### Eye Tracking Techniques for Public Displays

Classical eye tracking techniques require performing time-consuming and cumbersome calibration [13]. This prerequisite has slowed the adoption of gaze for public displays. Previous work overcame this by estimating gaze with low accuracy based on head tracking and face detection [5].

More sophisticated calibration-free eye tracking methods, that show promise in the domain of public displays, were recently introduced. Examples include the work by Zhang et al. [23] that relies on relative eye-movements (e.g. distance between the pupil and eye corner). Vidal et al. [22] proposed leveraging the smooth pursuit eye movement to enable spontaneous gaze-based interaction. Pursuits has been evaluated on a public display and was shown to be well-perceived by passersby [11]. The method has also been used for flexible calibration [18]. Nagamatsu et al. [17] enabled calibration-free eye tracking using a sophisticated hardware setup (2 cameras and 8 LEDs).

With our work we aim to identify which of the techniques are suitable for particular applications and situations.

### Addressing PD challenges using Gaze

Gaze can be used both implicitly and explicitly to enhance the user experience with public displays. By looking into existing work, six main challenges that hinder the uptake of interactive public displays have been identified, in this

section we shed light on why and how gaze can be superior over existing techniques in addressing these challenges.

#### *Detecting the user's attention*

Attracting and detecting the user's attention are core challenges at the outset of the interaction process with public displays [7]. Previous work presented readily interactive displays and tried to attract the passersby's attention using physical objects [10] or user representations such as mirrored silhouettes [16]. Although inferring "attention" to a display is complicated, the passerby's gaze indicates overt visual attention and in many cases precedes action [13]. Recent research used wearable eye trackers to detect visual attention to displays [6]. Also remote eye trackers can be used to detect a user's gaze, which can be a plausible indication of his/her attention to the display, particularly when combined with head orientation and body posture.

#### *Communicating Interactivity to the Passerby*

In order for the passers-by to distinguish interactive displays from static advertising screens, a public display needs to communicate that it is interactive [7]. Existing approaches include flipping an edge of the display [12], using *call-to-action* labels, using signs next to display [14] or assigning someone to invite passersby to interact [9]. Based on gaze-data, it is possible to show a concise *call-to-action* label right where the passerby is looking, at the moment s/he attends to the display. This makes it less likely to be overlooked compared to existing approaches.

#### *Accessibility and Social Embarrassment*

Another challenge is the accessibility of the displays. Touch-based interaction is not always possible due to the display's location (e.g. in many cases the display is behind a glass window or mounted above head-height for visibility and security [7]). Gesture-based interaction is often difficult due to the lack of a generally agreed-upon gesture-to-action map-

---

pings. Moreover, mid-air gestures were found to be embarrassing for users in public [4], particularly if visible from afar. By using remote eye trackers, interaction via gaze becomes very subtle and can be hardly recognized by others in public; thus overcoming the embarrassment problem, while maintaining the advantage of at-a-distance interaction.

#### *Immediate Usability*

When it comes to interaction, there is the requirement of immediate usability [7]. Interaction time with a public display is often short (in seconds) [15]. So far this has been addressed by using interaction concepts that require a low learning curve, and by using *call-to-action* labels [14]. Gaze-based interaction is fast [19] and intuitive [21]. Gaze can also be used alongside other interaction modalities to improve their usability. Combining midair gestures and/or touch interaction with gaze tracking in public displays holds promise by, for example, adapting UIs based on users' visual attention.

#### *Privacy in Public Spaces*

As displays become more interactive there is a need to enable personalization and to allow users to input data (e.g. add a post [2]). Consequently, displays need to deal with sensitive data (e.g. passwords), that users will be skeptical to provide in a public environment. This problem is currently mitigated by asking users to exchange sensitive data through their mobile devices [3]. Using gaze-based authentication was shown to be more secure than classical methods [8]. Moreover, previous work has demonstrated the feasibility of content exchange across devices via gaze [20]. This makes content-exchange less prone to observations and less likely to leave exploitable smudge traces.

#### *Gaze as a Performance Indicator*

Unlike websites, public displays have no equivalent of a user "click stream", which makes it difficult to track user ac-

tions for evaluation purposes [7]. Gaze can offer metrics to quantify the performance of displays. Such metrics include, among others, dwell time and number of fixations; these can be used as indicators of attention, perception, understanding, and interest.

#### **Limitations**

Eye tracking could be challenging outdoors as the trackers can be influenced by varying light conditions. Moreover, eye trackers have usually been intended for desktop settings, where a single user interacts at a time from the same distance. However, public displays expect multiple users of different heights to interact from different positions. Recent work suggested guiding passersby to certain positions in front of displays using on-screen visual cues [1].

#### **Conclusion**

In this paper we discussed why we believe gaze-based interaction to be a promising modality for tackling many challenges related to interactive public displays and introduced different gaze tracking techniques. In addition we provided pointers for future research in this area.

#### **References**

- [1] Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. GravitySpot: Guiding Users in Front of Public Displays Using On-Screen Visual Cues. In *Proc. UIST '15*. ACM.
- [2] Florian Alt, Nemanja Memarovic, Miriam Greis, and Niels Henze. UniDisplay - A Research Prototype to Investigate Expectations Towards Public Display Applications. In *Proc. PD-Apps'14*. IEEE.
- [3] Florian Alt, Alireza Sahami Shirazi, Thomas Kubitz, and Albrecht Schmidt. Interaction Techniques for Creating and Exchanging Content with Public Displays. In *Proc. CHI'13*. ACM, 1709–1718.

- [4] Harry Brignull and Yvonne Rogers. Enticing people to interact with large public displays in public spaces. In *Proc. INTERACT '03*. 17–24.
- [5] Frederik Brudy, David Ledo, Saul Greenberg, and Andreas Butz. Is Anyone Looking? Mitigating Shoulder Surfing on Public Displays Through Awareness and Protection. In *Proc. PerDis '14*. ACM, 1–6.
- [6] Nicholas S. Dalton, Emily Collins, and Paul Marshall. Display Blindness?: Looking Again at the Visibility of Situated Displays Using Eye-tracking. In *Proc. CHI '15*. ACM, 3889–3898.
- [7] Nigel Davies, Sarah Clinch, and Florian Alt. 2014. *Pervasive Displays - Understanding the Future of Digital Signage*. Morgan and Claypool Publishers.
- [8] Alexander De Luca, Martin Denzel, and Heinrich Hussmann. 2009. Look into My Eyes!: Can You Guess My Password?. In *Proc. SOUPS '09*. ACM, 7:1–7:12.
- [9] Giulio Jacucci, Ann Morrison, Gabriela T. Richard, Jari Kleimola, Peter Peltonen, Lorenza Parisi, and Toni Laitinen. Worlds of Information: Designing for Engagement at a Public Multi-touch Display. In *Proc. CHI'10*. ACM, 2267–2276.
- [10] Wendy Ju and David Sirkin. Animate Objects: How Physical Motion Encourages Public Interaction. In *Persuasive Technology*. Springer, 40–51.
- [11] Mohamed Khamis, Florian Alt, and Andreas Bulling. A Field Study on Spontaneous Gaze-based Interaction with a Public Display using Pursuits. In *Proc. UbiComp '15 Adjunct*. ACM.
- [12] Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Gonçalves, and Timo Ojala. What Makes You Click: Exploring Visual Signals to Entice Interaction on Public Displays. In *Proc. CHI '13*. ACM, 1699–1708.
- [13] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human-Computer Interaction*. Springer Publishing, 39–65.
- [14] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. Rethinking 'Multi-user': An In-the-wild Study of How Groups Approach a Walk-up-and-use Tabletop Interface. In *Proc. CHI '11*. ACM.
- [15] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and Design Space for Interactive Public Displays. In *Proc. MM '10*. ACM, 1285–1294.
- [16] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proc. CHI'12*. ACM, 297–306.
- [17] Takashi Nagamatsu, Kaoruko Fukuda, and Michiya Yamamoto. Development of Corneal Reflection-based Gaze Tracking System for Public Use. In *Proc. PerDis '14*. ACM, 194–195.
- [18] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proc. UIST '13*. ACM, 261–270.
- [19] Linda E. Sibert and Robert J. K. Jacob. Evaluation of Eye Gaze Interaction. In *Proc. CHI '00*. ACM.
- [20] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-Device Gaze-Supported Point-to-Point Content Transfer. In *Proc. ETRA '14*. 19–26.
- [21] Roel Vertegaal. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.
- [22] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proc. UbiComp '13*. ACM, 439–448.
- [23] Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proc. UbiComp '14*. ACM, 559–563.





# 7

**EyePACT**

**Eye-Based Parallax Correction on  
Touch-Enabled Interactive Displays.**

## EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays

MOHAMED KHAMIS, Ubiquitous Interactive Systems Group, LMU Munich, Germany

DANIEL BUSCHEK, Ubiquitous Interactive Systems Group, LMU Munich, Germany

TOBIAS THIERON, Ubiquitous Interactive Systems Group, LMU Munich, Germany

FLORIAN ALT, Ubiquitous Interactive Systems Group, LMU Munich, Germany

ANDREAS BULLING, Max Planck Institute for Informatics, Saarland Informatics Campus, Germany

The parallax effect describes the displacement between the perceived and detected touch locations on a touch-enabled surface. Parallax is a key usability challenge for interactive displays, particularly for those that require thick layers of glass between the screen and the touch surface to protect them from vandalism. To address this challenge, we present EyePACT, a method that compensates for input error caused by parallax on public displays. Our method uses a display-mounted depth camera to detect the user's 3D eye position in front of the display and the detected touch location to predict the perceived touch location on the surface. We evaluate our method in two user studies in terms of parallax correction performance as well as multi-user support. Our evaluations demonstrate that EyePACT (1) significantly improves accuracy even with varying gap distances between the touch surface and the display, (2) adapts to different levels of parallax by resulting in significantly larger corrections with larger gap distances, and (3) maintains a significantly large distance between two users' fingers when interacting with the same object. These findings are promising for the development of future parallax-free interactive displays.

CCS Concepts: • **Human-centered computing** → **User interface design**; **User centered design**;

Additional Key Words and Phrases: Public Displays; Parallax; Touch screens; Gaze

### ACM Reference Format:

Mohamed Khamis, Daniel Buschek, Tobias Thieron, Florian Alt, and Andreas Bulling. 2017. EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 146 (December 2017), 18 pages. <https://doi.org/10.1145/3161168>

## 1 INTRODUCTION

Interactive displays are becoming ubiquitous and are increasingly deployed in public areas, such as universities, train stations, airports, or shopping malls. While advances in sensing technology enable new types of interaction, such as with mid-air gestures [29] or gaze [20], touch remains among the most commonly employed modalities [4]. In the early days of touch screens, infrared sensors were mounted above the screen surface to detect where the user's finger was touching. Hence, a relatively large distance existed between the touch screen and the display

Authors' addresses: Mohamed Khamis, Ubiquitous Interactive Systems Group, LMU Munich, Amalienstr. 17, Munich, Bavaria, 80333, Germany; Daniel Buschek, Ubiquitous Interactive Systems Group, LMU Munich, Amalienstr. 17, Munich, Bavaria, 80333, Germany; Tobias Thieron, Ubiquitous Interactive Systems Group, LMU Munich, Amalienstr. 17, Munich, Bavaria, 80333, Germany; Florian Alt, Ubiquitous Interactive Systems Group, LMU Munich, Amalienstr. 17, Munich, Bavaria, 80333, Germany; Andreas Bulling, Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Saarland, 66123, Germany.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2017 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

2474-9567/2017/12-ART146 \$15.00

<https://doi.org/10.1145/3161168>

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 4, Article 146. Publication date: December 2017.

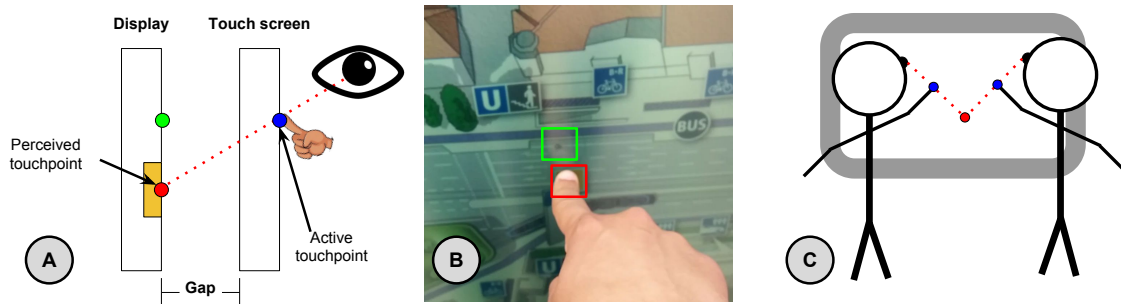


Fig. 1. The parallax effect commonly occurs on public displays where a protective glass layer separates the display and the touch screen (A) and results in the system recognizing a touch location (marked in green) that deviates from the intended one (red). To compensate for this error, EyePACT determines the expected touch location based on the position of the user's eyes and the touch point on the screen (blue) to trigger it at the perceived location (red).

that led to the so-called parallax effect [34], i.e., a difference between the perceived location at a particular UI element and users' actual touch location (see Figure 1A). This difference becomes larger as the angle between the display and the user's eyes increases (e.g., when looking at far corners of the display). The parallax effect can trigger unwanted input actions that can negatively impact usability and user experience [31].

Advances in capacitive touch screens have resulted in the screens becoming more compact, thereby reducing the parallax effect. However, the effect has become prominent again with the widespread use of public displays (see Figure 2). The reason is that on one hand, manufacturers typically protect displays against vandalism by using thick layers of glass. On the other hand, displays are often deployed behind shop windows and touch input is enabled through a transparent foil attached to the window. This increases the distance between touch screen and display and, in turn, introduces parallax (cf. Figure 1B).

Compensating for the error caused by the parallax effect is important because it makes interaction cumbersome and time-consuming which can, in turn, lead to users abandoning the displays [13, 44]. The negative impact on user experience is particularly prominent in interfaces that require several accurately located touch events, where an inaccurate touch event could result in unintended input [31]. Examples of such interfaces include interactive maps (see Figure 1B), on-screen keyboards used to provide text input or passwords (see Figure 2), calendars for choosing departure dates, etc. These interfaces would benefit from parallax correction to overcome inaccurate selections which impacts the user experience. Although a common approach is to use larger buttons aimed at covering the active and perceived touch areas, the use of larger buttons may not be feasible in the aforementioned applications. On the other hand, software-based solutions, such as training and using offset regression models like the one we discuss in Section 5, would not compensate for parallax accurately without knowledge of the user's height and position, as well as the number of users. Note that public displays often expect multiple users interacting simultaneously [11, 19, 29].

In this work we propose EyePACT, a simple, yet effective, method to overcome the parallax effect. The technique corrects for the parallax by estimating a gaze vector that intersects the user's touch point, hence estimating where on the screen the user *intended* to touch. Our approach relies on a depth camera, hence EyePACT is unlikely to require a significant cost, particularly since an increasing number of displays has integrated cameras to measure audience attention [2, 35], for security purposes (e.g., ATMs), and for interaction [29]. While similar ideas have been described in some patents [3, 5, 6, 18, 40], eye-based parallax compensation has never been evaluated and it therefore remains unclear how well such compensation can work and how well it is perceived by users. This work makes the following contributions: (1) We introduce the concept and implementation of

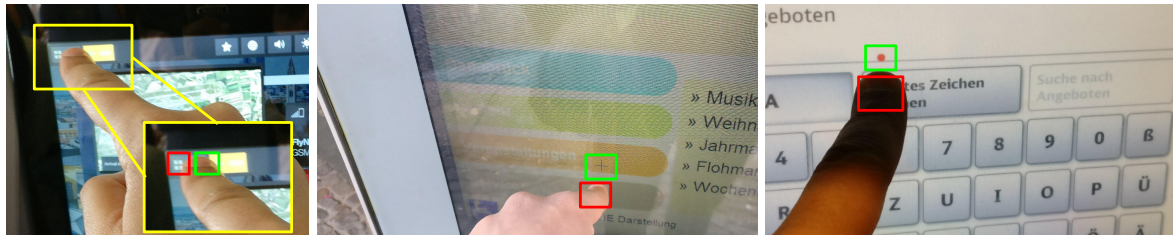


Fig. 2. The parallax effect has recently become prominent again with the widespread use of public displays that are augmented with thick layers of glass to protect against vandalism. The figure shows multiple cases where the parallax effect is prominent on real-world deployments. The red area is the touched area from the user's perspective, while the green one is the area the system mistakenly activated due to parallax.

EyePACT, a simple yet effective approach for compensating for parallax. (2) We report on our findings of an accuracy evaluation of EyePACT, which show that EyePACT brings the touch points significantly closer to the target even with varying degrees of parallax. (3) We demonstrate the effectiveness of EyePACT in allowing multiple users to interact on parallax-enabled displays without overlapping fingers.

## 2 BACKGROUND AND RELATED WORK

Our work builds on three strands of prior research, namely (1) parallax in HCI research, (2) parallax on touch-based public displays, and (3) eye detection on public displays.

### 2.1 Parallax in HCI research

While we focus on one way the parallax effect could occur, there are many conditions that could lead to a similar effect. As the distance between a touch surface and a display increases, so does the discrepancy between what each of the eyes sees. This is often referred to as binocular parallax [25]. Previous work showed that when pointing at 3D stereoscopic objects, users touch between their dominant and non-dominant eye position with a bias towards the dominant eye position [39]. Addressing binocular parallax is out of the scope of this work. (1) In our work, we focus on the parallax effect on public displays, where an increase in the distance between the touch surface and the display is rarely more than 7 cm [14, 15, 30, 49]. On the other hand, binocular parallax occurs at larger distances (e.g., 30 cm between the touchscreen and the UI element [25]). (2) Employing gaze or mid-air gestures would be more suitable for contexts where the public display is physically unreachable or too far away from the user [4, 20]. And (3), our participants did not report experiencing binocular parallax in our setup.

In contrast, motion parallax refers to the depth cues that humans perceive while moving; as we move or as objects move around us, we perceive closer objects to be moving faster than farther ones. Humans use this cue (i.e., the perceived speed of objects) to estimate how far the objects are compared to each other [32]. Parallax is often exploited for simulating depth perception by using so-called parallax barriers [23], which enables stereoscopic rendering by limiting the observer's point of view [41]. The idea is to use two stacked screens, with the front one showing an array of pinholes. By doing so, the setup exploits binocular and motion parallax cues to present the user with the illusion of depth.

### 2.2 Parallax in Public Displays

In contrast to the aforementioned types of parallax, we focus on the parallax error induced by large distances between the touchscreen and the display, which results in a displacement between the perceived and detected touch point on a touch-enabled public display. This was previously referred to as “parallax error” [28] or “visual

parallax” [24]. Parallax was a major issue in touch-based interaction in the early days of touchscreens [34]. The problem was significantly reduced with the advancements in Gorilla Glass and thin touchscreens. However public displays (see Figure 1B) are still affected by parallax due to the use of vandalism proof glass which increases the distance between the touch layer and the screen [14, 15, 30, 49].

A common approach used frequently for ATMs is to use larger buttons aimed to cover the active and perceived touch areas. Another approach is to incline the display to reduce the angle relative to the user’s view and hence reduce the parallax effect. Forcing designers to use bigger buttons is not an optimal solution and might not be feasible in interfaces where there is multitude of accurately located touch events (e.g., interactive map, month calendar, or on-screen keyboard). While inclining the display is not always feasible, in particular with large displays. Previous patents discussed solutions [3, 5, 6, 18, 40]. These methods utilize a front facing camera of a mobile device to estimate the viewing vector. However, it is not clear how well these methods perform and how they are perceived by users. Moreover, the use case covered by these patents involve a single user interacting with a touchscreen of a mobile device. In contrast, we study parallax correction on public displays for single and multiple users. Migge et al. presented a parallax-correction approach that relied on a Wiimote’s tracking IR-markers that are attached to the user’s head [28]. On the other hand, our approach does not augment the user but relies only on a remote camera, which is readily integrated into existing public displays.

### 2.3 Eye and Attention Detection on Public Displays

To detect the user’s eye position in front of a public display, it is essential to first detect the user’s face. Several approaches have been proposed for face detection, examples are the active shape model (ASM) [10], the active appearance model (AAM) [9], the gradient-based approach [38] and the Viola-Jones detection framework [43]. After detecting the user’s face, face landmarks are determined. These landmarks include features of the user’s eyes (e.g., eye corners, pupils, etc.). In our work, we use an optimized version of the AAM algorithm that incorporates depth-based face segmentation by using an RGBD camera (kinect) [36].

There is an increasing interest in the detection of eyes in the vicinity and attention to public displays [2, 35, 37]. Determining where on the screen the user is looking requires estimating a gaze vector. A time-consuming and tedious calibration process preceding the actual interaction is required to obtain highly accurate gaze estimates [27]. Therefore, recent work either proposed to blend calibration into the application (e.g., as users read text [21]), or to use gaze interaction techniques that do not require accurate gaze estimates and are therefore calibration-free [42, 47].

In summary, while highly accurate gaze estimates would enable perfect parallax correction, current state-of-the-art techniques either require calibration or estimate an inaccurate gaze vector. Instead, our approach does not require an accurate gaze estimate but uses the position of the eyes in 3D space in front of the display to generate a vector that intersects the user’s touch point and the display. The intersection is deemed to be the point the user is trying to touch.

## 3 EYEPACT– CONCEPT AND IMPLEMENTATION

To study the parallax effect, it was necessary to build an apparatus where the distance between the display and the touch screen is adjustable, hence recreating and controlling the parallax effect (see Figure 3). To do that, we laminated a touch foil<sup>1</sup> on a 110 cm × 60 cm × 8 mm Plexi glass. Figure 3 shows how we fixed the laminated touch foil to the display while keeping the distance between them adjustable. In particular, we fixed four metal rods into the stand of a 42 inch display (1366 × 768 pixels). We then used four 3D-printed holders to hold the display to the rods. The gap distance can be adjusted by sliding the rods and holders.

<sup>1</sup><http://www.visualplanet.biz/touchfoil/>

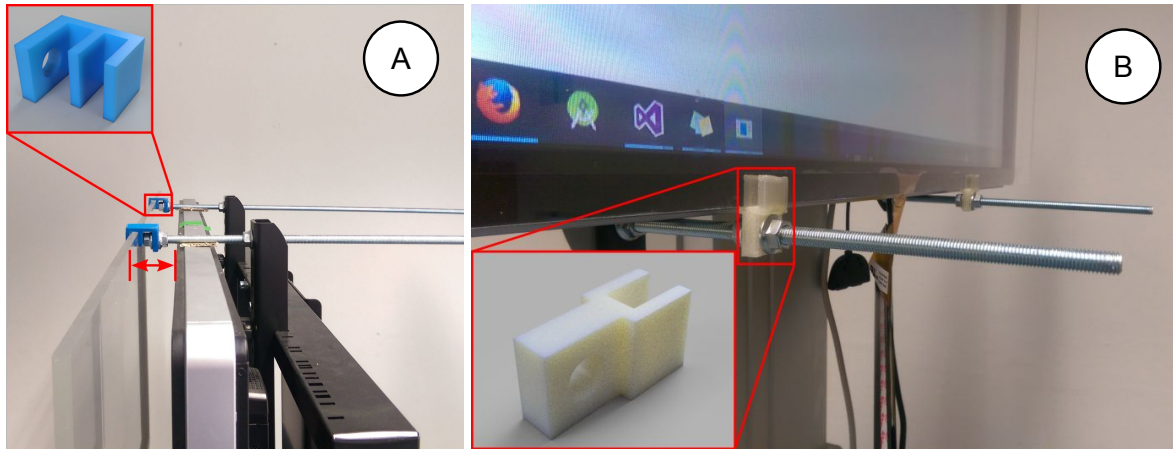


Fig. 3. The touchscreen was mounted parallel to the display using rods that are fixed to a 42-inch display stand. We used 3D-printed holders to hold the touchscreen to the rods. The upper rods (A) are adjustable by unloosing the nuts holding them to the stand, while lower ones (B) can be adjusted by unloosing the nuts holding them to the rods.

Our approach does not utilize accurate gaze point estimation and hence does not require calibration. Instead, a depth camera locates the user's eyes in the 3D space in front of the display. To do this, we mounted a Kinect 360 on a tripod positioned behind the display. We continuously track the user's pupils by employing the approach by Smolyanskiy et al., which augments the Active Appearance Model (AAM) algorithm by using depth-based face segmentation detected by an RGBD camera (kinect) [36].

Whenever a touch is detected on the touchscreen ( $Active_{2D}$ ), the point is converted from a 2D point on the touchscreen surface to a 3D point in the 3D space in front of the display ( $Active_{3D}$ ). Afterwards, a 3D ray is extended from the middle of both eyes and intersecting point  $Active_{3D}$  to eventually intersect the display at point  $Perceived_{3D}$ .  $Perceived_{3D}$  is then converted from a 3D point to a 2D point ( $Perceived_{2D}$ ) on the display's surface. The system blocks the original touch event that is orthogonal to  $Active_{2D}$  (green dot in Figure 1A), and triggers a touch event at  $Perceived_{2D}$  instead (red dot in Figure 1A). Visual feedback appears where  $Perceived_{2D}$  is triggered.

When correcting parallax for multiple users (Figure 1C), the system has to determine which user performed the action to estimate the perceived point from his/her perspective. A straightforward approach might seem to correct for the user closest to the target. This approach would fail if, for example, a user extends his/her arm to touch a point that is closer to another user.

Instead, EyePACT determines which pair of eyes to correct for depending on the closest arm to the display at the time of the touch event. This is done by utilizing skeletal tracking [48], which provides the position of users' joints in the 3D space in front of the display. The positions of the hand, wrist and elbow joints are then used to determine which user is closest to the touchscreen at the time of the touch action. Once the user is determined, EyePACT corrects for parallax from that user's perspective.

#### 4 STUDY1: ACCURACY EVALUATION

To evaluate EyePACT's parallax correction accuracy we asked participants to touch the center of crosshairs with different levels of parallax, target states, and user positions (Figures 4A and 4B).



146:6 • M. Khamis et al.

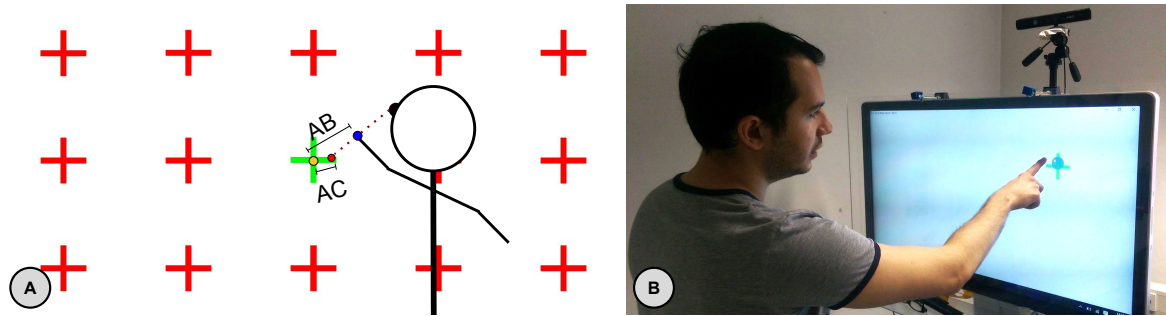


Fig. 4. In the accuracy evaluation study, participants were asked to touch the center of multiple crosshairs. (A) In the stationary target condition, one out of 15 equally distributed crosshairs was highlighted at a time. Participants had to aim at touching the center of the highlighted one (yellow dot) twice before another untouched crosshair was highlighted. The blue dot is the uncorrected touch point ( $Active_{2D}$ ), while the red dot is the corrected touch point inferred by our system ( $Perceived_{2D}$ ). (B) In the moving target condition, participants had to touch a moving crosshair, that bounced against edges, twice before it reappeared at a different starting position. We found that EyePACT results in significantly shorter distances to the target (Line AC) compared to the uncorrected distances (Line AB).

#### 4.1 Design

We experimentally manipulated several independent variables that could impact the performance of EyePACT:

- (1) The user's state: the participant could be stationary (standing 50 cm away from the display center) or move freely.
- (2) The target's state: the target could be (T1) stationary or (T2) moving (Figures 4A and 4B respectively).
- (3) The gap distance: the distance between the touch screen and the display could be (G1) short, (G2) medium, or (G3) long (4 cm, 5.4 cm and 6.8 cm respectively).

The experiment was designed as a repeated measures study, where all participants performed all conditions. To make sure participants are not influenced by the predefined *stationary user* position, all participants started with the *moving user* condition. To minimize learning effects, we alternated the order at which participants performed the conditions of the target state. For example, the first participant started with the stationary target condition, while the second user started with the moving one. The same was done for the third variable.

In the stationary target condition, a random crosshair out of 15 equally distributed ones was highlighted at a time. Participants had to touch the center of the highlighted one (yellow dot in Figure 4A) twice before another untouched crosshair was highlighted. In the moving target condition, participants had to touch a moving crosshair twice before it reappeared at a different starting position. Each participant touched 15 moving crosshairs twice. All crosshairs moved at a constant speed, but in different directions with different starting positions.

We covered three gap distances to experiment with different levels of parallax. The manufacturer's recommendation is to use a distance of 4.0 cm between the display and the touch screen to overcome the electrical noise caused by the display, which could interfere with the capacitive touch sensors. We used this as our minimum distance in addition to two larger distances: 5.4 cm and 6.8 cm. These values were determined by adding the recommended air-gap distance and the typical thicknesses of current vandalism-resistant glass. For example, at the time of submission of this paper, companies manufacture security glass that range in thickness from 6.5 mm to 25 mm [14], and from 11.5 mm to 39 mm [30].

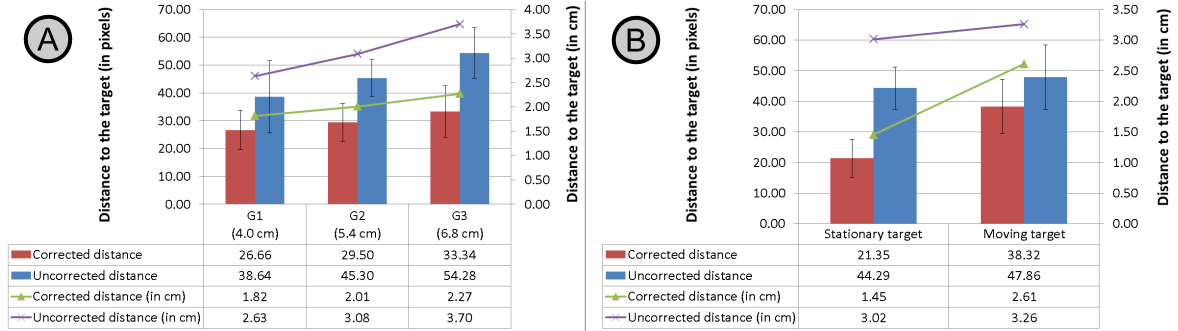


Fig. 5. Figure A shows that: (1) The uncorrected distance (blue) increases depending on the gap distance. This indicates that parallax was successfully created in our setup. (2) The corrected distance to the target (red) is always shorter than the uncorrected one. This means that EyePACT significantly improves accuracy for all gaps. (3) The portion of the uncorrected distance that is decreased (i.e., difference between the red and blue bars) becomes larger as the gap distance increases. This means that EyePACT's correction adapts to the size of the gaps. Figure B shows that although EyePACT reduces the distance to both moving and stationary targets significantly, the improvement is significantly higher with stationary targets than with moving ones.

## 4.2 Participants

We recruited 18 participants (5 females) with ages between 21 and 36 ( $M = 25.2$ ,  $SD = 4.4$ ) through mailing lists and social network groups. The height of the participants ranged from 172 cm to 189 cm ( $M = 178.3$ ,  $SD = 5.1$ ). They were compensated with online shop vouchers or participation points.

## 4.3 Procedure

The experimenter explained the study. Each participant performed 3 blocks with each block covering one of the three gap distances and 4 conditions (2 user states  $\times$  2 target states), amounting to 12 conditions  $\times$  15 targets  $\times$  2 touches = 360 touches. The study was concluded with a questionnaire and a semi-structured interview.

## 4.4 Results

At every touch event, we logged the uncorrected touch point, the corrected touch point, and the center of the target, which correspond to the blue dot, red dot, and yellow region/dot in Figures 1A and 4A respectively. We logged 7406 touch points – slightly more than what we expected, because participants sometimes performed more than two touches when uncertain if they hit the target. This mostly happened for moving targets. To address this we measured the average Euclidean distance between the points and the target for each condition. We compared the Euclidean distance between the uncorrected touch point and the target (Line AB in Figure 4A) and between touch point corrected by EyePACT and the target (Line AC in Figure 4A). Before analyzing the data, we excluded 122 out of 7406 measurements as outliers ( $> \mu + 2.5 \times StandardDeviation$ ).

For analysis we used repeated measures ANOVA to test for significance in case of parametric data. Post-hoc pair-wise comparisons were done using t-tests. For non-parametric data (e.g., not normalized), we used Friedman's test for significance, and Wilcoxon signed-rank tests with for pair-wise comparisons. In all cases, the p-value was corrected using Bonferroni correction to counteract the multiple comparisons problem.

*Effect of Gap Distance on Accuracy.* A significant main effect was found for the gap distance on the distance to the target  $F_{2,34} = 22.28$ ,  $p < 0.001$ . Post-hoc analysis revealed significant differences between all pairs ( $p < 0.05$ ). This means that our setup successfully recreated a parallax effect that significantly affects accuracy.



*Effect of EyePACT Correction on Accuracy.* A significant main effect was found for EyePACT correction on distance to target  $F_{1,17} = 87.06, p < 0.001$ . Post-hoc analysis showed a significant difference ( $p < 0.001$ ) between corrected distance ( $M = 29.84, SD = 13.15$ ) and uncorrected distance ( $M = 46.07, SD = 15.08$ ). This means that EyePACT results in a significantly shorter distance to the target.

*Effect of EyePACT Correction on Accuracy With Respect to the Gap Distances.* To study the effect of the gap distance further, we ran paired-samples t-tests and found that the corrected distance to the target is significantly shorter compared to the uncorrected one in all gap distances: G1  $t(17) = -6.059, p < 0.001$ , G2  $t(17) = -8.154, p < 0.001$  and G3  $t(17) = -8.89, p < 0.001$ .

We calculated the difference between the corrected and uncorrected distances (*Diff*) which is represented by the green bar in Figure 5A. We also found that *Diff* is significantly different depending on the gap distance  $\chi^2(2) = 20.333, p < 0.001$ . Post-hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction showed significant differences between all pairs ( $p \leq 0.01$ ). The largest correction was apparent in  $Diff_{G3}$  ( $M = 20.94, SD = 9.99$ ), followed by  $Diff_{G2}$  ( $M = 15.81, SD = 8.22$ ) then  $Diff_{G1}$  ( $M = 11.97, SD = 8.38$ ). This implies that EyePACT adapts to different levels of parallax, resulting in larger corrections when the parallax effect is higher.

The results are visualized in Figure 5A, and summarized in its caption.

*Effect of Target State on Accuracy.* A significant main effect was found for target state on distance to the target  $F_{1,17} = 51.2, p < 0.001$ . Post-hoc analysis revealed that the distance in case of stationary targets ( $M = 32.82, SD = 16.73$ ) is significantly shorter ( $p < 0.001$ ) than in the case of moving targets ( $M = 43.09, SD = 14.14$ ). This motivated us to further investigate how well EyePACT performs with respect to the target state.

Running paired-samples t-tests, we found significant differences between the corrected distance ( $M = 21.35, SD = 8.92$ ) and the uncorrected distance ( $M = 44.29, SD = 14.74$ ) in the case of a stationary target  $t(17) = -13.084, p < 0.001$ , and also between the corrected distance ( $M = 38.32, SD = 11.07$ ) and the uncorrected distance ( $M = 47.86, SD = 15.1$ ) in the case of a moving target  $t(17) = -4.416, p < 0.001$ . This means that EyePACT significantly reduces the distance to stationary and moving targets. However the improvement is significantly higher in case of a stationary target. Results are visualized in Figure 5B and summarized in its caption.

*Effect of User's Height on Accuracy.* Compared to short users, taller ones are expected to experience a stronger parallax effect due to the steeper angle to the target. Indeed, we found a strong positive correlation between the participant's height and the distance between the uncorrected point and the center of the target  $|r| = 0.535$  ( $p < 0.05$ ) – data was shown to be linear according to a Shapiro-Wilk's test ( $p > 0.05$ ). On the other hand, we did not find any indicators that the height of the user, or distance between the eyes and the screen, influences or correlates with the accuracy of EyePACT. A Pearson's product-moment correlation was run and showed a small correlation ( $|r| = 0.105$ ), that was not statistically significant ( $p > 0.05$ ), between the participant's height and the distance between the corrected touchpoint and target.

On the other hand, a repeated measures ANOVA showed significant main effect of the position of the target's row on the accuracy of EyePACT  $F_{2,34} = 10.7, p < 0.001$ . Pairwise comparisons using Bonferroni correction showed that only the accuracy at the topmost row ( $M = 18.8, SD = 4.5$ ) and at the lowermost row ( $M = 24.9, SD = 9.9$ ) are significantly different ( $p < 0.005$ ). This implies that EyePACT's accuracy drops as the target's position is towards the bottom of the display. This is because the parallax grows further as the target is farther away from the user (see Figure 6).

In summary, we neither found an influence of height on EyePACT's accuracy, nor a relationship between height and our method's accuracy. However, not finding significant differences does not mean that there is none. In fact, the significant effect of the position of the target on accuracy means that the distance between the user's eyes and the target has an influence. Hence whether the user's height influences EyePACT's accuracy remains an

open question for future research. This can be investigated by a large scale study with balanced heights, and performing statistical power analysis to estimate the probability of accepting the null hypothesis (i.e., that the user's height has no influence on accuracy) [12].

#### 4.5 Observations and Feedback

When asked about the accuracy of the system on a 5-point scale (1=very imprecise; 5=very precise), responses ranged from 2 to 5 and were more inclined towards high precision ( $M = 3.28$ ,  $SD = 0.83$ ). Almost all participants indicated to find it easier to select stationary targets, which is reflected in the quantitative analysis. P9 mentioned that touching the moving crosshair became easier over time. P3 and P4 noted that they found the system to be highly accurate. P3 was surprised the system worked despite the touch screen being far from the display. P11 and P18 remarked that they perceived no difference in accuracy between the smallest and medium gap distances. This is in-line with our findings which show that EyePACT corrections are equally good for different gap distances (see Figure 5A).

Additionally we asked participants if they attempted to compensate for the parallax effect themselves. While some needed an explanation of what parallax is, some others indicated that although they did at the beginning, they quickly realized that the system responds more intuitively if they touch where they see the target. On a 5-point scale (1=very rarely; 5=very often), responses varied from 1 to 4 and indicate that participants rarely corrected for parallax ( $M = 2.56$ ,  $SD = 1.04$ ).

Although participants were explicitly asked to move freely in the *moving user* conditions, participants often remained in the center. Some highlighted that they were too lazy to move and targets were easily reachable from the center. Others did not think that they would perform better by moving. While P12 said he would avoid walking while looking at the screen lest he trips or steps on someone's feet. This explains the lack of significant effects of user state on the distance to the target.

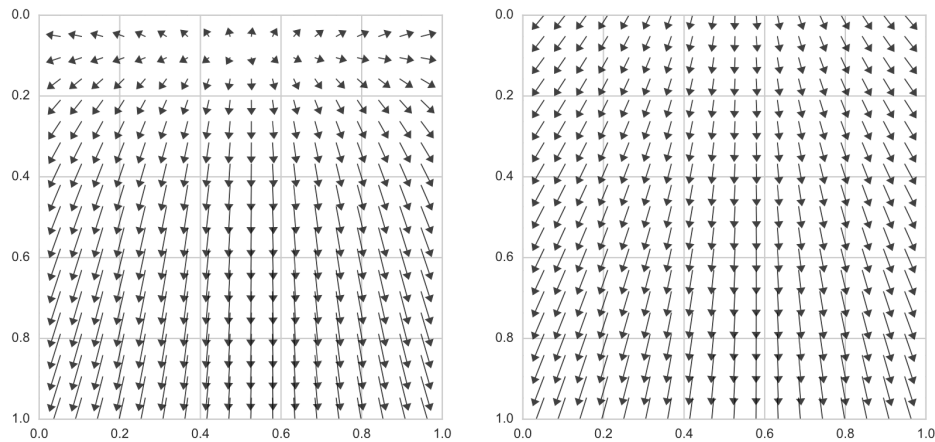


Fig. 6. Offset regression models for stationary (left) and moving targets (right). Arrows point from touch to target locations, revealing 2D offset patterns across the screen.

### 5 TOUCH OFFSET REGRESSION ANALYSIS

Beyond overall accuracy, it is also insightful to evaluate the *offset patterns* underlying the observed touch interactions. Related work analyzed such patterns on mobile devices with regression models [7, 8, 45, 46]. We also trained linear offset models (see [7, 8]) on the data of the accuracy study to analyze offset patterns.

Figure 6 visualizes models for stationary and moving targets: Users had large vertical offsets, which grow with the target location towards the bottom of the screen. Smaller horizontal offsets grow towards the left/right screen edges. This is explained by geometry – parallax grows the further users have to look down and to the side. Thus, these patterns visually explain why knowing the eye position relative to the display is highly valuable information for correcting parallax.

Since offset models map touches to targets, they can also be applied for parallax correction. We used leave-one-user-out cross-validation to train and test the models. ANOVA showed that the factor *model* (with levels: no model, hardware, software, both) has a significant effect on touch offsets (corrected by Greenhouse-Geisser:  $F_{1.92, 32.64} = 175.54, p < 0.001$ ; moving:  $F_{1.71, 29.04} = 51.29, p < 0.001$ ). Post-hoc tests revealed that both hardware and software approach significantly reduce offsets, and that combining their predictions significantly reduces offsets further. This results in a total improvement in offset RMSE of 56.6% (over the baseline) for fixed crosshairs and 29.7% for moving crosshairs.

Note, however, that the software approach requires (1) training data for the models, (2) might not work for users with heights different from those the data was trained with, and (3) cannot distinguish multiple users.

## 6 STUDY2: SUPPORTING MULTIPLE USERS

In a following step, we investigated how well EyePACT corrects parallax for multiple users and in turn allowing them to interact with the same on-screen object. We invited participants in pairs to participate in a collaborative multiplayer game where balloons would appear at the bottom and float to the top (Figure 7). The task was to collaboratively pop the balloons by (1) one player touching the balloon to *stop* it and (2) the second player touching it to *pop* it.

### 6.1 Design

This study was designed as a repeated measures experiment where all participants went through all conditions. We studied the effect of two independent variables:

- (1) Target size: we experimented with small, medium and large balloons with diameters 30, 60 and 90 pixels ( $2.15^\circ$ ,  $4.29^\circ$ ,  $6.44^\circ$  of visual angle respectively).
- (2) Gap distance: similar to the accuracy study, we experimented with 4 cm, 5.4 cm and 6.8 cm gap distances.

The order of conditions was counterbalanced using a Latin-square. Participants swapped roles, that is, if participant A *stopped* a balloon, and participant B *popped* it, the following time a balloon appeared participant B *stopped* it, while participant A *popped* it. Each pair popped 8 balloons per condition.

### 6.2 Participants

We invited 20 participants (10 pairs, 19-55 years, ( $M = 25.7, SD = 7.77$ )), 160 to 191 cm tall ( $M = 177.3, SD = 8.7$ ). Participants were recruited using mailing lists and social networks. They were compensated with online shop vouchers or participation points.

### 6.3 Procedure

Participants filled in consent forms and were explained the study. Pairs were then asked to take positions (Figure 7), which they maintained till the end of the study. Each pair performed 3 blocks with each block covering one of the three gap distances and 3 conditions. This means each pair performed 3 gap distances  $\times$  3 target sizes  $\times$  8 balloons = 72 balloon bursts.

### 6.4 Results

At each *stop* and *pop* action, we logged the uncorrected and corrected touch points. This allowed us to measure:

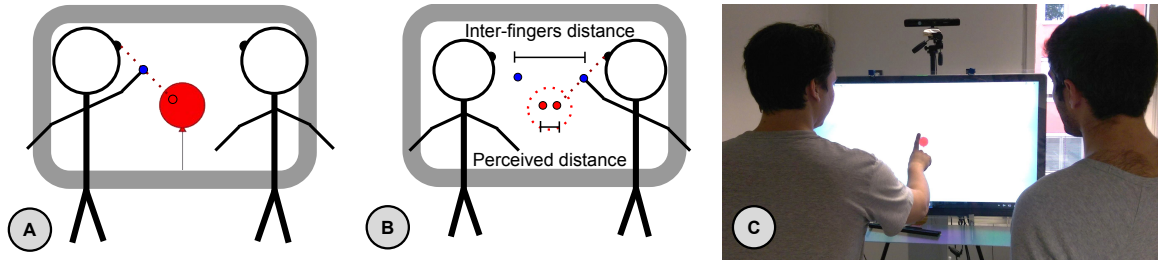


Fig. 7. We invited pairs of participants to the multiplayer study to play a multiplayer game where players collaboratively stop and pop balloons that are floating to the top of the screen. In Figure (A) the participant on the left touches the balloon to stop it, which allows the second participant (B) to pop it by touching it another time. We measured the distance between the two uncorrected points (inter-fingers distance) and between the two corrected points (perceived distance). Figure (C) demonstrates the successful parallax correction by EyePACT; although the participant is not touching the red balloon from the camera's perspective, EyePACT is accurately correcting the touch point for the left participant based on his eyes' position, allowing him to successfully touch the balloon's center from his perspective.

- The *inter-fingers distance*: the distance between the two uncorrected points.
- The *perceived distance*: the distance between the two corrected points.

We investigated how the conditions influence these two distances and the difference between them (see Figure 7B).

**Effect of Parallax Correction.** A repeated measures ANOVA showed significant main effect of EyePACT correction on the distance between the two points  $F_{1,9} = 56.61, p < 0.001$ . Post-hoc analysis using Bonferroni correction showed that the *perceived distance* ( $M = 23.26, SD = 13.07$ ) is significantly shorter than the *inter-fingers distance* ( $M = 58.12, SD = 23.83$ ). This means that EyePACT brings the two touch points significantly closer to each other when multiple participants touch the same object.

**Effect of Gap Distance.** After excluding 7 outliers ( $> \mu + 3 \times \text{StandardDeviation}$ ), a repeated measures ANOVA showed significant main effect of the gap size on the *inter-fingers distance*  $F_{2,18} = 11.2944, p < 0.005$ . Post-hoc analysis using Bonferroni correction showed that there is a significant difference ( $p < 0.005$ ) between the shortest ( $M = 45.2, SD = 14.86$ ) and the largest gap distance ( $M = 72.3, SD = 23.62$ ). The other pairs were not significantly different. Yet, Figure 8 suggests that the larger the gap distance, the larger the *inter-fingers distance*. From this we conclude that the larger the parallax, the farther users' fingers will be when touching the screen while looking at the same object.

On the other hand, no significant main effects were found for the gap distance on the *perceived distance* ( $p > 0.05$ ). This means that there is no evidence that the gap distance influences the distance between the corrected points. Figure 8 shows that the distance between the two touch points corrected by EyePACT is almost the same across the different gap distances. This suggests that EyePACT corrects equally well for multiple users at different levels of parallax.

**Effect of Target Size on Perceived and Inter-fingers distances.** A repeated measures ANOVA showed significant main effect of target size on the perceived distance  $F_{2,18} = 106.32, p < 0.001$ . Post-hoc analysis using Bonferroni correction revealed significant differences between all pairs ( $p < 0.001$ ), indicating that the distance was shortest for small targets ( $M = 12.56, SD = 4.67$ ), followed by medium targets ( $M = 22.86, SD = 7.53$ ), and longest for large targets ( $M = 34.38, SD = 9.95$ ). This is expected as both corrected touch points need to be closer to each other to fit in smaller targets. We found no significant effect of target size on distance between the two uncorrected points (inter-fingers distance). This means there is no evidence that the target size influences the distance between the fingers.

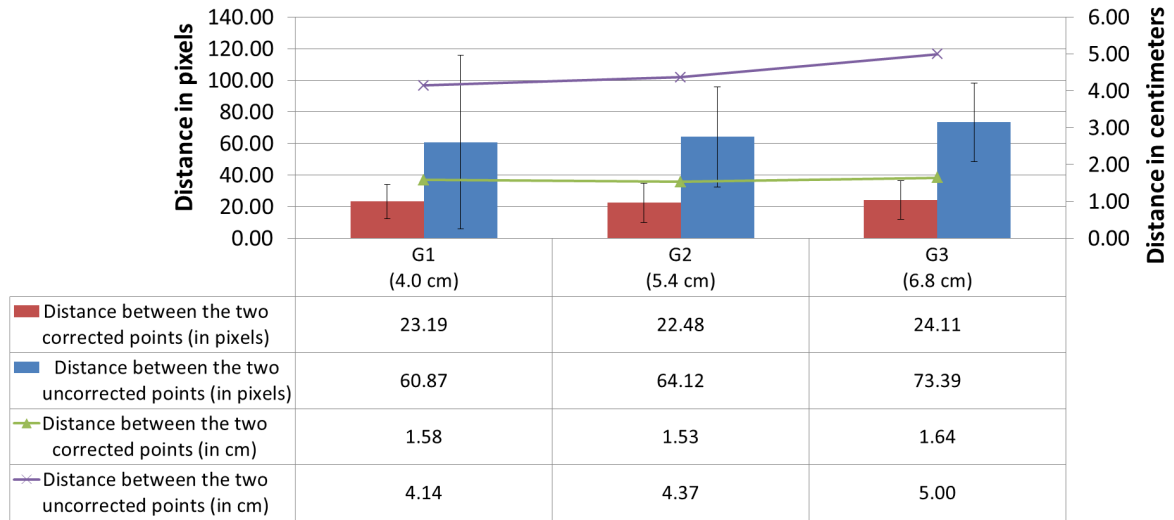


Fig. 8. The figure shows that as the gap between the touch screen and the display increases, so does the distance between the users fingers when they both touch the same object (inter-fingers distance). The figure also shows that despite the increasing gap distance, which in turn results in a stronger parallax effect, the distance between the two corrected touch points (perceived distance) does not vary a lot.

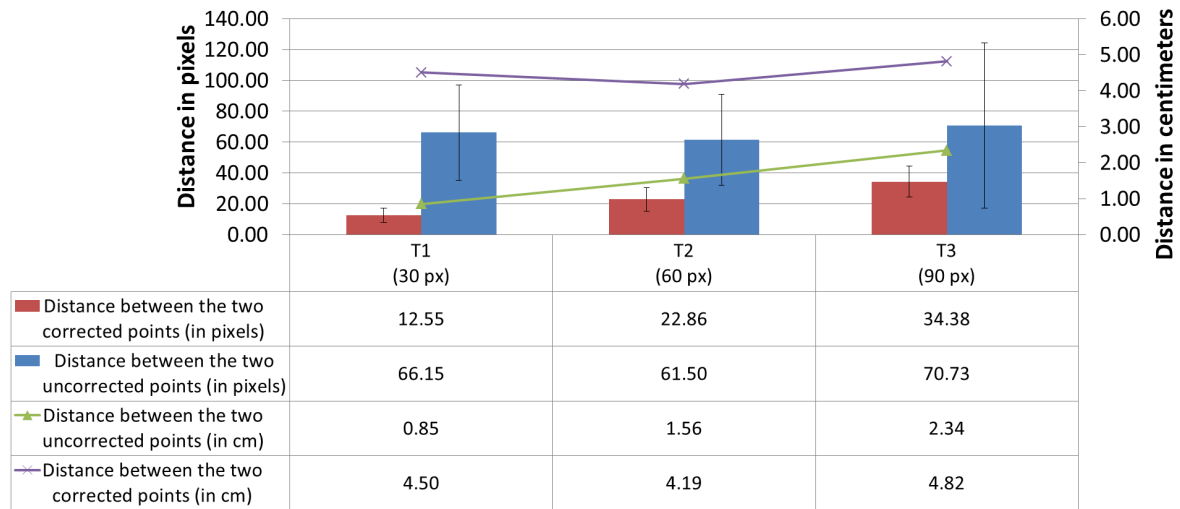


Fig. 9. The figure shows the effect of the target's size on the distance between the two corrected points (perceived distance), and the distance between the two uncorrected points (inter-fingers distance). As expected, the smaller the target, the smaller is the distance between the two corrected points within the target. The figure also suggests that the Inter-fingers distances are random and are not influenced by the target's size. The value between brackets denotes the distance in centimeters.

*Effect of Target Size on the Distance to Target Center.* Although the main aim of the second study was to evaluate EyePACT in case of multiple users, it was also possible to study the influence of the target size on EyePACT's accuracy for each individual user, since the task involved touching balloons of different sizes. We did not find any

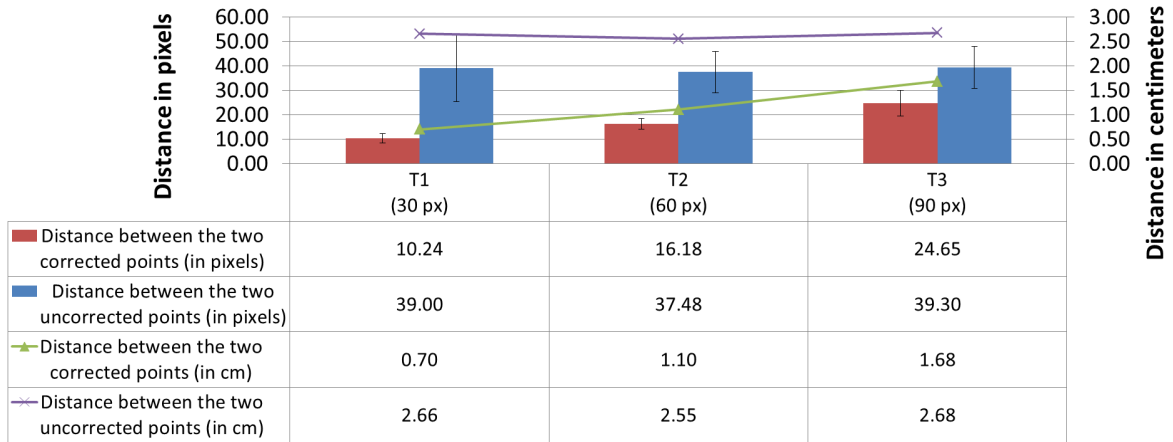


Fig. 10. The figure shows that the distance between the uncorrected touchpoint and the target center remained almost unchanged, while the distance between the corrected touchpoint and the target center is significantly shorter when touching smaller targets. This means that EyePACT adapts to smaller targets without requiring users to be more precise.

significant effect of target size on the distance between the target center and the uncorrected touchpoint ( $p > 0.05$ ). On the other hand, a repeated measures ANOVA with Greenhouse-Geisser correction revealed a significant main effect of target size on distance between the target center and the corrected touchpoint  $F_{1,2,10,9} = 39.97, p < 0.001$ . Pairwise comparisons with Bonferroni correction showed significant differences between all pairs. The distance between the corrected touchpoint and the small target ( $M = 10.16, SD = 4.63$ ) was the shortest, followed by that between the corrected touchpoint and the medium target ( $M = 15.45, SD = 8.34$ ), and finally that between the corrected touchpoint and the large target ( $M = 24.13, SD = 14.07$ ).

This means that although the distance between the uncorrected touchpoint and the target center remained almost unchanged (see Figure 10), the distance between the corrected touchpoint and the target center adapted based on the target size. This provides evidence for an additional usability advantage made available by EyePACT. That is, although smaller targets typically require more precise pointing by users, EyePACT users do not need to be more precise in order to touch smaller targets. Figure 10 illustrates this advantage: our participants did not need to be more precise when touching smaller targets compared to larger ones (blue bars); EyePACT made their touchpoints more precise and corrected them to trigger on the targets (red bars).

## 7 DISCUSSION

Our findings from the first study show that EyePACT (1) brings the touch point significantly closer to the target for all gap distances, (2) results in significantly higher accuracy for stationary targets, (3) adapts to parallax by resulting in significantly larger corrections when the gap distance is increased. The second study shows that EyePACT (1) maintains a significantly large distance between two users' fingers when interacting with the same object, and (2) is flexible with regard to targets of different sizes.

*Technical Realization.* While we are starting to see a rising adoption of interaction via mid-air gestures and gaze, touch remains the most prominent interaction technique for public displays [11]. Many of these displays have readily integrated cameras. For example, ATM machines are often equipped with cameras for security reasons. This makes the adoption of EyePACT straightforward; our approach only requires detection of the user's eyes. We use a depth camera for multi-user scenarios to identify which arm is closest to the display at the time of interaction. Although achieving the same with high accuracy using an RGB camera would result in higher



processing demand, which could in turn result in a delay in the system's response, advances in hardware and processing power make the use of an RGB camera for the same purpose also feasible in the near future.

*Eliminating the Error Completely.* Although EyePACT significantly improves touch accuracy based on the user's eye position, it does not completely eliminate the error caused by parallax. This is due to several reasons. One reason is that even with perfect eye tracking, knowing the exact pixel the user is focused on is almost impossible; users can move attention within the fovea, which is around 2 degrees of visual angle, without moving their eyes [27]. Another reason is that the user's touch tends to be biased towards the dominant eye [39]. An interesting direction for future work is to estimate the user's dominant eye based on these biases and adapt the parallax correction accordingly.

*Manual Parallax Correction and Showing Visual Feedback.* In general, users do not expect parallax correction. This was observed in our studies where several users upon the first touch overcompensated since they expected a parallax effect. In the two studies visual feedback was provided. Hence, participants quickly understood where the touch point was registered and stopped compensating. From this we learn that upon deployment in a public space, visual feedback is crucial in order not to confuse users. Furthermore, future versions of EyePACT can detect such over compensations and use them as a form of calibration to improve the correction.

*Other Application Opportunities.* In this work we focused on addressing the problem of the displacement between the perceived and detected touch points on a public display. However, this is not the only issue resulting from the large distance between the touch surface and the display. For example, binocular parallax occurs when the distance between the touchscreen and UI elements is very large (e.g., 30 cm [25]). While this is not a realistic setting for public displays that are in the scope of this work, and binocular parallax was not reported by our participants, EyePACT can be optimized to address binocular parallax as well. In particular, previous work showed that when pointing at 3D stereoscopic objects, users touch between their dominant and non-dominant eye position with a bias towards the dominant eye position [39]. Valkov et al. developed guidelines to address this issue for interacting with 3D stereoscopic objects, extrapolating their 3D solution to the 2D case could be a direction for future work to optimize EyePACT for contexts where this issue occurs (e.g., transparent displays [25, 26]).

*Reaching Farther Objects.* While the parallax effect comes with its shortcomings, it could also offer opportunities that can improve the user experience. For example, by intentionally introducing a large gap between the touch surface and the display, the angle between the user's gaze vector and the touch surface plane becomes wider. This can, in turn, reduce the user's effort to reach out to distant objects since the user does not have to be in a position that is orthogonal to the object (see Figures 1A, 1C, and Figure 7). A typical problem with touch screens is the reachability of distant content [16, 17]. This problem is amplified as large displays become more common [1, 22, 33], which results in hardly reachable on-screen objects. The use of EyePACT with a large gap can also allow users to interact with a larger portion of the display while being stationary. Nevertheless, interacting with far-away objects is likely to cause binocular parallax. Furthermore, a large gap can result in making it difficult for the user to focus on his/her fingers and the displayed object at the same time. An interesting direction for future work is to investigate starting at which gap distances focusing becomes uncomfortable for the user.

### 7.1 Limitations and Future Work

In our experiments, the speed and direction of the moving target was randomly decided at the beginning of each trial. Future work could investigate the impact of speed and direction of moving targets on the accuracy of EyePACT. Furthermore, as indicated in section 4.4, we plan to conduct a large scale study in which we balance the height of participants to better understand the impact of the user's height on the parallax error and correct it accordingly. Finally, an interesting direction for future work is to study how knowledge about the user's dominant

eye and dominant hand can help in improving the accuracy of EyePACT. In the future we plan to further extend EyePACT to address different types of parallax (e.g., Binocular parallax). Furthermore, EyePACT can be exploited to support interaction by, for example, allowing users to reach farther objects, addressing binocular parallax, and allowing novel collaborative multi-user interactions by eliminating finger interferences.

## 8 CONCLUSION

In this work we presented EyePACT, a low-computation yet effective approach for overcoming parallax on public displays. First, we discussed why parallax is an issue on public displays nowadays and the need for research in this space. Second, we described a prototype with which we created and controlled the parallax effect. Third, we described the implementation of EyePACT. Fourth, we demonstrated EyePACT's accuracy in a first user study to show that it (1) significantly improves accuracy in the presence of parallax, (2) adapts to different levels of parallax, (3) improves accuracy for both stationary and moving targets, but the improvement is more significant for stationary ones, and (4) although future work is needed to determine if it is not affected by the user's height, we did not find any significant effects in our study. Fifth, we described a software-based regression model that was trained using the data from the first study, and described why it is not sufficient in the case of users with different heights and in multiuser scenarios. Finally, in a second user study we showed that EyePACT (1) significantly improves accuracy for multiple users and also adapts to different levels of parallax, (2) enables multiple users to interact with the same objects while keeping their fingers far apart, (3) does not require users to be more precise when touching smaller targets. In addition we utilized this method to enable novel ways of interaction that are otherwise infeasible.

## ACKNOWLEDGEMENTS

This work was funded, in part, by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI) at Saarland University, Germany, and by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B).

## REFERENCES

- [1] Christopher Ackad, Martin Tomitsch, and Judy Kay. 2016. Skeletons and Silhouettes: Comparing User Representations at a Gesture-based Large Display. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2343–2347. DOI: <http://dx.doi.org/10.1145/2858036.2858427>
- [2] Florian Alt, Andreas Bulling, Lukas Mecke, and Daniel Buschek. 2016. Attention, Please!: Comparing Features for Measuring Audience Attention Towards Pervasive Displays. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 823–828. DOI: <http://dx.doi.org/10.1145/2901790.2901897>
- [3] N.R. Andrysco. 2015. Gaze-assisted touchscreen inputs. (May 7 2015). <https://www.google.de/patents/WO2015065478A1?cl=en> WO Patent App. PCT/US2013/068,125.
- [4] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. 2015. Interaction with Large Displays: A Survey. *ACM Comput. Surv.* 47, 3, Article 46 (Feb. 2015), 38 pages. DOI: <http://dx.doi.org/10.1145/2682623>
- [5] Leo B. Baldwin. 2014. Correcting for parallax in electronic displays. (Nov. 11 2014). <https://www.google.de/patents/US8884928> US Patent 8,884,928. Filed January 26, 2012.
- [6] S. Bathiche, J.R. Cheatham, P.H. Dietz, M.G. Dyor, P.A. Eckhoff, A. Gupta, K.P. Hinckley, R.A. Hyde, M.Y. Ishikawa, J.T. Kare, and others. 2015. Systems and methods for parallax compensation. (June 2 2015). <https://www.google.com/patents/US9047002> US Patent 9,047,002. Filed March 15, 2013.
- [7] Daniel Buschek and Florian Alt. 2015. TouchML: A Machine Learning Toolkit for Modelling Spatial Touch Targeting Behaviour. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 110–114. DOI: <http://dx.doi.org/10.1145/2678025.2701381>
- [8] Daniel Buschek, Simon Rogers, and Roderick Murray-Smith. 2013. User-specific Touch Models in a Cross-device Context. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 382–391. DOI: <http://dx.doi.org/10.1145/2493190.2493206>



- [9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. 1998. *Active appearance models*. Springer Berlin Heidelberg, Berlin, Heidelberg, 484–498. DOI: <http://dx.doi.org/10.1007/BFb0054760>
- [10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. 1995. Active Shape Models & Their Training and Application. *Comput. Vis. Image Underst.* 61, 1 (Jan. 1995), 38–59. DOI: <http://dx.doi.org/10.1006/cviu.1995.1004>
- [11] Nigel Davies, Sarah Clinch, and Florian Alt. 2014. *Pervasive Displays: Understanding the Future of Digital Signage* (1st ed.). Morgan & Claypool Publishers.
- [12] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2 (01 May 2007), 175–191. DOI: <http://dx.doi.org/10.3758/BF03193146>
- [13] Miriam Greis, Florian Alt, Niels Henze, and Nemanja Memarovic. 2014. I Can Wait a Minute: Uncovering the Optimal Delay Time for Pre-moderated User-generated Content on Public Displays. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1435–1438. DOI: <http://dx.doi.org/10.1145/2556288.2557186>
- [14] Double Glazing Info. 2017. Webpage. (2017). <http://www.double-glazing-info.com/Choosing-your-windows/Types-of-glass/Security-glass> Retrieved February 11, 2017.
- [15] InterelectronIX. 2017. Webpage. (2017). <https://www.interelectronix.com/en/products/touch-screen/custom/laminated-glass.html> Retrieved February 11, 2017.
- [16] Mikkel R. Jakobsen and Kasper Hornbæk. 2016. Negotiating for Space?: Collaborative Work Using a Wall Display with Mouse and Touch Input. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2050–2061. DOI: <http://dx.doi.org/10.1145/2858036.2858158>
- [17] Mikkel R. Jakobsen, Yvonne Jansen, Sebastian Boring, and Kasper Hornbæk. 2015. *Should I Stay or Should I Go? Selecting Between Touch and Mid-Air Gestures for Large-Display Interaction*. Springer International Publishing, Cham, 455–473. DOI: [http://dx.doi.org/10.1007/978-3-319-22698-9\\_31](http://dx.doi.org/10.1007/978-3-319-22698-9_31)
- [18] Van Lier Jan. 2015. Touchscreen device with parallax error compensation. (Aug. 6 2015). <https://www.google.com/patents/US20150220207> US Patent App. 14/426,105. Filed September 25, 2013.
- [19] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2015. A Field Study on Spontaneous Gaze-based Interaction with a Public Display Using Pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. ACM, New York, NY, USA, 863–872. DOI: <http://dx.doi.org/10.1145/2800835.2804335>
- [20] Mohamed Khamis, Andreas Bulling, and Florian Alt. 2015. Tackling Challenges of Interactive Public Displays Using Gaze. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. ACM, New York, NY, USA, 763–766. DOI: <http://dx.doi.org/10.1145/2800835.2807951>
- [21] Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 12. DOI: <http://dx.doi.org/10.1145/2971648.2971679>
- [22] Azam Khan, George Fitzmaurice, Don Almeida, Nicolas Burtnyk, and Gordon Kurtenbach. 2004. A Remote Control Interface for Large Displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 127–136. DOI: <http://dx.doi.org/10.1145/1029632.1029655>
- [23] Douglas Lanman, Matthew Hirsch, Yunhee Kim, and Ramesh Raskar. 2010. Content-adaptive Parallax Barriers: Optimizing Dual-layer 3D Displays Using Low-rank Light Field Factorization. *ACM Trans. Graph.* 29, 6, Article 163 (Dec. 2010), 10 pages. DOI: <http://dx.doi.org/10.1145/1882261.1866164>
- [24] David Lee, KyoungHee Son, Joon Hyub Lee, and Seok-Hyung Bae. 2012. PhantomPen: Virtualization of Pen Head for Digital Drawing Free from Pen Occlusion; Visual Parallax. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 331–340. DOI: <http://dx.doi.org/10.1145/2380116.2380159>
- [25] Joon Hyub Lee and Seok-Hyung Bae. 2013. Binocular Cursor: Enabling Selection on Transparent Displays Troubled by Binocular Parallax. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3169–3172. DOI: <http://dx.doi.org/10.1145/2470654.2466433>
- [26] David Lindlbauer, Klemen Lilija, Robert Walter, and Jörg Müller. 2016. Influence of Display Transparency on Background Awareness and Task Performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1705–1716. DOI: <http://dx.doi.org/10.1145/2858036.2858453>
- [27] Päivi Majaranta and Andreas Bulling. 2014. Eye Tracking and Eye-Based Human-Computer Interaction. In *Advances in Physiological Computing*. Springer, 39–65.
- [28] Bastian Migge and Andreas Kunz. 2010. User Model for Predictive Calibration Control on Interactive Screens. In *Cyberworlds (CW), 2010 International Conference on*. 32–37. DOI: <http://dx.doi.org/10.1109/CW.2010.18>

- [29] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 297–306. DOI : <http://dx.doi.org/10.1145/2207676.2207718>
- [30] Pilkington Optilam. 2017. Webpage. (2017). <http://www.pilkington.com/products/bp/bybenefit/safety-security/optilam/> Retrieved February 11, 2017.
- [31] Richard L. Potter, Linda J. Weldon, and Ben Shneiderman. 1988. Improving the Accuracy of Touch Screens: An Experimental Evaluation of Three Strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '88)*. ACM, New York, NY, USA, 27–32. DOI : <http://dx.doi.org/10.1145/57167.57171>
- [32] Brian Rogers and Maureen Graham. 1979. Motion Parallax as an Independent Cue for Depth Perception. *Perception* 8, 2 (1979), 125–134. DOI : <http://dx.doi.org/10.1068/p080125> arXiv:<http://dx.doi.org/10.1068/p080125> PMID: 471676.
- [33] Constantin Schmidt, Jörg Müller, and Gilles Bailly. 2013. Screenfinity: Extending the Perception Area of Content on Very Large Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1719–1728. DOI : <http://dx.doi.org/10.1145/2470654.2466227>
- [34] Andrew Sears, Catherine Plaisant, and Ben Shneiderman. 1992. Advances in Human-computer Interaction (Vol. 3). Ablex Publishing Corp., Norwood, NJ, USA, Chapter A New Era for High Precision Touchscreens, 1–33. <http://dl.acm.org/citation.cfm?id=159433.159436>
- [35] Brian A. Smith, Qi Yin, Steven K. Feiner, and Shree K. Nayar. 2013. Gaze Locking: Passive Eye Contact Detection for Human-object Interaction. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 271–280. DOI : <http://dx.doi.org/10.1145/2501988.2501994>
- [36] Nikolai Smolyanskiy, Christian Huitema, Lin Liang, and Sean Eron Anderson. 2014. Real-time 3D Face Tracking Based on Active Appearance Model Constrained by Depth Data. *Image Vision Comput.* 32, 11 (Nov. 2014), 860–869. DOI : <http://dx.doi.org/10.1016/j.imavis.2014.08.005>
- [37] Yusuke Sugano, Xucong Zhang, and Andreas Bulling. 2016. AggreGaze: Collective Estimation of Audience Attention on Public Displays. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 821–831. DOI : <http://dx.doi.org/10.1145/2984511.2984536>
- [38] Fabian Timm and Erhardt Barth. 2011. Accurate Eye Centre Localisation by Means of Gradients.. In *VISAPP*, Leonid Mestetskiy and JosÁl Braz (Eds.). SciTePress, 125–130. <http://www.inb.uni-luebeck.de/publikationen/pdfs/TiBa11b.pdf>
- [39] Dimitar Valkov, Frank Steinicke, Gerd Bruder, and Klaus Hinrichs. 2011. 2D Touching of 3D Stereoscopic Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1353–1362. DOI : <http://dx.doi.org/10.1145/1978942.1979142>
- [40] L.J. Van. 2014. Touchscreen device with parallax error compensation. (April 10 2014). <https://www.google.de/patents/WO2014053369A1?cl=en> WO Patent App. PCT/EP2013/069,895.
- [41] Klen Čopič Pucihar, Paul Coulton, and Jason Alexander. 2013. Creating a Stereoscopic Magic-lens to Improve Depth Perception in Handheld Augmented Reality. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 448–451. DOI : <http://dx.doi.org/10.1145/2493190.2494660>
- [42] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 439–448. DOI : <http://dx.doi.org/10.1145/2493432.2493477>
- [43] Paul Viola and Michael Jones. 2001. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Vol. 1. I–511–I–518 vol.1. DOI : <http://dx.doi.org/10.1109/CVPR.2001.990517>
- [44] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 841–850. DOI : <http://dx.doi.org/10.1145/2470654.2470774>
- [45] Daryl Weir, Daniel Buschek, and Simon Rogers. 2013. Sparse selection of training data for touch correction systems. In *15th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '13, Munich, Germany, August 27 - 30, 2013*. 404–407. DOI : <http://dx.doi.org/10.1145/2493190.2493241>
- [46] Daryl Weir, Simon Rogers, Roderick Murray-Smith, and Markus Löchtefeld. 2012. A user-specific machine learning approach for improving touch accuracy on mobile devices. In *The 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, Cambridge, MA, USA, October 7-10, 2012*. 465–476. DOI : <http://dx.doi.org/10.1145/2380116.2380175>
- [47] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 851–860. DOI : <http://dx.doi.org/10.1145/2470654.2470775>
- [48] Zhengyou Zhang. 2012. Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia* 19, 2 (Feb 2012), 4–10. DOI : <http://dx.doi.org/10.1109/MMUL.2012.24>
- [49] Zytronic Zybrid. 2017. Webpage. (2017). <http://zytronic.co.uk/support/product-info/zybrid/> Retrieved February 11, 2017.

## **7 EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays.**

---

146:18 • M. Khamis et al.

Received February 2017; revised August 2017; accepted October 2017

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 4, Article 146. Publication date: December 2017.



# 8

**GTmoPass**  
**Two-factor Authentication on Public Displays**  
**Using GazeTouch passwords and Personal**  
**Mobile Devices.**

# GTmoPass: Two-factor Authentication on Public Displays Using Gaze-Touch passwords and Personal Mobile Devices

Mohamed Khamis<sup>1</sup>, Regina Hasholzner<sup>1</sup>, Andreas Bulling<sup>2</sup>, Florian Alt<sup>1</sup>

<sup>1</sup>Ubiquitous Interactive Systems Group, LMU Munich, Germany

<sup>2</sup>Max Planck Institute for Informatics, Saarland Informatics Campus, Germany  
{mohamed.khamis, florian.alt}@ifi.lmu.de, bulling@mpi-inf.mpg.de

## ABSTRACT

As public displays continue to deliver increasingly private and personalized content, there is a need to ensure that only the legitimate users can access private information in sensitive contexts. While public displays can adopt similar authentication concepts like those used on public terminals (e.g., ATMs), authentication in public is subject to a number of risks. Namely, adversaries can uncover a user's password through (1) shoulder surfing, (2) thermal attacks, or (3) smudge attacks. To address this problem we propose GTmoPass, an authentication architecture that enables Multi-factor user authentication on public displays. The first factor is a knowledge-factor: we employ a shoulder-surfing resilient multimodal scheme that combines gaze and touch input for password entry. The second factor is a possession-factor: users utilize their personal mobile devices, on which they enter the password. Credentials are securely transmitted to a server via Bluetooth beacons. We describe the implementation of GTmoPass and report on an evaluation of its usability and security, which shows that although authentication using GTmoPass is slightly slower than traditional methods, it protects against the three aforementioned threats.

## Author Keywords

Multi-factor Authentication, Pervasive Displays, Eye Gestures

## ACM Classification Keywords

K.6.5 Computing Milieux: Security and Protection: Authentication

## INTRODUCTION

Public displays deliver various kinds of tangible benefits and are now deployed in train stations, airports, and streets. Meanwhile, there is an increasing demand for displays to offer personalized, context-specific content [7, 21, 23].

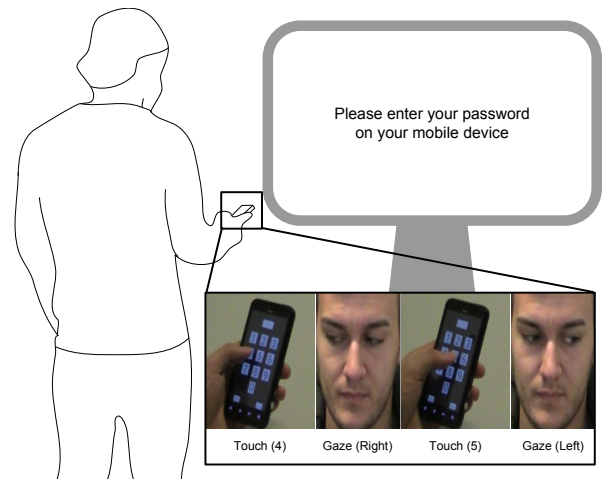
There are many cases in which users need to securely authenticate at public displays. For example, while a group of tourists examine places to visit on a large public display, the system could allow users to buy tickets for museums, buses, etc. One

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

PerDis '17, June 07-09, 2017, Lugano, Switzerland

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-5045-7/17/06... \$15.00

DOI: <http://dx.doi.org/10.1145/3078810.3078815>



**Figure 1.** We propose a Multifactor authentication architecture for public displays. Users authenticate by entering a shoulder-surfing resilient GazeTouch password (knowledge factor) on their personal mobile device (possession factor). Users simply need to launch the GazeTouch-Pass app [19] that we modified so that it would securely communicate the password to an authentication server, whose URL is received by the phone through an Eddystone-URL broadcasted using a BLE Beacon.

or more users could then authenticate in parallel by entering their passwords on their mobile devices. Using the mobile device's MAC address and the provided password, the system validates the user's credentials and charges the correct account for the ticket fee. While there are scenarios where it might be acceptable to continue the purchase and interaction on the mobile device, in many cases it is favorable to keep the user at the display to resume the primary task. In the aforementioned example, tourists could then be shown further suggestions for activities at the given place.

When exchanging sensitive data with a public display (e.g., login credentials), users are prone to several types of threats. Namely, adversaries can uncover a user's password in public space through: (1) *Shoulder surfing attacks*: observing the user while authenticating [14], (2) *Thermal attacks*: exploiting the heat traces resulting from the user's interaction with the interface [1, 24], and (3) *Smudge attacks*: exploiting the oily residues left after authenticating on touchscreens [3]. While the latter two risks were demonstrated to be feasible, shoulder surfing was shown to occur in daily contexts [14].

In this work we introduce GTmoPass, an authentication architecture that enables multi-factor user authentication on public

## 8 GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices.

displays. GTmoPass uses a shoulder-surfing resilient multimodal scheme that combines **G**aze and **T**ouch for password entry as a knowledge factor. Additionally it uses personal **m**obile devices as a possession factor (see Figure 1). After entering the password on the mobile device, the password is then securely transferred to an authentication server whose URL is communicated to the mobile device via Bluetooth beacons. The use of BLE beacons alleviates the need to manually enter URLs, or scan QR-codes. This means that when interacting with public display that employs GTmoPass for the first time, users do not have to do anything other than launching the app and entering the password.

The results of our evaluation show that users authenticate relatively fast (2.8 – 4.9 seconds), and that the authentication process is resilient to the aforementioned threats. Even if the password is leaked, the architecture of GTmoPass requires the adversary to additionally acquire the user's mobile device.

### BACKGROUND AND RELATED WORK

#### Authentication Factors

Researchers and practitioners have developed different ways for users to authenticate in order to be granted access to Private or sensitive information. Three of the most popular authentication forms are (1) knowledge-based authentication, (2) possession-based authentication, and (3) inference-based authentication (also known as biometric authentication).

#### Knowledge Factor

Knowledge-based authentication schemes rely on “something the user knows”. It is perhaps the most commonly used factor [2]. Examples are passwords, PINs, and graphical passwords. Researchers also developed ways to authenticate using eye movements [8, 12, 13, 16], mid-air gestures [25], and by recalling photographs [26]. Knowledge-based schemes allow changing passwords, and can be integrated into any system that accepts any kind of user input. On the other hand, knowledge-based passwords can be guessed by illegitimate users. Attackers can find the password by observing users during authentication [14]. Smudge attacks are possible when passwords are entered through touchscreens [3]. Graphical passwords such as Android lock patterns are particularly vulnerable to smudge attacks [29, 40]. Furthermore, many knowledge-based schemes are also vulnerable to thermal attacks, where heat traces resulting from the user's interactions with the interface are exploited to find the password [1].

#### Possession Factor

The possession factor relies on “something the user possesses”. Physical keys and scannable personal IDs are examples of possession-based authentication. Researchers experimented with identifying users on public displays through the MAC address of their smartphone [28]. Davies et al. [7] exploit the user's mobile device to identify the list of applications the user wants to interact with on a display. Others approaches include using bluetooth devices [6] and wearable shutter glasses [32]. While this type of schemes does not require users to remember passwords, it requires keeping possession of the access token. A drawback of using this factor alone, is that if an attacker

gets hold of the access token, the attacker can impersonate the user and gain unauthorized access.

#### Inference Factor

The inference factor relies on biometric data, such as fingerprints, user behavior (e.g., behavior on a touchscreen [10] or eye-gaze behavior [33]) and face detection [15]. While biometric authentication can be easy and fast to use, it is accompanied with a number of problems. Biometric passwords cannot be changed; once a user's biometric data (e.g., fingerprint or iris scan) is leaked, there is no way the user can invalidate the leaked data. Face recognition can be bypassed using pictures of the legitimate user [22], and users leave fingerprints everywhere as they interact with surrounding objects. Furthermore, users are oftentimes concerned about disclosing their biometric data to third-party companies [27], especially after it was found that biometric data can be stolen remotely [34, 42].

#### Multifactor Authentication

Multifactor authentication refers to the use of two or more of the aforementioned factors for improved security. This approach is employed by ATM machines; users have to *know* a PIN, and have to *possess* an ATM card. The approach has also been adopted by multiple Internet services, such as Google, Facebook, Microsoft, and Dropbox; users have to *know* their username and password, and have to *possess* a previously identified mobile device on which they receive an additional one-time password, or confirm their log-in attempt.

Researchers developed systems where users authenticate at ATMs by entering PINs on their mobile phones [4, 30, 31]. De Luca and Frauendienst introduced PocketPIN, where users can enter credit card data on their phones before being securely transmitted to public terminals [9].

The advantage of GTmoPass is that it employs multimodal authentication as a knowledge factor, and personal mobile devices as a possession factor. Multimodal authentication was shown to be highly resilient to shoulder surfing [19]. Furthermore, thermal and smudge attacks normally require the attacker to inspect the interface after the user had left [1, 3]. Our architecture complicates these attacks by relying on the user's mobile device for input. This means that an attacker can only perform these attacks by stealing the mobile device fast enough before the heat or smudge traces can no longer be traced. And even by doing so, the attacker would not be able to identify the gaze-input.

#### Protecting Privacy on Public Displays

In addition to the aforementioned works by Davies et al. [6, 7] and Schaub et al. [28], other works exploited proxemics for showing content on public displays. For example, Vogel and Balakrishnan show private content on public displays only when the user is very close to the display [36]. Brudy et al. proposed some concepts to hide private data on public displays by partially hiding the private data from the observer's view estimated by a Kinect [5].

#### GTMO PASS

GTmoPass is an authentication architecture that enables secure multifactor user authentication on public displays. In the



following we describe the concept and implementation of GTmoPass, and which threat models it is optimized against.

### Concept

GTmoPass relies on two factors (1) a knowledge factor: we use a multimodal authentication scheme that combines gaze and touch input, and (2) a possession factor: users enter the multimodal passwords on their personal mobile devices.

#### The Knowledge Factor

For the knowledge factor in GTmoPass, we employ a modified version of GazeTouchPass [19], a state-of-the-art authentication scheme that is highly resilient to shoulder surfing attacks. GazeTouchPass is a multimodal scheme that employs combinations of touch-based PINs (0-9) and eye movements (gaze gestures to the left and to the right). This means that it uses a theoretical password space of  $12^n$  (10 digits + 2 gaze gestures) where  $n$  is the length of the password. A password could be: Gaze (left), Touch (2), Gaze (right), Touch (1).

The strength of GazeTouchPass lies in its use of two input modalities. This adds complexity to shoulder surfing attacks, as it requires the attacker to observe (1) the user's input on the touchscreen, and (2) the eye movements of the user.

#### The Possession Factor

For the possession factor in GTmoPass, we rely on the user's personal mobile device. The multimodal passwords are entered on the mobile device; touch input is detected through the touchscreen, and gaze input is detected through the front-facing camera. The mobile device then communicates securely with an authentication server, that validates the password and signals the display to show the private information.

### Implementation

GTmoPass consists of two main components: (1) the authentication server, and (2) the user's mobile device (client).

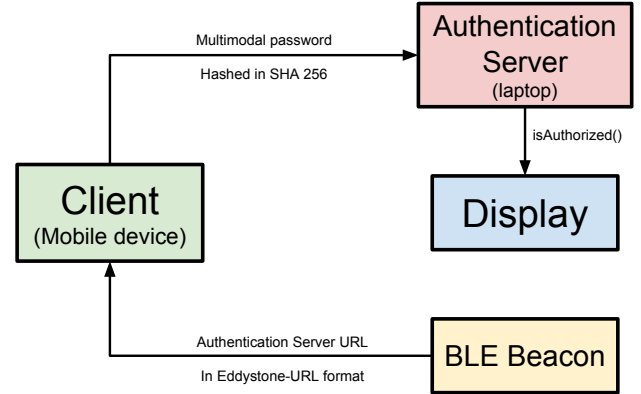
#### Authentication Server

Implemented in NodeJS, the authentication server is configured to receive HTTP requests. The server runs on a computer (e.g., in our setup we used a laptop) that is connected to a WiFi router. The IP address of the server is broadcasted using a BEEKS BLE beacon<sup>1</sup> in Google's Eddystone protocol<sup>2</sup>. The IP is broadcasted in Eddystone-URL data packets at a 10 Hz rate (i.e., it broadcasts once every 100 ms), with a transmission power of 0 dBm ( $\approx$  50 meters).

#### The Client

GazeTouchPass [19] was implemented as an Android application. It uses the OpenCV library<sup>3</sup> and the Viola-Jones classifier [35] to detect the user's face and eyes. Afterwards, in a manner similar to previous work [41, 43], the gaze direction is estimated depending on the distance between each eye and the center of the user's face.

We further extended GazeTouchPass to communicate with the authentication server. As soon as the modified app launches, it



**Figure 2.** As a user approaches the display to be authenticated at, the user would take out his/her personal mobile device and launch the modified GazeTouchPass app. The app receives the Eddystone-URL broadcasted by the beacon and presents the log in screen. The user then enters the multimodal password. The password is hashed using SHA 256 and then securely transferred to the authentication server. The server validates the log in credentials and signals the display that the credentials were correct/incorrect.

scans for beacons and connects to the nearest one that broadcasts Eddystone URLs. The URL is then saved in the local memory. Whenever the user provides four inputs, the system hashes the input using SHA 256 and sends an HTTP request to the server (see Figure 2). Similar to previous work [28], the mobile device is uniquely identified through its MAC address.

### EVALUATION

We previously evaluated the usability and observation resistance of GazeTouchPass, to find that although it requires 1.6 more seconds than PINs [39], it is significantly more secure against shoulder surfing compared to traditional PINs [19]. We also found that the structure of a multimodal password has an influence on its security. Namely, passwords that contain several switches from one modality to another, are more difficult to observe compared to those that have less switches. For example, a password such as “left-2-right-1” has three modality switches, and is hence harder to observe compared to “left-right-2-1”, which has only one modality switch. The reason is that attackers would have to switch attention between the user's fingers and the user's eyes more often in the case of passwords with more modality switches.

Another interesting insight from our previous evaluation of GazeTouchPass, is that multimodal passwords that start or end with gaze input were perceived by participants to be more difficult to observe. While our previous study was designed to focus on the effect of the number of switches in input modalities on the observability of the password, in the currently presented work we focus on the influence of the *position* of the gaze input in the password on the usability and security of the GazeTouch password.

#### Usability Study

The goals of this study are to: (1) collect feedback about the use of GazeTouchPass in the proposed setup, and (2) understand the influence of the gaze-input's *position* on usability.

<sup>1</sup><http://bluvision.com/beeks/>

<sup>2</sup><https://developers.google.com/beacons/eddytone-eid>

<sup>3</sup><http://opencv.org/platforms/android.html>



## 8 GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices.

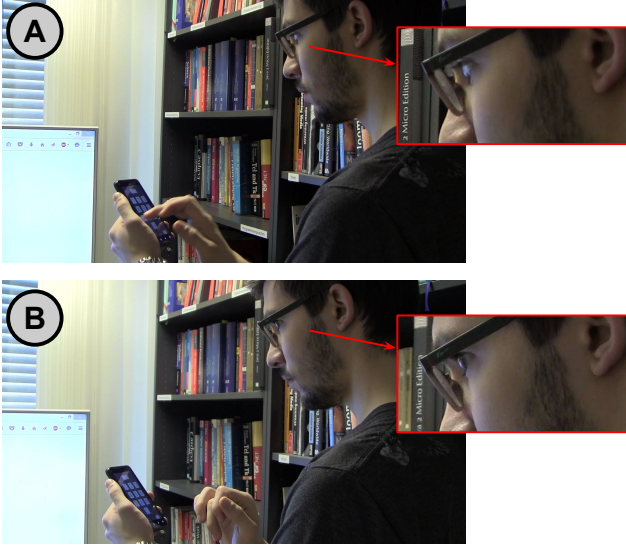


Figure 3. Participants were recorded during the usability study as they enter passwords using an HD video camera. The recorded videos were used in the subsequent security study to simulate shoulder surfing attacks. Figure A shows a user entering a touch input *Touch(7)*, while figure B shows a user performing a gaze gesture *Gaze(Left)*.

### Apparatus

We used a 48 inch Samsung TV (1920×1080 px) as a display. We connected a laptop computer running a NodeJS server that accepts HTTP requests. The server validates the received passwords and updates the display’s content accordingly. We recorded participants as they enter passwords from the side (see Figure 3). An HD camera was positioned such that it is close enough to show the touchscreen, and also the user’s eyes. These videos were recorded to be used in the subsequent security study, to simulate shoulder surfing attacks.

### Design

Since we wanted to investigate the influence of the position of the gaze input in the password, we experimented with four conditions: (1) passwords that start with gaze input (*GazeStart*), (2) passwords that end with gaze input (*GazeEnd*), (3) passwords that start and end with gaze input (*GazeStartEnd*), and (4) passwords with gaze input in the middle (*GazeMiddle*).

The study was designed as a repeated measures experiment. Each participant performed 16 authentications (4 passwords × 4 conditions) using randomly generated passwords. In case of conditions *GazeStart* and *GazeEnd*, participants entered two passwords with one gaze input at the start/end of the password, while the other two passwords had two gaze inputs at the start/end of the password. Table 1 shows sample passwords.

### Participants

We invited 16 participants to our lab (6 females), recruited through mailing lists and social networks. Participants were awarded with online shop vouchers or participation points. All participants had normal or corrected-to-normal vision.

### Procedure

The experimenter first described the study and asked the participants to sign a consent form. She then handed the participant a mobile device with the modified version of GazeTouchPass

Condition		Format				Example
GazeStart	startOne	Gaze	Touch	Touch	Touch	L123
	startTwo	Gaze	Gaze	Touch	Touch	RL34
GazeEnd	endOne	Touch	Touch	Touch	Gaze	943L
	endTwo	Touch	Touch	Gaze	Gaze	53RR
GazeMiddle		Touch	Gaze	Gaze	Touch	7LL3
GazeStartEnd		Gaze	Touch	Touch	Gaze	R82L

Table 1. Previous work reported that the position of the gaze input was perceived by participants to have an influence on the password’s observability [19]. Hence we experimented with the above conditions to cover possible positions of the gaze input in the password.

installed, and explained how it works. Each participant was allowed to perform a training run per condition to get acquainted with the system. The trial attempts were not included in the analysis. At each authentication attempt, the experimenter read out the password to be entered according to a previously generated list that was randomized. The participant would then enter the password, and observe the feedback on the display, which indicated whether or not the correct password was detected. Afterwards participants were interviewed to learn about their feedback, ideas and concerns about GTmoPass.

### Results

To measure the impact of the position of gaze input on the usability of the passwords, we measured the *entry time* and the *error count* while authenticating.

**Entry time** was measured starting from the first input until the last input was recognized. Figure 4 illustrates the time taken to enter a password at every condition. A repeated measures ANOVA (with Greenhouse-Geisser Correction due to violation of sphericity) showed a significant effect of the gaze input’s position on the time it takes to enter a password ( $F_{1.8,27.5} = 9.1$ ,  $p < 0.05$ ). Post-hoc analysis with Bonferroni correction ( $\alpha = 0.05 / 6$  comparisons = 0.0083) showed significant differences in entry time between *GazeStart* ( $M = 2863$  ms,  $SD = 1525$  ms) and *GazeMiddle* ( $M = 4959$  ms,  $SD = 3141$  ms), between *GazeEnd* ( $M = 3892$  ms,  $SD = 3045$  ms) and *GazeMiddle* ( $M = 4959$  ms,  $SD = 3141$  ms), and between *GazeStartEnd* ( $M = 3757$  ms,  $SD = 3852$  ms) and *GazeMiddle* ( $M = 4959$  ms,  $SD = 3141$  ms). The other pairs were not significantly different ( $p > 0.05$ ). This means that passwords with gaze in the middle are significantly slower than other cases.

**Error count** reflects the number of times the participant entered the password incorrectly. Errors could occur either due to entering the wrong gaze or touch input, or due to the system detecting an incorrect gaze input due to poor lighting conditions. Figure 5 shows the number of errors at each condition.

**Qualitative Feedback** collected at the end of the study through semi-structured interviews revealed positive feedback towards the system. Many participants reported that they liked the idea of detecting eye movements through the smartphone’s camera, and would imagine using it to authenticate on ATMs instead of using cards and PINs. Some participants suggested using the system to open security doors. One participant suggested using it for authentication on other digital devices.

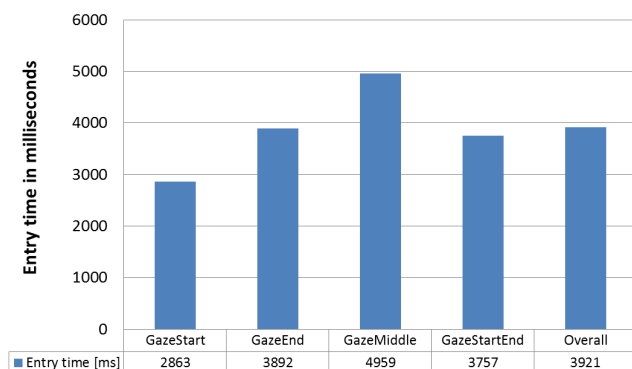


Figure 4. Mean authentication times.

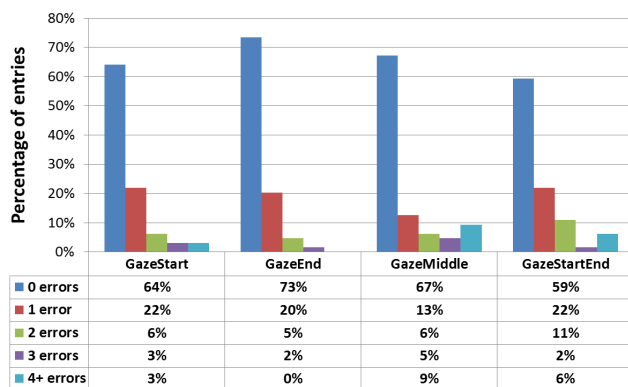


Figure 5. Number of attempts before a successful entry.

## Security Study

Unlike existing systems, GTmoPass is resilient to thermal and smudge attacks by design. Heat traces and oily residues can only uncover the *touch* part of the password, but not the gaze input. Therefore in this study we focus on GTmoPass's resilience to observation attacks.

Although previous work evaluated the impact of the number of modality switches on the security of passwords [19]. Our aim in this study was to understand the influence of gaze input's position on the observation resistance of the password.

Using the video material produced in the first study, we conducted a second *observability study* that simulated a shoulder surfing attack against a user authenticating using GTmoPass. To do this, we invited 16 different participants and asked them to simulate shoulder surfers by watching the recorded videos, and trying to find the entered passwords (see Figure 6).

### Threat Model

In our threat model, the user and the attacker are in public space. The attacker is familiar with GTmoPass and how it works. The attacker observes the user from an angle that allows seeing both the touch input on the mobile device, and the eye movements (see Figure 3). The distance between the attacker and the user is close enough to see the touchscreen, but far enough to reduce the effort of switching focus back and forth between the user's eyes and the device's touchscreen. After unveiling the password, the attacker tries to get hold of the device and authenticate at the display.

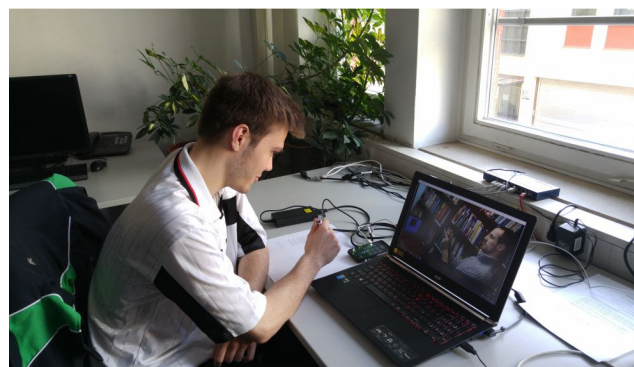


Figure 6. Participants of the security study watched HD videos of the usability study participants as they authenticated. The task was to find the correct password. Participants were explained how the system works, had a chance to try the application themselves, and were allowed to take notes while watching the videos.

### Participants and Reward mechanism

We invited 16 participants (9 females) through mailing lists and social networks. None of them had participated in the usability study. All participants were awarded either an online shop voucher or participation points. In addition, they took part in a lottery for an additional 10 EUR online shop voucher, where the chance of winning increases as the participant successfully attacks more passwords. This was done to encourage participants to put a lot of effort in finding the passwords.

### Design

This study also followed a repeated measures within subjects design. Each participant watched 4 videos of successful authentications from each condition (4 videos  $\times$  4 conditions = 16 videos in total), each of which is a recording from a different usability study participant.

### Procedure and Apparatus

After arriving at our lab and filling out the consent form, the experimenter explained the concept and the reward mechanism. The videos were displayed and controlled by the experimenter on a computer (1920 px  $\times$  1080 px). Participants were given a pen and a paper and were allowed to take notes while watching the videos (see Figure 6). Since each video was attacked once, it was watched once and hence the duration of the attack depends on the length of the video. They were also allowed to try the application themselves. After watching the video once, they provided up to three guesses, but were not told whether their guesses were correct or not to avoid influencing the perceived difficulty. We concluded with a questionnaire to learn more about the perceived difficulty of attacks.

### Results

To understand the impact of the position of gaze input on the observability of the passwords, we measured the *binary success rate* and the *Levenshtein distance*.

*Success rate* reflects how many passwords of each condition were successfully observed. On the other hand, the *Levenshtein distance* indicates how close the guess is to the original password [1, 19, 38]. Since our participants provided 3 guesses against each password they observed, only the guess

## 8 GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices.

Mean values	GazeStart	GazeEnd	GazeMiddle	GazeStartEnd
Binary success rate (0%=all incorrect;100%=all correct)	19%	16%	13%	16%
Levenshtein distance (0=similar;4=completely different)	1.36	1.18	1.21	1.27
Perceived Difficulty (1=very easy;4=very hard)	4	3	3.5	4

**Table 2.** We measured the binary success rate (i.e., whether or not the attacker’s guess is correct) and the Levenshtein distance (i.e., how similar the guesses are to the correct password). A low success rate and high Levenshtein distance indicate high observation resistance and hence higher security. The most secure passwords are the GazeMiddle, where only 13% of the passwords were observed successfully, and guesses were 55% similar to the correct ones. Participants reported the perceived difficulty of attacks before knowing whether their guesses were correct.

with the shortest distance (i.e., highest similarity) to the correct password was considered for analysis.

Table 2 shows the average success rate and Levenshtein distance for each condition. While participants succeeded the most when attacking GazeStart, their guesses had the longest average distance from the correct password. Guesses against GazeEnd were closest to the correct password (i.e., low distance) possibly because the attackers knew that each password consists of 4 inputs, and after seeing two or three touch-inputs they foresaw that the following inputs are gaze-based.

We also collected the *perceived difficulty* which participants indicated through a Likert scale (5-points;1=very easy;5=very hard). Table 2 shows that participants found all types difficult to observe, but GazeStart and GazeStartEnd were perceived to be more difficult compared to GazeEnd and GazeMiddle. These values are in-line with the Levenshtein distances.

### DISCUSSION

The proposed architecture enables usable and secure authentication on public displays. Users simply approach the display with their unmodified personal mobile device that has our app installed. The app retrieves the IP of the server that is broadcasted by the BLE beacon, allowing the user to directly authenticate without entering URLs or scanning QR codes.

A usability study shows that the use of multimodal passwords in that setup is feasible, well perceived, and is only slightly slower than the less secure PINs (von Zezschwitz et al. report 1.5 seconds for PIN-entry [39]). While users make few errors, previous work has shown that users are willing to correct errors on public displays [20]. A security study showed that authentication is robust against observation attacks (only 13%-19% successful attacks in optimal conditions), which is more secure than PINs and several recently proposed systems. For example, attacks against EyePassShapes [8], EyePIN [12], GazeTouchPass [19] and XSide [11] had success rates of 42%, 55%, 19% and 9% – 38% respectively. Furthermore, the fact that gaze-input does not leave traces on the device makes GTmoPass secure against thermal and smudge attacks even if the attacker gets hold of the mobile device.

### Trade-off between Usability and Security

While authentication using PINs is fast, it is known to be insecure and highly vulnerable to observation attacks [19, 37] and thermal attacks [1]. On the other hand authentication using multimodal passwords is more secure, but takes longer time

compared to PINs. While even a slight increase in authentication time on mobile devices has a big impact considering that users unlock their mobile devices more than 50 times a day [18], we argue that authentication on public displays does not happen as often and hence a slight increase (between 1.3 and 2.5 seconds in our case) is not very significant.

In addition to the overall trade-off, we found that having gaze-input in the middle of the password (GazeMiddle) is the least likely to be successfully attacked, but also requires the longest time to enter. In general, it was found that providing consecutive gaze inputs results in longer authentication times. This was the case in GazeStartTwo and GazeEndTwo and GazeMiddle (see examples in Table 1). This is due to the time it takes to perform a gaze gesture, look to the front again, then perform another gaze gesture. On the other hand, guesses against GazeEnd are the closest to the actual password. We expect that after observing two or three touch inputs, participants foresaw that the following inputs could be gaze-based.

### Perceived Difficulty of Shoulder Surfing

It is interesting that the perceived difficulty of attacks reported by participants was more in-line with the Levenshtein distances rather than with the binary success rate. The Levenshtein distance metric evaluates how similar a guess is to the actual password, which means that it also reflects how many times digits or gaze gestures were observed correctly. This means that unlike the binary success rate, participant’s confidence in identifying particular inputs can be a valid indicator of low Levenshtein distances.

It is not surprising that the final success rate does not correlate with the perceived difficulty. In fact, previous work reported that attackers often underestimate the perceived difficulty of shoulder surfing. For example, in the work by George et al. [17], the perceived difficulty of performing shoulder surfing attacks changed drastically after trying to perform attacks.

### CONCLUSION

In this work we showed that GTmoPass offers a secure authentication architecture for public displays. A usability and a security study showed that GTmoPass is usable and secure against shoulder surfing. We also discussed how thermal and smudge attacks are infeasible by design.

In the future, we want to evaluate more complicated threat models. For example, a combination of a thermal attack to uncover touch input and an observation attack to uncover gaze input, or multiple consecutive observations by insiders (e.g., family members or work colleagues). Another interesting threat model is the case of having two attackers: one observing the eyes, while the other observes the touchscreen. We also intend to conduct a field study to better understand how users perceive GTmoPass in the wild. A further direction for future work is to include a third inheritance factor. This can be done by scanning the finger print or by face detection using the front-facing camera.

### ACKNOWLEDGEMENT

This work was supported by a Google IoT Technology Research Award.

## REFERENCES

1. Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA. DOI: <http://dx.doi.org/10.1145/3025453.3025461>
2. M. D. H. Abdullah, A. H. Abdullah, N. Ithnin, and H. K. Mammi. 2008. Towards Identifying Usability and Security Features of Graphical Password in Knowledge Based Authentication Technique. In *2008 Second Asia International Conference on Modelling & Simulation (AMS)*. 396–403. DOI: <http://dx.doi.org/10.1109/AMS.2008.136>
3. Adam J. Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. 2010. Smudge Attacks on Smartphone Touch Screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies (WOOT'10)*. USENIX Association, Berkeley, CA, USA, 1–7. <http://dl.acm.org/citation.cfm?id=1925004.1925009>
4. Andrea Bianchi. 2011. Authentication on Public Terminals with Private Devices. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '11)*. ACM, New York, NY, USA, 429–430. DOI: <http://dx.doi.org/10.1145/1935701.1935815>
5. Frederik Brudy, David Ledo, Saul Greenberg, and Andreas Butz. 2014. Is Anyone Looking? Mitigating Shoulder Surfing on Public Displays Through Awareness and Protection. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. ACM, New York, NY, USA, Article 1, 6 pages. DOI: <http://dx.doi.org/10.1145/2611009.2611028>
6. Nigel Davies, Adrian Friday, Peter Newman, Sarah Rutledge, and Oliver Storz. 2009. Using Bluetooth Device Names to Support Interaction in Smart Environments. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services (MobiSys '09)*. ACM, New York, NY, USA, 151–164. DOI: <http://dx.doi.org/10.1145/1555816.1555832>
7. Nigel Davies, Marc Langheinrich, Sarah Clinch, Ivan Elhart, Adrian Friday, Thomas Kubitza, and Bholanathsingh Surajbali. 2014. Personalisation and Privacy in Future Pervasive Display Networks. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2357–2366. DOI: <http://dx.doi.org/10.1145/2556288.2557287>
8. Alexander De Luca, Martin Denzel, and Heinrich Hussmann. 2009. Look into My Eyes!: Can You Guess My Password?. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. ACM, New York, NY, USA, Article 7, 12 pages. DOI: <http://dx.doi.org/10.1145/1572532.1572542>
9. Alexander De Luca and Bernhard Frauendienst. 2008. A Privacy-respectful Input Method for Public Terminals. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges (NordCHI '08)*. ACM, New York, NY, USA, 455–458. DOI: <http://dx.doi.org/10.1145/1463160.1463218>
10. Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. 2012. Touch Me Once and I Know It's You!: Implicit Authentication Based on Touch Screen Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 987–996. DOI: <http://dx.doi.org/10.1145/2207676.2208544>
11. Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. 2014. Now You See Me, Now You Don't: Protecting Smartphone Authentication from Shoulder Surfers. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2937–2946. DOI: <http://dx.doi.org/10.1145/2556288.2557097>
12. Alexander De Luca, Roman Weiss, and Heiko Drewes. 2007. Evaluation of Eye-gaze Interaction Methods for Security Enhanced PIN-entry. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (OZCHI '07)*. ACM, New York, NY, USA, 199–202. DOI: <http://dx.doi.org/10.1145/1324892.1324932>
13. Alexander De Luca, Roman Weiss, Heinrich Hussmann, and Xueli An. 2008. Eyepass - Eye-stroke Authentication for Public Terminals. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, New York, NY, USA, 3003–3008. DOI: <http://dx.doi.org/10.1145/1358628.1358798>
14. Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA. DOI: <http://dx.doi.org/10.1145/3025453.3025636>
15. FaceLock. 2013. FaceLock. <http://www.facelock.mobi/>. (2013). accessed 10 January 2017.
16. Alain Forget, Sonia Chiasson, and Robert Biddle. 2010. Shoulder-surfing Resistance with Eye-gaze Entry in Cued-recall Graphical Passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1107–1110. DOI: <http://dx.doi.org/10.1145/1753326.1753491>
17. Ceenu Goerge, Mohamed Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *Proceedings of the*

## 8 GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices.

- Network and Distributed System Security Symposium (USEC '17)*. NDSS. DOI: <http://dx.doi.org/10.14722/usec.2017.23028>
18. Marian Harbach, Alexander De Luca, and Serge Egelman. 2016. The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4806–4817. DOI: <http://dx.doi.org/10.1145/2858036.2858267>
  19. Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016a. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 2156–2164. DOI: <http://dx.doi.org/10.1145/2851581.2892314>
  20. Mohamed Khamis, Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, and Florian Alt. 2016b. EyeVote in the Wild: Do Users Bother Correcting System Errors on Public Displays?. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 57–62. DOI: <http://dx.doi.org/10.1145/3012709.3012743>
  21. Thomas Kubitz, Sarah Clinch, Nigel Davies, and Marc Langheinrich. 2013. Using Mobile Devices to Personalize Pervasive Displays. *SIGMOBILE Mob. Comput. Commun. Rev.* 16, 4 (Feb. 2013), 26–27. DOI: <http://dx.doi.org/10.1145/2436196.2436211>
  22. Yan Li, Yingjiu Li, Qiang Yan, Hancong Kong, and Robert H. Deng. 2015. Seeing Your Face Is Not Enough: An Inertial Sensor-Based Liveness Detection for Face Authentication. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. ACM, New York, NY, USA, 1558–1569. DOI: <http://dx.doi.org/10.1145/2810103.2813612>
  23. Nemanja Memarovic. 2015. Public Photos, Private Concerns: Uncovering Privacy Concerns of User Generated Content Created Through Networked Public Displays. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*. ACM, New York, NY, USA, 171–177. DOI: <http://dx.doi.org/10.1145/2757710.2757739>
  24. Keaton Mowery, Sarah Meiklejohn, and Stefan Savage. 2011. Heat of the Moment: Characterizing the Efficacy of Thermal Camera-based Attacks. In *Proceedings of the 5th USENIX Conference on Offensive Technologies (WOOT'11)*. USENIX Association, Berkeley, CA, USA, 6–6. <http://dl.acm.org/citation.cfm?id=2028052.2028058>
  25. Shwetak N. Patel, Jeffrey S. Pierce, and Gregory D. Abowd. 2004. A Gesture-based Authentication Scheme for Untrusted Public Terminals. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 157–160. DOI: <http://dx.doi.org/10.1145/1029632.1029658>
  26. Trevor Perring, Murali Sundar, John Light, and Roy Want. 2003. Photographic Authentication Through Untrusted Terminals. *IEEE Pervasive Computing* 2, 1 (Jan. 2003), 30–36. DOI: <http://dx.doi.org/10.1109/MPRV.2003.1186723>
  27. Alexander P. Pons and Peter Polak. 2008. Understanding User Perspectives on Biometric Technology. *Commun. ACM* 51, 9 (Sept. 2008), 115–118. DOI: <http://dx.doi.org/10.1145/1378727.1389971>
  28. Florian Schaub, Peter Lang, Bastian Könings, and Michael Weber. 2013. PriCal: Dynamic Privacy Adaptation of Collaborative Calendar Displays. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp '13 Adjunct)*. ACM, New York, NY, USA, 223–226. DOI: <http://dx.doi.org/10.1145/2494091.2494163>
  29. Stefan Schneeeggass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2014. SmudgeSafe: Geometric Image Transformations for Smudge-resistant User Authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 775–786. DOI: <http://dx.doi.org/10.1145/2632048.2636090>
  30. Julian Seifert, Alexander De Luca, and Enrico Rukzio. 2012. Don't Queue Up!: User Attitudes Towards Mobile Interactions with Public Terminals. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*. ACM, New York, NY, USA, Article 45, 4 pages. DOI: <http://dx.doi.org/10.1145/2406367.2406422>
  31. Richard Sharp, James Scott, and Alastair R. Beresford. 2006. *Secure Mobile Computing Via Public Terminals*. Springer Berlin Heidelberg, Berlin, Heidelberg, 238–253. DOI: [http://dx.doi.org/10.1007/11748625\\_15](http://dx.doi.org/10.1007/11748625_15)
  32. Garth B. D. Shoemaker and Kori M. Inkpen. 2001. Single Display Privacyware: Augmenting Public Displays with Private Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 522–529. DOI: <http://dx.doi.org/10.1145/365024.365349>
  33. Chen Song, Aosen Wang, Kui Ren, and Wenya Xu. 2016. "EyeVeri: A Secure and Usable Approach for Smartphone User Authentication". In *IEEE International Conference on Computer Communication (INFOCOM'16)*. San Francisco, California, 1 – 9.
  34. Martin Stokkenes, Raghavendra Ramachandra, and Christoph Busch. 2016. Biometric Authentication Protocols on Smartphones: An Overview. In *Proceedings of the 9th International Conference on Security of Information and Networks (SIN '16)*. ACM, New York, NY, USA, 136–140. DOI: <http://dx.doi.org/10.1145/2947626.2951962>



- 
35. Paul Viola and MichaelJ. Jones. 2004. Robust Real-Time Face Detection. *International Journal of Computer Vision* 57, 2 (2004), 137–154. DOI: <http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb>
  36. Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 137–146. DOI: <http://dx.doi.org/10.1145/1029632.1029656>
  37. Emanuel von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. 2015. SwiPIN: Fast and Secure PIN-Entry on Smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1403–1406. DOI: <http://dx.doi.org/10.1145/2702123.2702212>
  38. Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann. 2013. *Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition*. Springer Berlin Heidelberg, Berlin, Heidelberg, 460–467. DOI: [http://dx.doi.org/10.1007/978-3-642-40477-1\\_28](http://dx.doi.org/10.1007/978-3-642-40477-1_28)
  39. Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. 2013a. Patterns in the Wild: A Field Study of the Usability of Pattern and Pin-based Authentication on Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 261–270. DOI: <http://dx.doi.org/10.1145/2493190.2493231>
  40. Emanuel von Zezschwitz, Anton Koslow, Alexander De Luca, and Heinrich Hussmann. 2013b. Making Graphic-based Authentication Secure Against Smudge Attacks. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*. ACM, New York, NY, USA, 277–286. DOI: <http://dx.doi.org/10.1145/2449396.2449432>
  41. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2014. Pupil-canthy-ratio: A Calibration-free Method for Tracking Horizontal Gaze Direction. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 129–132. DOI: <http://dx.doi.org/10.1145/2598153.2598186>
  42. Yulong Zhang, Zhaofeng Chen, Hui Xue, and Tao Wei. 2015a. Fingerprints On Mobile Devices: Abusing and leaking. In *Black Hat Conference*.
  43. Yanxia Zhang, Ming Ki Chong, Jörg Müller, Andreas Bulling, and Hans Gellersen. 2015b. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing* 19, 5 (2015), 967–981. DOI: <http://dx.doi.org/10.1007/s00779-015-0866-8>

# V

## CHALLENGES OF GAZE-ENABLED PUBLIC DISPLAYS





# 9

**EyeScout**

**Active Eye Tracking for Position and  
Movement Independent Gaze Interaction with  
Large Public Displays.**

# EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays

Mohamed Khamis<sup>1</sup>, Axel Hoesl<sup>1</sup>, Alexander Klimczak<sup>1</sup>, Martin Reiss<sup>1</sup>,  
Florian Alt<sup>1</sup>, Andreas Bulling<sup>2</sup>

<sup>1</sup>Ubiquitous Interactive Systems Group, LMU Munich, Germany

<sup>2</sup>Max Planck Institute for Informatics, Saarland Informatics Campus, Germany  
{mohamed.khamis, axel.hoesl, florian.alt}@ifi.lmu.de, bulling@mpi-inf.mpg.de

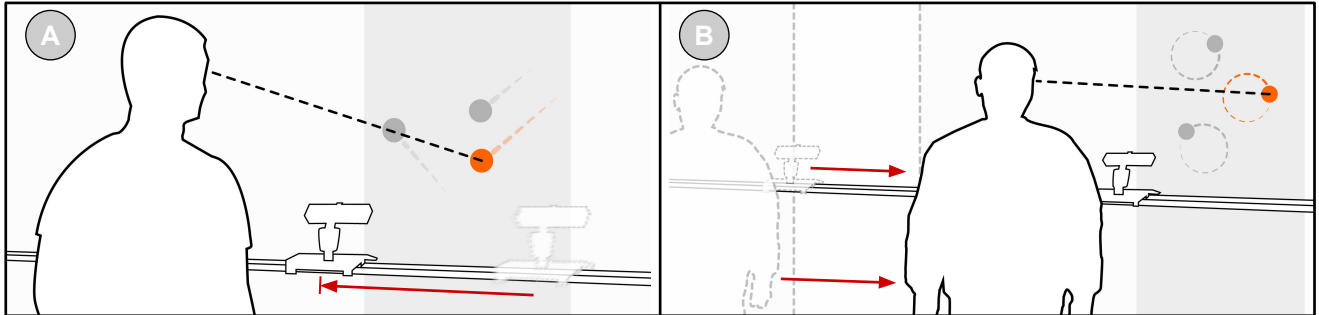


Figure 1. EyeScout is an active eye tracking system that enables gaze interaction with large public displays. It supports two interaction modes: In “Walk then Interact” the user can walk to a location in front of the display and the system positions itself accurately to enable gaze interaction (A). In “Walk and Interact” the user can walk along the display and the system follows the user, thereby enabling gaze interaction while on the move (B).

## ABSTRACT

While gaze holds a lot of promise for hands-free interaction with public displays, remote eye trackers with their confined tracking box restrict users to a single stationary position in front of the display. We present EyeScout, an active eye tracking system that combines an eye tracker mounted on a rail system with a computational method to automatically detect and align the tracker with the user’s lateral movement. EyeScout addresses key limitations of current gaze-enabled large public displays by offering two novel gaze-interaction modes for a single user: In “Walk then Interact” the user can walk up to an arbitrary position in front of the display and interact, while in “Walk and Interact” the user can interact even while on the move. We report on a user study that shows that EyeScout is well perceived by users, extends a public display’s sweet spot into a sweet line, and reduces gaze interaction kick-off time to 3.5 seconds – a 62% improvement over state of the art solutions. We discuss sample applications that demonstrate how EyeScout can enable position and movement-independent gaze interaction with large public displays.

## ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI)

## Author Keywords

Gaze Estimation; Body Tracking; Gaze-enabled Displays

## INTRODUCTION

Over the last years we have witnessed a significant increase in the number and size of displays deployed in public. Large displays are now commonly found in public communal spaces, such as shopping malls or transit areas in airports and train stations [13]. At the same time, as sensing technologies are becoming cheaper and more robust, various modalities for interacting with these displays have been explored. A particularly promising interaction modality is gaze, given that it is fast, natural, and intuitive to use [41].

However, in contrast to common desktop settings, public displays afford ad-hoc use over short periods of time by “passersby”, i.e. users who move in front of the display [26, 27]. These characteristics pose three unique challenges that have so far forestalled wider adoption of gaze interaction on large public displays: 1) Gaze-enabled public displays cannot afford time-consuming eye tracker calibration for each user prior to interaction [21], 2) they have to allow passersby to interact from different positions [36] and 3) they have to support interactions while on the move [30]. Previous work mainly addressed the first challenge [22, 34, 40]. To date, addressing the latter two currently requires augmentation of each individual user with head-mounted eye trackers as well as interconnected displays [24].

To address the last two challenges we present EyeScout, a novel active eye tracking system that enables gaze interaction for a single user on large public displays from different lateral positions in front of the display and while on the move. EyeScout consists of a body detection and tracking module using a depth sensor, and an eye tracking module using an eye tracker mounted on a rail system. Our system detects

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

UIST 2017, October 22–25, 2017, Quebec City, QC, Canada

©2017 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-4981-9/17/10...\$15.00

<https://doi.org/10.1145/3126594.3126630>

## 9 EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.

the user's position in front of the display and then moves the eye tracker to face and follow the user. EyeScout thereby enables gaze interaction with large public displays that are (1) *position-independent*: the user can interact from different positions within 90 cm in front of the display and along the display's full extent, and (2) *movement-independent*: the user can interact via gaze while passing by the display. EyeScout builds on existing work that employed Pursuits [34], a popular calibration-free gaze interaction technique. It is important to note that EyeScout readily supports other gaze interaction techniques, such as gaze gestures [14] or pupil-canthy-ratio [40].

The specific contributions of this work are three-fold: First, we introduce the design and implementation of EyeScout, a novel active eye tracking system that enables gaze interaction for a single user on large displays. Second, we report on our findings from a controlled laboratory study (N=23) to evaluate the performance of EyeScout. We evaluate EyeScout for scenarios in which users "Walk *then* Interact" (to test for position independence), as well as "Walk *and* Interact" (to test for movement independence). Findings from our study show that EyeScout is well-suited for both interaction modes and well-perceived by users. In particular, EyeScout reduces the time required to kick-off gaze interaction (i.e., the time starting from the moment the user appears in front of the display until the time the user interacts) to 3.5 seconds – a 62% improvement over state-of-the-art methods [1, 41]. Finally, we discuss how active eye tracking using EyeScout can enable novel gaze-based applications that are not possible with current systems, such as gaze interaction with non-planar displays, and on escalators and moving walkways.

### RELATED WORK

Our work builds on three strands of prior work: (1) Gaze interaction with public displays and (2) active eye tracking.

#### Gaze Interaction With Public Displays

Gaze holds particular promise for interaction with public displays given that it, for example, overcomes the embarrassment problems associated with mid-air gestures [8], reflects attention [32], and allows at-a-distance interactions [17]. However, gaze-enabled public displays also face several unique challenges. First, although eye tracker calibration may be acceptable in desktop settings where users interact for long periods of time, interaction time on public displays is short [26, 27], which makes time-consuming tasks, in particular calibration, undesirable [21]. Recent works have therefore either tried to improve calibration [22] or employed calibration-free interaction techniques [19, 34, 40].

Second, in contrast to desktop settings, users approach public displays from different directions and want to interact with them from different locations, distances, and relative orientations [36]. However, existing gaze-enabled public displays restrict users' position; users have to position themselves within the tracking box of the eye tracker for their eyes to be detected [20, 22, 34, 41]. A common approach to address this problem is to guide passersby to the right position in front of the eye tracker, for example, using on-screen mirrored video feeds [41], markers on the floor [20], or on-screen visual cues

that are adapted based on the user's distance to the display [1]. In contrast, EyeScout moves the eye tracker to the user as soon as they approach the display or as they walk along it. An alternative approach is to use head-mounted eye tracking that allows for freedom of movement. However, this approach requires the eye tracker to (1) identify the position and borders of surrounding displays, (2) map gaze estimates to on-screen positions, and (3) communicate gaze data to the display. Prior work attached printed markers on the display [39] or used on-screen visual markers [15] to locate the display and map the gaze estimates onto it. These approaches usually rely on a tethered connection to the display. Lander et al. used visual feature tracking to determine the positions of surrounding displays, and exchanged gaze data over Wifi [24].

Although mobile trackers are starting to become ubiquitous and integrated into eyewear computers and despite the vision of pervasive display networks [13], pervasive integration on such a big scale would require taking concepts from lab settings to the field. In-field application is currently challenging, as participants need to be explicitly hired and asked to wear mobile eye trackers [12]. Until passersby wearing mobile eye trackers becomes the norm, there is a need to study user behavior on gaze-enabled public displays using other means, such as remote eye trackers.

#### Active Eye Tracking

One way to achieve position-independent gaze interaction with public displays is by using active eye tracking, i.e. systems that adapt to the user's eye position rather than restricting their head and/or body movements. Active eye tracking is particularly popular in medicine; it is used in eye surgery to account for eye and body movements during lasik operations [25]. A common approach is to use a single [9, 10, 11, 29] or multiple pan-and-tilt cameras [6], or pan-and-tilt mirrors [28] to adapt to the user's head position. Hennessey and Fiset used a Kinect to detect faces and adjust the angle of an eye tracker mounted on a pan and tilt mechanism accordingly [17]. While all of these methods demonstrated the potential of active eye tracking, EyeScout is first to *move* the eye tracker rather than only panning and tilting it in a single fixed position. This way, EyeScout actively accommodates for the user's body movements along large displays.

#### THE EYESCOUT SYSTEM

When interacting with large, cylindrical or spherical displays, users approach from different directions and do not necessarily interact from a static position in front of the display [1]. Instead, passersby expect to be able to walk-up to the display and interact from any position. We refer to this interaction mode as "Walk *then* interact". Similarly, passersby move at different speeds and often interact with displays while moving [30]. We refer to this interaction mode as "Walk *and* interact". The key challenges in both interaction modes are that the system needs to detect the user's eyes at arbitrary stationary positions or while the user moves in front of the display.

We designed EyeScout for single user gaze interaction specifically with these two interaction modes and associated challenges in mind. The design was inspired by camera motion

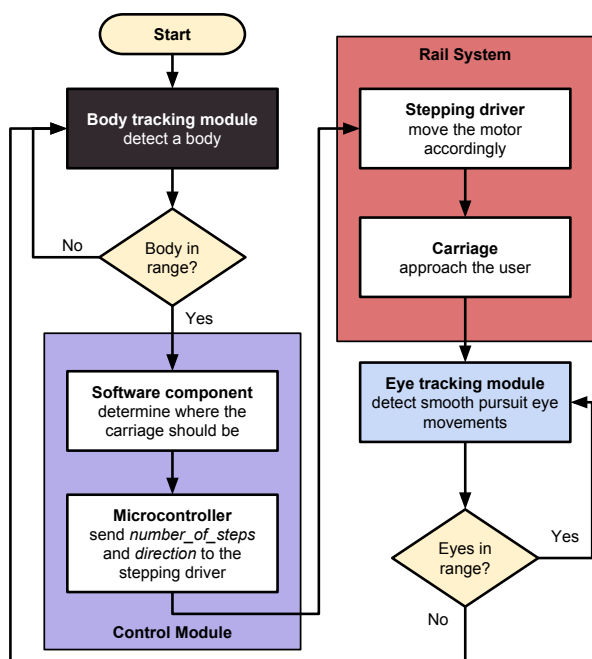


Figure 2. The body tracking module lies at the outset of EyeScout. Once a body is detected, the control module determines the target position of the carriage which carries the eye tracker, and instructs the rail system to move it to the body's position. The eye tracking module then tracks the user's eyes. If eyes are no longer detected, the system determines where the user is and moves the carriage to the new position.

technologies used in cinematography [18], and by actuated displays [2, 31]. EyeScout consists of four main modules (see Figure 2 and Figure 3): a body tracking module, a rail system, an eye tracking module, and a control module. Each module can be run and debugged independently allowing it to be modified or replaced. This maximizes ease of assembly and allows straightforward integration of new modules. In the following we describe each module in details.

### Body Tracking Module

The body tracking module consists of a motion sensing camera (Microsoft Kinect One). The Kinect is mounted on a tripod (150 cm high) aligned with the center of the rail. This module determines the user's position in front of the display. We used the skeletal joints provided by the Kinect API to detect the user's position. The Kinect detects the user immediately as soon as they are in range. We placed the Kinect at the opposite side of the display to cover a large area (see Figure 3).

### Rail System

The rail system consists of a 4-meter twin track aluminium rail<sup>1</sup> and a carriage<sup>2</sup> to move the eye tracker. Two 3D-printed end pieces were attached at both ends of the rail (see Figures 4A and 4C). The end pieces serve two main purposes: (1) To hold switches that are activated once the carriage reaches any of the ends. The switches are used for a one-time

<sup>1</sup>igus® drylin® Double rail [http://www.igus.eu/wpck/2003/drylin\\_w\\_doppelschiene](http://www.igus.eu/wpck/2003/drylin_w_doppelschiene)

<sup>2</sup>igus® drylin® Double rail/carriage [http://www.igus.eu/wpck/8980/drylin\\_w\\_Slider\\_Schienen](http://www.igus.eu/wpck/8980/drylin_w_Slider_Schienen)

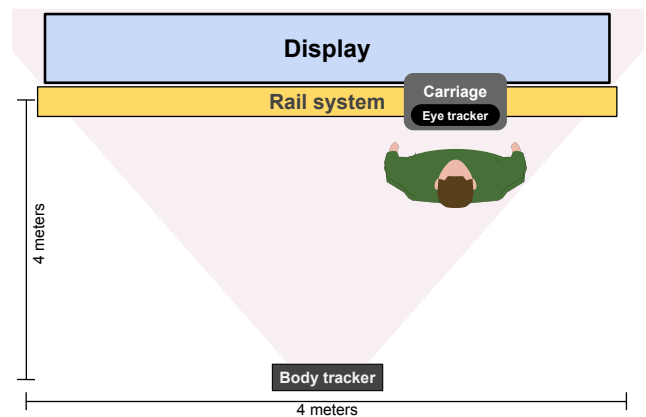


Figure 3. An illustration of the top view of the apparatus (not to scale). The body tracking module (Kinect One) is positioned 4 meters away from a 4 meters wide projected display. The carriage, which carries the eye tracker (Tobii REX), moves along the rail system according to the commands received from the body tracking module.

system calibration that determines the bounds of the rail's range to prevent the carriage from colliding with other components of the end pieces. (2) Each end piece harbors a steel axis that holds a pulley. A motor, whose axis is connected to the steel axis, is mounted on one of the end pieces. Thus, when the motor spins, the axis is spun and moves a tightened timing belt that moves the carriage. In addition to a DC motor<sup>3</sup>, we used a 2-phase digital stepper driver<sup>4</sup> to convert digital signals to commands that can control the motor. The entire rail system was mounted above three evenly distributed tripods (height: 113 cm), as depicted in Figure 5A.

### Eye Tracking Module

The eye tracking module consists of a remote eye tracker (Tobii Rex) and a custom 3D-printed mount. The mount is attached to a tripod head that allows adjusting the angle of the eye tracker. The tripod head is in turn attached to another 3D-printed base that is screwed into the carriage (Figure 4B). This module is responsible for tracking the eyes once they are in range. The minimum and maximum range of the eye tracker (40 cm to 90 cm in our case) are predefined in the body tracking module. This allows the body tracking module to detect when users are too close or too far away from the eye tracker. Similarly, the eye tracking module continuously detects whether or not the user's eyes are detected. This information can then be used to provide feedback to the user.

For gaze interaction with our system we use Pursuits [34], which has been widely adopted recently for calibration-free gaze interaction. Pursuits checks for motion correlation [33] between user's eye movements and trajectories of on-screen targets. The strength of the method lies in its ability to determine which object the user is gazing at without the need for calibrating the eye tracker to each user. As public displays require immediate usability and cannot afford the time spent for eye tracker calibration [21], Pursuits is well-suited for use

<sup>3</sup>igus® drylin® step motor NEMA23 [http://www.igus.eu/wpck/7663/N11\\_6\\_14\\_2\\_Schrittmotor\\_NEMA23](http://www.igus.eu/wpck/7663/N11_6_14_2_Schrittmotor_NEMA23)

<sup>4</sup>Leadshine DM556 <http://www.leadshine.com/UploadFile/Down/DM556m.pdf>

## 9 EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.

in the context of public displays. Moreover, as users will be moving, gaze estimates can be expected to have low accuracy as humans naturally bob up and down while walking. Its robustness to inaccurate gaze data makes Pursuits even more suitable for our deployment.

Our implementation of Pursuits is based on prior work; we used Pearson's product-moment correlation with a threshold of 0.85 and a window size of 500 ms [34]. This means that the system computes Pearson's correlation every 0.5 s. This is similar to previous work, some of which used a 0.5 s window size [20, 34], while Orbits [16] and TextPursuits [22] used 1 s and 2 s respectively. The stimulus whose movement correlates the most with the eye movement is deemed to be the one being looked at, if the correlation is higher than 85%.

### Control Module

The control module handles the logic of EyeScout and the communication between the body tracking module and the rail system. It consists of a software component written in C# and a microcontroller (Arduino Due). The software component runs on a Microsoft Surface Pro that is connected to the Kinect via USB. It receives the coordinates of the user's body from the body tracking module. According to the predefined distance between the Kinect and the rail system (4 meters in our implementation), the software computes an optimal position for the carriage at which the user's eyes would be in the eye tracker's range. Based on the coordinates that are received from the eye and body tracking modules, text prompts are shown to instruct the user to stand back or come closer to the system if necessary. These coordinates can be sent to the microcontroller via Bluetooth or USB.

Given the current position of the carriage, the Arduino Due maps the new coordinates received from the software component to a number of motor steps in a specific direction. These values are then forwarded to the stepping motor, which moves the motor accordingly. The microcontroller also interacts with the switches that are attached to the end pieces. In the aforementioned calibration process, the microcontroller determines the bounds of the rail's range and updates them internally if necessary. After successful calibration, the carriage is never instructed to move far enough to touch the switches again (see Figure 4A). For example, if a user moves out of range the carriage will stop right before it touches the switch. For additional security, the microcontroller issues an emergency stop command in case the carriage touches the switches after calibration, and resets the boundary values. Although the tasks of the microcontroller could also be performed by the software module, we opted for separating the component that interacts with the rail system and the one that interacts with the body tracking module to further minimize the dependencies between different modules. Additionally, while the heavy traffic generated by the body tracking module often requires a tethered connection to the computer (in our case a Kinect One is connected via USB), microcontrollers can be communicated with wirelessly (e.g., via Bluetooth), which would not necessitate long cables between the Kinect and the rail system; the computer could stay next to the Kinect, and the microcontroller could stay next to the rail system (see Figure 3).

### EVALUATION

We designed a controlled laboratory study to evaluate the performance of EyeScout for both interaction modes: (1) *Walk then interact*: the user approaches the display then interact while *stationary* and (2) *Walk and interact*: the user interacts with the display while *moving* at different speeds.

### Participants

We recruited 26 participants (11 females) aged 19 to 37 years ( $M = 26.77$ ,  $SD = 4.46$ ). All participants had normal or corrected-to-normal vision. Four had previous experience with body tracking devices such as Kinect, out of which three participants had prior experience with eye tracking. One participant was not detected by the Kinect due to wearing a black outfit, and thus was excluded from the analysis.

### Apparatus

We deployed EyeScout in one of our lab spaces (7.15 m  $\times$  5.65 m). The system was placed parallel to the wall at a distance of 117 cm (see Figure 5A). We used a short throw projector (1920  $\times$  1080 pixels) positioned 92 cm from the wall. The eye tracker angle was adjusted at 50° to the display.

### Study Design and Procedure

The study was split into two experiments, each evaluating one of the two interaction modes. All participants took part in both experiments. Half of the participants started with "Walk then Interact" while the other half started with "Walk and Interact". Each experiment also followed a within-subjects design in which all participants performed all conditions.

The experimenters started by introducing the study and asking the participant to sign a consent form. In both experiments, the system showed a white vertical rectangle (the "interaction frame") in which three dots moved in either linear or circular trajectory (see Figure 5). The participant's task was to select the red moving dot from among the grey ones via Pursuits, i.e., by simply following it with their eyes. We picked two trajectory types that are commonly used in implementations of Pursuits: circular [16, 20] and linear [22, 34] trajectories. All simultaneously shown moving dots were selectable and followed the same trajectory. We predefined 8 arrangements for moving dots (see sample arrangement in Figure 5B), 4 of which followed linear trajectories, while the other 4 followed circular trajectories. Figures 5C and 5D show one example of each. The participant was shown one arrangement (i.e., one set of three selectable dots moving in the same trajectory) at a time. In each selection, participants had to select 1 of 3 targets. Dots disappeared after being selected. After participants had performed all selections in both experiments, they filled in a questionnaire and participated in a semi-structured interview.

### Experiment 1: Walk then Interact

To study the interaction mode where passersby approach a random spot in front of the display and then interact while stationary, the interaction frame appeared at a random position on the display in this experiment. The participant was asked to walk to the frame and then select the red dot via Pursuits. The carriage approached the participant as he/she approached the frame, to ultimately position the eye tracker in front of

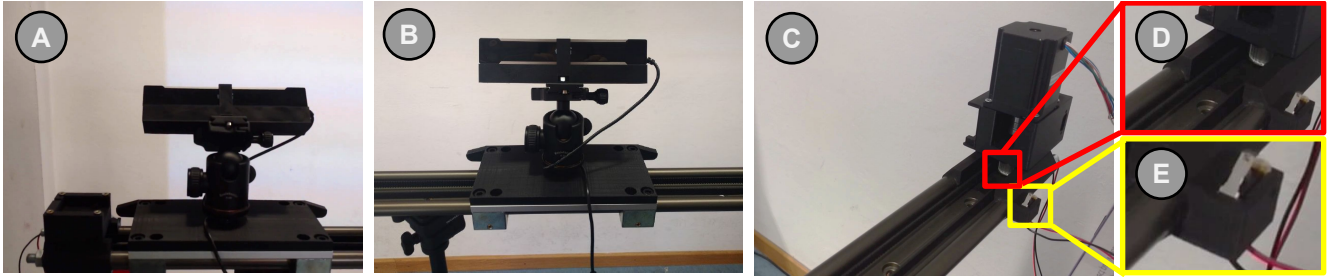


Figure 4. (A) and (C) show the 3D-printed end pieces attached at the ends of the rail. Each end piece harbors a pulley that moves the timing belt (D), and a switch (E) that prevents the carriage from accidentally colliding with the end piece. One end piece carries the motor (C). A 3D-printed base is screwed into the carriage (B), on which a tripod head is attached. The head allows the adjustment of the eye tracker, held by a 3D-printed holder.

the user. After a successful selection, the frame reappeared at another random position that was at least 50 cm away from the previous one. This was done to ensure that the participant had to approach a different spot in front of the display, rather than performing the selection from the current position. Each participant performed 3 blocks, each of which covered one selection per trajectory arrangement. Thus, every participant performed  $8 \text{ trajectory arrangements} \times 3 \text{ blocks} = 24 \text{ selections}$ . We consider these blocks an additional independent variable, referred to in the following as “repetitions”. Repetitions were studied to investigate learning or fatigue effects. The order of conditions was counter-balanced using a Latin-square.

#### Experiment 2: Walk and Interact

To study the interaction mode where passersby interact while moving, in this experiment the participant, the interaction frame, and the carriage were all moving. We focus on scenarios in which users interact with content that moves with them, as typically done in large interactive displays intended for moving users [30]. To evaluate if the carriage’s speed had an impact on detection accuracy we introduced an independent variable “carriage speed” with three conditions: 0.36 m/s (maximum speed of EyeScout), 0.3 m/s, and 0.24 m/s. The interaction frame would follow the participant, but the participant would be able to make a selection only when in range of the eye tracker. Each participant performed 24 selections in this experiment ( $8 \text{ arrangements} \times 3 \text{ speeds}$ ). The order of conditions was counter-balanced using a Latin-square.

#### Limitations

In the current version of our prototype, taller participants are asked by means of the aforementioned textual prompts to step back in order for the eye tracker to detect them. In our study, the angle of the eye tracker to the head was between  $35^\circ$  and  $50^\circ$ . The exact value depends on the user’s height and distance from the tracker. Future systems can adjust the eye tracker’s angle dynamically to be within this range.

Another limitation of gaze interaction while on the move is that it might be affected by motion blur. Although we did not face this problem in our study; the performance of EyeScout was almost similar in “Walk then Interact” (baseline) compared to “Walk and Interact”. However we acknowledge that higher carriage speeds might result in less accurate gaze data.

Like current stationary eye trackers, EyeScout only supports a single user. Multi-user support is one of the most important directions for future work. This can be realized by mounting

multiple eye trackers on different belts or by using appearance-based gaze estimation methods [32] that use multiple or a single wide angle RGB camera. In the current implementation of EyeScout, there are three possible scenarios in which a person other than the user appears in range of EyeScout: 1) passersby show up near the user, 2) passersby step between the user and the Kinect, 3) passersby step between the user and the eye tracker. The eye tracker locks onto the user whose eyes are detected even if a passerby occludes the user from the Kinect’s view. This prevents disrupting interaction in case of glitching position tracking, and means that 1) and 2) do not influence the system’s usability. The Kinect can be placed at the top of the interaction area to account for 2) when the user is moving. However 3), while unlikely, would result in the passerby taking over the interaction.

Finally, we can validate EyeScout’s performance with higher walking speeds only after upgrading its motor. However users are likely to slow down to interact when moving [30].

#### Quantitative Results

Although the Kinect performed fairly well in our setup, it failed to detect one participant wearing black. We further excluded the data of two more participants: One found the task to be overwhelming; he struggled to walk around and look at multiple objects at the same time. The second squinted his eyes too often, resulting in very few collected gaze points.

We measured the *cruise time* in the “Walk then Interact” experiment, i.e., the time it took the carriage from the moment the interaction window appeared till the moment eyes were detected. Because this is the first time a commercial IR-PCR eye tracker is used for active eye tracking, we wanted to investigate whether there is any degradation in the eye tracker performance when it is in motion. This was done by logging the *gaze points per second* during the Walk and Interact experiment. In both experiments, we additionally measured the *error count*, which we define as the number of grey dots that were selected before the red dot. There can be 0, 1, or 2 errors before selecting the red dot.

#### Error Count

As shown in Table 1, errors decreased in the “Walk then Interact” experiment as participants performed more repetitions. This suggests that there could be a learning effect, i.e., participants adapted to the system. In the “Walk and Interact” experiment we found a slight increase in the number of errors as the carriage moved at higher speeds.



## 9 EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.

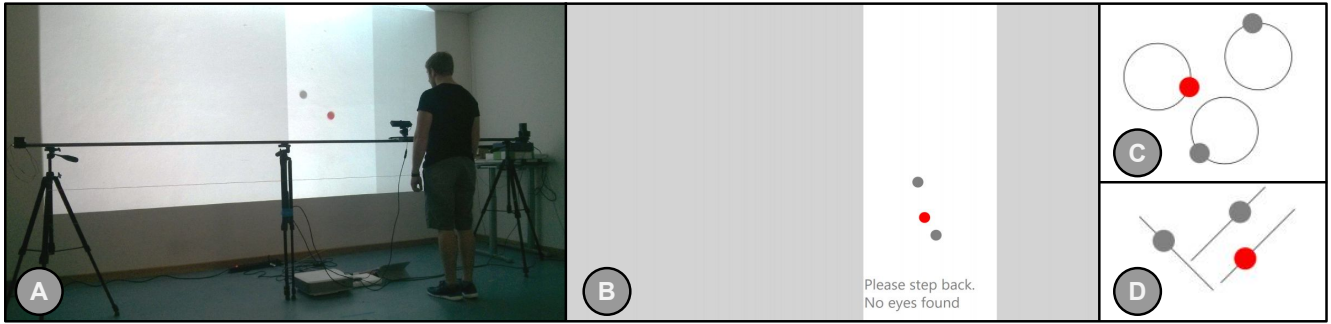


Figure 5. An “interaction window” appeared on the projected display with an arrangement showing three selectable moving dots. The dots moved either in circular trajectories (C) or in linear trajectories (D). The participant’s task was to select the red dot. In the “Walk *then* Interact” experiment, the window appeared at a random place on the display; the participant had to walk to the window then perform the selection. While in the “Walk *and* Interact” experiment, the window moved along the display; the participant had to walk along the window and perform selections while moving.

$\epsilon$	Walk <i>then</i> Interact <sub>(baseline)</sub>			Walk <i>and</i> Interact		
	Repetition 1	Repetition 2	Repetition 3	Speed 1	Speed 2	Speed 3
0	73.9%	83.7%	79.4%	79.3%	76.1%	72.3%
1	26.1%	15.8%	20.1%	19.6%	21.7%	23.9%
2	0%	0.5%	0.5%	1.1%	2.2%	3.8%

Table 1. The table shows the percentage of times a successful selection of a red dot was preceded by 0, 1, or 2 errors (i.e., selection of a gray dot). As participants performed more selections, errors in “Walk *then* Interact” decreased. This suggests that there could be a learning effect. In “Walk *and* Interact”, errors increased slightly with higher speeds.

$\mu$	Walk <i>then</i> Interact			Walk <i>and</i> Interact		
	Repetition 1	Repetition 2	Repetition 3	Speed 1	Speed 2	Speed 3
$\mu$	2.4 s	2.9 s	2.9 s	4.7 s	4.7 s	5.4 s
$\sigma$	2.1 s	2.2 s	2.2 s	6.1 s	5.8 s	7.0 s

Table 2. The table shows the mean selection time and standard deviation. Selection times in “Walk *and* Interact” are longer than in “Walk *then* Interact” due to tracking distortions caused by walking. While in “Walk *and* Interact” mean selection time is only slightly more than in previous work (e.g., 1.5 s to 2.0 s [20]).

### Mean Correlation Coefficient

The mean correlation coefficient that was calculated during the experiments is 0.91 (Threshold = 0.6). This is comparable to previous work. The mean correlation coefficient in previous work were 0.8 (Threshold = 0.6) in TextPursuits [22], 0.86 (Threshold = 0.7) in EyeVote [23], and 0.89 (Threshold = 0.8) in a study by Khamis et al. [20]. This shows that the accuracy of EyeScout is comparable to static systems.

### Cruise Time

Although the maximum speed of the carriage was 0.36 meters per second, the carriage needs to accelerate and decelerate at the beginning and end of its cruise. The average cruise speed in the “Walk *then* Interact” experiment was 0.2 meters per second ( $SD=0.09$ ), while the average distance between two consecutive selection areas was 0.7 meters ( $SD=0.34$ ). This means that the overall mean cruise time was 3.5 seconds.

### Gaze Points per Second

The eye tracker used in our prototype system returns a maximum of 30 gaze points every second. By measuring the number of collected gaze points at each carriage speed, we did not find large differences between the different conditions used in the “Walk *and* Interact” experiment. The eye tracker collected

an average of 21.59 ( $SD = 7.84$ ), 22.78 ( $SD = 7.3$ ), and 22.57 ( $SD = 7.33$ ) gaze points per second during the slow, medium and fast carriage speeds respectively. This means that we did not find any evidence that the performance of EyeScout is degraded by movements of up to 0.36 meters per second.

### Selection time

Table 2 summarizes the selection times. The overall mean selection time in “Walk *then* Interact” is 2.7 seconds, which is only slightly more than in previous work (e.g., 1.5 s to 2.0 s [20]). In “Walk *and* Interact” mean selection time is 4.9 s, which is longer due to tracking distortions caused by walking.

### Questionnaire

We asked participants about the perceived easiness and precision of selections on a 5-point Likert scale. In the “walk *then* interact” experiment, participants felt that walking up to the display *then* making selections was easy ( $Mdn = 4$ ,  $SD = 0.59$ ) and precise ( $Mdn = 4$ ,  $SD = 1.08$ ) (see Figures 6 and 7). They also agreed that the eye tracker was positioned properly in front of them ( $Mdn = 4$ ,  $SD = 0.66$ ). A Friedman test showed statistically significant differences in perceived easiness of selections depending on carriage speed  $\chi^2(2) = 9.414$ ,  $p = 0.009$  in the “walk *and* interact” experiment. Post hoc analysis with Wilcoxon signed-rank tests was conducted with Bonferroni correction, indicating significant differences ( $p < 0.017$ ). Median perceived easiness of selection levels for the slow, medium, and fast carriage speeds were 5 (4 to 5), 4 (4 to 5) and 4 (4 to 4), respectively. There were no significant differences between medium and slow carriage speeds ( $Z = -0.535$ ,  $p = 0.593$ ). However, there was a statistically significant reduction in perceived easiness in fast vs slow carriage speed ( $Z = -2.555$ ,  $p = 0.011$ ), and in fast carriage speed vs medium carriage speed ( $Z = -2.517$ ,  $p = 0.012$ ). This means that selections done during slower and medium carriage speeds are perceived to be easier to perform (Figure 6). We performed a Friedman test to determine if there were differences in perceived precision across the different carriage speeds. Perceived precision was consistent ( $Mdn = 4$ ) across the different carriage speeds and there were no statistically significant differences  $\chi^2(2) = 3.073$ ,  $p = 0.215$ . This means there is no evidence that perceived precision changes depending on the carriage speed (Figure 7).

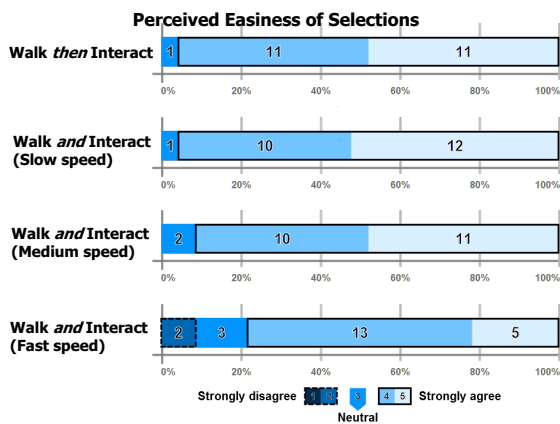


Figure 6. Overall, participants found it easy to make selections when using EyeScout. However, making selections on medium and slow speeds is perceived to be easier than on fast speeds.

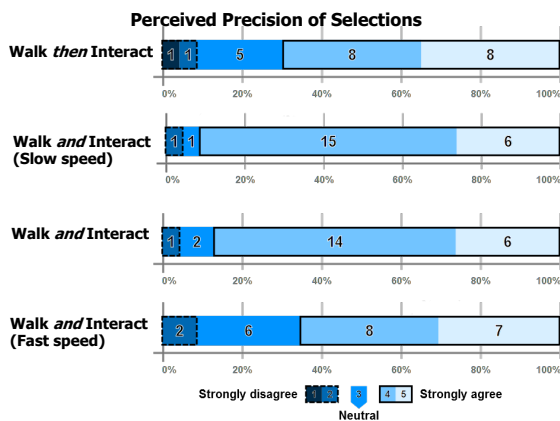


Figure 7. Participants perceived precision to be generally high. It seems that selections at higher speeds in the “Walk and Interact” experiment are perceived to be slightly less precise, however we found no significant differences to support this.

### Observations and Qualitative Feedback

The overall feedback about the experience was very positive. Participants mentioned that they found the system “interesting” and thought it was “working surprisingly well”. One participant found the experience of being followed by the eye tracker to be “futuristic”. Another participant commented that he found the idea “novel” and “sci-fi”. This is in-line with the novelty effect often experienced when interacting via gaze.

Participants also mentioned some aspects of EyeScout that could be improved.

#### Hardware Improvements

Although carefully placed to avoid tripping, three participants reported that the legs of the tripods sometimes distracted them. This suggests that future versions of EyeScout should be mounted differently. For example, the rail can be engraved into the wall underneath the display. Five participants reported that the cable that connects the eye tracker to the PC distracted them shortly when they saw it the first time. In field deployments of EyeScout, wireless technology (e.g., WiFi or Bluetooth) should be utilized instead. To further reduce

distraction, the tracker could be embedded into a case with semi-transparent glass so that the current position of the tracker is not visible to users. Yet, knowledge about the position of the eye tracker might positively influence the position of the user – hence, this needs to be subject to future investigation.

A female participant with long dark hair was not always correctly detected by the Kinect due to the aforementioned problem with detecting dark objects. While she reported that the eye tracker was consequently not always in front of her, she did not report any problems regarding the responsiveness of the system. A possible direction of improvement is to place the motion sensing device at the top of the interaction area, or embed a wide range sensor into the display.

#### Interaction

Two participants reported they were uncomfortable with performing Pursuits against circular trajectories. Some participants also reported feeling tired after performing 48 selections using Pursuits. This feedback is in line with findings reported from lab studies of Pursuits [22]. Given that interactions in real deployments would not involve as many Pursuit selections as in our study, this fatigue effect can be expected to play a minor effect in real deployments. However, applications that expect multiple selections (e.g., games) should be designed with the fatigue effect in mind.

One participant reported not having noticed the text prompts used to guide the user closer or farther from the eye tracker. Previous work has shown that public display users sometimes miss on-screen content, and are more likely to notice it if it is attached to their user representation [35]. Another participant was not confident that the system recognized his eyes, which led him to look at the eye tracker during his first trials. However that was only before he realized that the projected screen shows feedback when eyes are not detected. This suggests that the system should always provide feedback to indicate that the eyes are detected, rather than only when they are not in range. Furthermore, future work should consider different visual feedback methods, such as continuously showing eye symbols on the screen and adapting them depending on the state of eye detection. Similar to previous work [1], the content can be adapted to subconsciously guide the user by making it completely visible only when the user is at a particular distance from the display.

#### Walking Strategies

When asked how to move, the experimenters told the participants that they are free to move however they liked. We noticed that participants walked in different ways in the “Walk and Interact” experiment. While the majority walked naturally with their head turned towards the display, some walked sideways with their entire body facing the display. Participants who walked sideways reported that it was uncomfortable, but they walked that way thinking that the system would not detect them otherwise. One participant tried both and eventually settled on the natural walk. User interfaces of commercial eye trackers explicitly tell their users to relax and act naturally; EyeScout can similarly provide such feedback when unnatural moving behavior is detected.



## 9 EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.

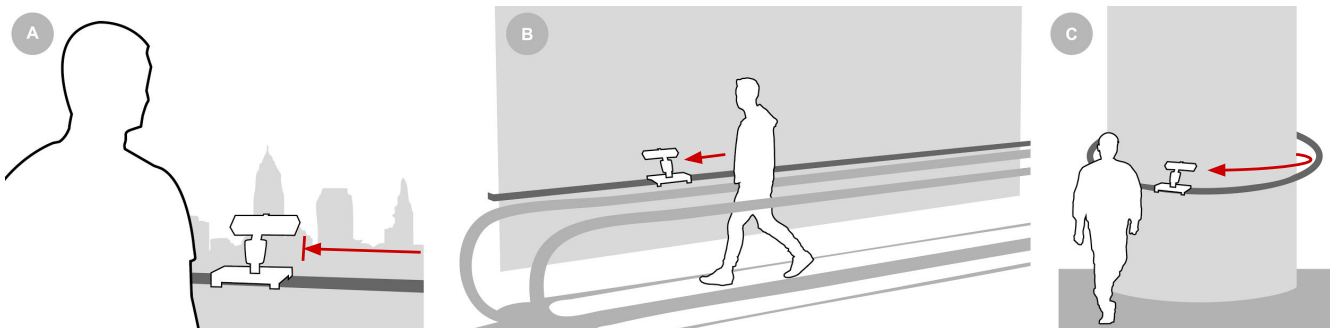


Figure 8. We envision multiple scenarios in which EyeScout can create new possibilities for gaze interaction with large displays. For example, EyeScout can enable interaction with objects or scenery (A); a user is standing at a touristic vantage point (e.g., observation deck), EyeScout can detect and track the user's gaze along the scene and provide information about the buildings and landmarks being gazed at. Users can interact while on a moving walkway (e.g., at an airport), where the tracker actively follows the moving user (B). EyeScout could also enable gaze-based interaction on non-planar displays such as cylindrical displays; the user approaches the display from any direction, then EyeScout moves the eye tracker to a position from which it can detect the user's eyes (C).

### SAMPLE APPLICATIONS

We envision EyeScout to be used in a variety of scenarios. We describe three examples in which the use of EyeScout enlarges the interaction space and opens up novel possibilities for interaction via gaze that are otherwise infeasible.

#### Gaze Interaction with Scenery

EyeScout could be used for gaze-based interaction with a scenery (see Figure 8A). For example, observation decks and towers offer vantage points for scenic overviews (e.g., of a city or a certain landmark). On these platforms, tourists and visitors overlook a scene and are often provided with audio guides that, in a sequential manner, describe what can be seen from the platform (for example, “To your left, next to the red building, you can see the townhall.”). However, this sequential feeding of information eliminates the exploratory nature of these platforms, thus hindering the tourist's experience. Even with the presence of a human tour guide, pointing at something (for example, a building or a landmark) and asking for information is not straight forward; communicating a point of interest to others by pointing is also inefficient.

We propose augmenting these platforms with EyeScout to enable eye tracking across the whole platform as shown in Figure 8A. In this scenario, the eye tracker would follow the user as he/she walks on the platform, the system would then be able to detect which buildings/landmarks the user is looking at, or allow the user to select landmarks. In-situ information about the area of interest can then be shown in the form of visual overlays or audio messages. The displayed information could be predefined in the system and loaded based on the positions of the areas of interest relative to the platform, and the position of the user. Since this application requires a precise gaze point, calibration might be required. A direction for future work is to investigate how well calibration-free techniques such as TextPursuits [22] perform in scenarios where users are moving. Another alternative is to use gaze gestures (e.g., right and left) to allow users to select the landmark they want to learn about.

#### Eye Tracking on Moving Walkways and Escalators

Moving walkways (aka travelators) and escalators can be found in large numbers in airports, supermarkets, ski resorts, museums, and public transport stations. People using them are

usually presented with static content at one or both sides of the walkways, such as advertisements. Gaze interaction or attention measurements in walkways is infeasible using current systems and techniques, unless each passersby is augmented with a head-mounted eye tracker.

Walkways and escalators could be augmented with EyeScout, such that the eye tracker would follow the user (see Figure 8B). In static non-interactive contexts, eye tracking enabled through EyeScout could provide information about the passersby's attention (for example, which advertisements passersby look at). The content displayed on one or two sides can also be interactive, in this case interacting with a UI while standing on a moving walkway via touch would be challenging unless the UI moves with the user. On the other hand, in cases where the user “passes by” the content (for example, moving walkway surrounded by stationary exhibits<sup>5</sup>), interaction via gaze extends the user's reach, as the user's gaze vector could reach farther areas compared to interaction via touch or via mid-air gestures.

#### Gaze-based Interaction with Non-Planar Displays

There has been a recent interest in deploying and studying passersby behavior in front of non-planar displays such as cylindrical [4, 5, 7] and spherical [3, 37, 38] displays. Due to their form factor, the requirement of adapting to users approaching and interacting from different directions and positions becomes even more prominent. To date, the strict positioning constraints imposed by eye trackers make gaze interaction infeasible with such non-planar displays.

Although we evaluated EyeScout only in the context of a large planar display, the same concept is applicable to non-planar ones by using a circular rail system (Figure 8C). A camera mounted on the top of the display could detect surrounding motion and move the eye tracker to intercept passersby as they approach the display. Gaze could then be tracked to understand which content the passersby attend to, or to enable interaction as they move around the display. We believe such displays to be particularly useful in guiding users to less crowded areas of a public space.

<sup>5</sup><https://www.nga.gov/exhibitions/villarealinfo.shtml>

## DISCUSSION

Findings from our study show that EyeScout successfully overcomes the positioning requirements imposed by classical eye tracking systems, and is flexible to lateral movements in front of the display at different speeds. We also found that it is well perceived by users, who reported finding it generally easy and precise to perform selections using Pursuits.

### Improvement over State-of-the-art

#### *62% – 87% Faster in Kickstarting Gaze Interaction*

State-of-the-art methods for guiding passersby to the sweet spot – which is, in our case, the area in which the user is detected by the eye tracker – were reported to require 4.5 to 23 seconds [1]. After reaching the sweet spot, users typically need to align their face to the correct position in front of the gaze-enabled display. Recent work reported that users required 4.8 seconds for the face alignment on a gaze-enabled display [41]. By adding these values, we can expect that even for state-of-the-art methods, passersby need 9.3 to 27.8 seconds before they can start interaction via gaze. EyeScout reduces this time to 3.5 seconds, which represents a 62% to 87% improvement. This improvement is due to EyeScout not requiring users to move to the sweet spot, nor to align their faces, but instead “doing the work for them”. EyeScout still requires less time compared to previous approaches although it informs participants if they are too far from or too close to the eye tracker, and asks them to reposition accordingly.

Previous work showed that unless displays are immediately usable, users abandon them [26, 27]. Hence, we expect an increase in conversion rates due to EyeScout’s reduction of kick-off time. In future work, this increase can be quantified through a field study.

#### *From “Sweet Spot” to “Sweet Line”*

EyeScout maximizes the horizontal flexibility of public displays. While previous work report an optimal interaction spot (the sweet spot [27]), our work extends the sweet spot to a *sweet line*: an area with the width of the display, and the length of the eye tracker’s range. Future work can further extend the distance to the screen by incorporating 3D vector rigs<sup>6</sup>.

### Gaze-based Interaction on the Move

Although the Tobii REX eye tracker that we used is intended for stationary settings, it performed fairly well when in motion. The number of collected gaze points stayed almost the same across the different cruise speeds and was sufficient to perform Pursuits-based selections. Although there is a slight increase in error when using faster cruise speeds compared to slower ones (see Table 1), the accuracy of selections achieved in the “Walk *then* Interact” experiment do not differ much from those in the “Walk *and* Interact” experiment. We furthermore found that participants generally felt selections to be easy and precise but results were in favor of slower speeds compared to faster ones in the “Walk *and* Interact” experiment. The differences between the perception of easiness and precision of both experiments were not significant. Figures 6 and 7 suggest that participants perceived selections positively in all modes.

<sup>6</sup><http://www.vector-cam.com/services.html>

### Gaze Interaction Techniques Other than Pursuits

We opted to use Pursuits because it is the state-of-the-art method for calibration-free gaze interaction with public displays and it addresses the first of the three challenges mentioned at the beginning of this paper. However, EyeScout is not limited to this technique. It is straight forward to replace Pursuits by other calibration-free techniques such as eye gestures [14] or pupil-cantheni-ratio [40]. Furthermore, future work can experiment with calibrating the eye tracker implicitly while users are interacting as in TextPursuits [22] to collect more accurate gaze points.

However, we cannot claim that all methods can be adapted into active eye tracking scenarios; Pursuits is robust to uncalibrated gaze points, which is likely one of the reasons it performs well while the eye tracker was in motion. It will be interesting to see whether techniques that require accurate gaze estimates, such as dwell time, can be used with EyeScout or whether the applications enabled by our system will remain infeasible if these techniques are used.

### Upgrading EyeScout

The gaze-interaction technique is not the only upgradable component of EyeScout. The way EyeScout is designed allows straightforward upgrades and improvements to the different hardware and software modules. Basically all modules are upgradable, including the motor, the eye tracker, the body tracker, and even the control unit. For example, a stronger motor can be used to increase the cruise speed, a wide-angle RGB camera can be used instead of infrared-based eye trackers, and body positions can be detected via on-body sensors.

## CONCLUSION

In this work we introduced the design and implementation of EyeScout, a novel active eye tracking system that addresses two challenges that were unsolved in research on gaze-enabled public displays to date. Findings from a user study show that EyeScout is not only well-perceived but also allows passersby to interact with large displays (1) from different positions and (2) while one the move. We furthermore introduced several sample applications that demonstrate how active eye tracking can enable new interactions with gaze that were not possible before. Our detailed description of EyeScout’s implementation is valuable for researchers and practitioners alike who would like to employ active eye tracking into their public displays.

## ACKNOWLEDGEMENTS

Thanks to Patrick Mörwald for his help in developing EyeScout. Work on this project was partially funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B).

## REFERENCES

1. Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. 2015. GravitySpot: Guiding Users in Front of Public Displays Using On-Screen Visual Cues. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST ’15)*. ACM, New York, NY, USA, 47–56. DOI : <http://dx.doi.org/10.1145/2807442.2807490>

## 9 EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.

2. Gilles Bailly, Sidharth Sahdev, Sylvain Malacria, and Thomas Pietrzak. 2016. LivingDesktop: Augmenting Desktop Workstation with Actuated Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5298–5310. DOI : <http://dx.doi.org/10.1145/2858036.2858208>
3. Hrvoje Benko, Andrew D. Wilson, and Ravin Balakrishnan. 2008. Sphere: Multi-touch Interactions on a Spherical Display. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. ACM, New York, NY, USA, 77–86. DOI : <http://dx.doi.org/10.1145/1449715.1449729>
4. Gilbert Beyer, Florian Alt, Jörg Müller, Albrecht Schmidt, Karsten Isakovic, Stefan Klose, Manuel Schiewe, and Ivo Haulsen. 2011. Audience Behavior Around Large Interactive Cylindrical Screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1021–1030. DOI : <http://dx.doi.org/10.1145/1978942.1979095>
5. Gilbert Beyer, Florian Köttner, Manuel Schiewe, Ivo Haulsen, and Andreas Butz. 2013. Squaring the Circle: How Framing Influences User Behavior Around a Seamless Cylindrical Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1729–1738. DOI : <http://dx.doi.org/10.1145/2470654.2466228>
6. David Beymer and Myron Flickner. 2003. Eye gaze tracking using an active stereo head. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, Vol. 2. II–451–8 vol.2. DOI : <http://dx.doi.org/10.1109/CVPR.2003.1211502>
7. John Bolton, Peng Wang, Kibum Kim, and Roel Vertegaal. 2012. BodiPod: Interacting with 3D Human Anatomy via a 360° Cylindrical Display. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 1039–1042. DOI : <http://dx.doi.org/10.1145/2212776.2212380>
8. Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *In Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT 2003)*. 17–24.
9. Chao-Ning Chan, Shunichiro Oe, and Chern-Sheng Lin. 2007a. Active Eye-tracking System by Using Quad PTZ Cameras. In *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE*. 2389–2394. DOI : <http://dx.doi.org/10.1109/IECON.2007.4460366>
10. Chao Ning Chan, Shunichiro Oe, and Chern-Sheng Lin. 2007b. Development of an active gaze tracking system in unrestricted posture. In *International Conference on Control, Automation and Systems, 2007. ICCAS '07*. 1348–1353. DOI : <http://dx.doi.org/10.1109/ICCAS.2007.4406547>
11. Dong-Chan Cho and Whoi-Yul Kim. 2013. Long-Range Gaze Tracking System for Large Movements. *IEEE Transactions on Biomedical Engineering* 60, 12 (Dec 2013), 3432–3440. DOI : <http://dx.doi.org/10.1109/TBME.2013.2266413>
12. Nicholas S. Dalton, Emily Collins, and Paul Marshall. 2015. Display Blindness?: Looking Again at the Visibility of Situated Displays Using Eye-tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3889–3898. DOI : <http://dx.doi.org/10.1145/2702123.2702150>
13. Nigel Davies, Sarah Clinch, and Florian Alt. 2014. *Pervasive Displays: Understanding the Future of Digital Signage* (1st ed.). Morgan & Claypool Publishers.
14. Heiko Drewes and Albrecht Schmidt. 2007. *Interacting with the Computer Using Gaze Gestures*. Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488. DOI : [http://dx.doi.org/10.1007/978-3-540-74800-7\\_43](http://dx.doi.org/10.1007/978-3-540-74800-7_43)
15. Marc Eaddy, Gabor Blasko, Jason Babcock, and Steven Feiner. 2004. My own private kiosk: privacy-preserving public displays. In *Eighth International Symposium on Wearable Computers*, Vol. 1. 132–135. DOI : <http://dx.doi.org/10.1109/ISWC.2004.32>
16. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 457–466. DOI : <http://dx.doi.org/10.1145/2807442.2807499>
17. Craig Hennessey and Jacob Fiset. 2012. Long Range Eye Tracking: Bringing Eye Tracking into the Living Room. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 249–252. DOI : <http://dx.doi.org/10.1145/2168556.2168608>
18. Axel Hoesl, Julie Wagner, and Andreas Butz. 2015. Delegation Impossible?: Towards Novel Interfaces for Camera Motion. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1729–1734. DOI : <http://dx.doi.org/10.1145/2702613.2732904>
19. Shahram Jalaliniya and Diako Mardanbegi. 2016. EyeGrip: Detecting Targets in a Series of Uni-directional Moving Objects Using Optokinetic Nystagmus Eye Movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5801–5811. DOI : <http://dx.doi.org/10.1145/2858036.2858584>



20. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2015. A Field Study on Spontaneous Gaze-based Interaction with a Public Display Using Pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. ACM, New York, NY, USA, 863–872. DOI : <http://dx.doi.org/10.1145/2800835.2804335>
21. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2016a. Challenges and Design Space of Gaze-enabled Public Displays. In *Adjunct Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 10. DOI : <http://dx.doi.org/10.1145/2968219.2968342>
22. Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016b. TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 12. DOI : <http://dx.doi.org/10.1145/2971648.2971679>
23. Mohamed Khamis, Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, and Florian Alt. 2016c. EyeVote in the Wild: Do Users Bother Correcting System Errors on Public Displays?. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 57–62. DOI : <http://dx.doi.org/10.1145/3012709.3012743>
24. Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. 2015. GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 395–404. DOI : <http://dx.doi.org/10.1145/2807442.2807479>
25. Michael Mrochen, Mostafa Salah Eldine, Maik Kaemmerer, Theo Seiler, and Werner Htz. 2001. Improvement in photorefractive corneal laser surgery results using an active eye-tracking system. *Journal of Cataract & Refractive Surgery* 27, 7 (2001), 1000 – 1006. DOI : [http://dx.doi.org/10.1016/S0886-3350\(00\)00884-1](http://dx.doi.org/10.1016/S0886-3350(00)00884-1)
26. Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 1285–1294. DOI : <http://dx.doi.org/10.1145/1873951.1874203>
27. Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 297–306. DOI : <http://dx.doi.org/10.1145/2207676.2207718>
28. Borna Nouredin, Peter D. Lawrence, and C.F. Man. 2005. A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding* 98, 1 (2005), 52 – 82. DOI : <http://dx.doi.org/10.1016/j.cviu.2004.07.005> Special Issue on Eye Detection and Tracking.
29. Takehiko Ohno and Naoki Mukawa. 2004. A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-based Interaction. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA '04)*. ACM, New York, NY, USA, 115–122. DOI : <http://dx.doi.org/10.1145/968363.968387>
30. Constantin Schmidt, Jörg Müller, and Gilles Bailly. 2013. Screenfinity: Extending the Perception Area of Content on Very Large Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1719–1728. DOI : <http://dx.doi.org/10.1145/2470654.2466227>
31. Julian Seifert, Sebastian Boring, Christian Winkler, Florian Schaub, Fabian Schwab, Steffen Herrdum, Fabian Maier, Daniel Mayer, and Enrico Rukzio. 2014. Hover Pad: Interacting with Autonomous and Self-actuated Displays in Space. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 139–147. DOI : <http://dx.doi.org/10.1145/2642918.2647385>
32. Yusuke Sugano, Xucong Zhang, and Andreas Bulling. 2016. AggreGaze: Collective Estimation of Audience Attention on Public Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '16)*. ACM, New York, NY, USA. DOI : <http://dx.doi.org/10.1145/2984511.2984536>
33. Eduardo Velloso, Marcus Carter, Joshua Newn, Augusto Esteves, Christopher Clarke, and Hans Gellersen. 2017. Motion Correlation: Selecting Objects by Matching Their Movement. *ACM Trans. Comput.-Hum. Interact.* 24, 3, Article 22 (April 2017), 35 pages. DOI : <http://dx.doi.org/10.1145/3064937>
34. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 439–448. DOI : <http://dx.doi.org/10.1145/2493432.2493477>
35. Robert Walter, Andreas Bulling, David Lindlbauer, Martin Schuessler, and Jörg Müller. 2015. Analyzing Visual Attention During Whole Body Interaction with Public Displays. In *Proceedings of the 2015 ACM*

## 9 EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays.

- International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 1263–1267. DOI : <http://dx.doi.org/10.1145/2750858.2804255>
36. Miaosen Wang, Sebastian Boring, and Saul Greenberg. 2012. Proxemic Peddler: A Public Advertising Display That Captures and Preserves the Attention of a Passerby. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*. ACM, New York, NY, USA, Article 3, 6 pages. DOI : <http://dx.doi.org/10.1145/2307798.2307801>
37. Julie R. Williamson, Daniel Sundén, and Jay Bradley. 2015a. GlobalFestival: Evaluating Real World Interaction on a Spherical Display. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 1251–1261. DOI : <http://dx.doi.org/10.1145/2750858.2807518>
38. Julie R. Williamson, John Williamson, Daniel Sundén, and Jay Bradley. 2015b. Multi-Player Gaming on Spherical Displays. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 355–358. DOI : <http://dx.doi.org/10.1145/2702613.2725447>
39. Lawrence H Yu and Moshe Eizenman. 2004. A new methodology for determining point-of-gaze in head-mounted eye tracking systems. *IEEE Transactions on Biomedical Engineering* 51, 10 (2004), 1765–1773.
40. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2014. Pupil-canthy-ratio: A Calibration-free Method for Tracking Horizontal Gaze Direction. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 129–132. DOI : <http://dx.doi.org/10.1145/2598153.2598186>
41. Yanxia Zhang, MingKi Chong, Jörg Müller, Andreas Bulling, and Hans Gellersen. 2015. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing* 19, 5-6 (2015), 967–981.



# 10

**TextPursuits**

**Using Text for Pursuits-Based Interaction and  
Calibration on Public Displays.**

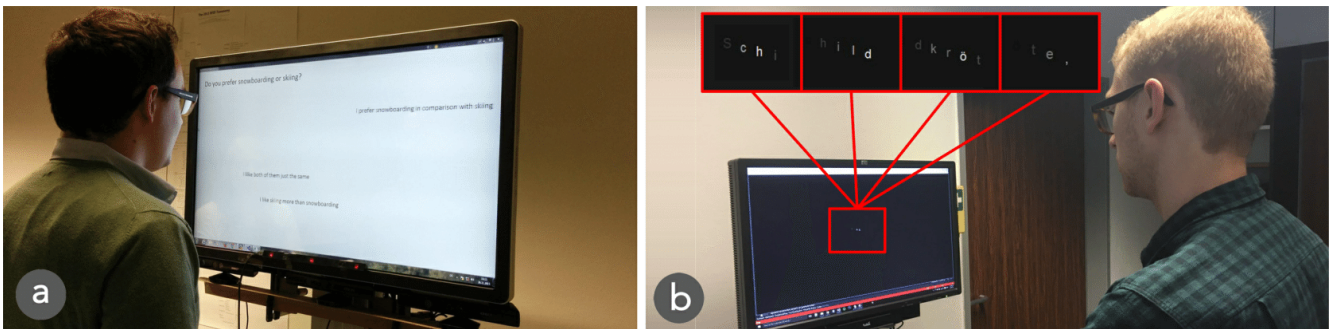
# TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays

Mohamed Khamis<sup>1</sup>, Ozan Saltuk<sup>1</sup>, Alina Hang<sup>2</sup>, Katharina Stolz<sup>1</sup>, Andreas Bulling<sup>3</sup>, Florian Alt<sup>1</sup>

<sup>1</sup> LMU Munich, Germany

<sup>2</sup> Sonix Group, Munich, Germany

<sup>3</sup> Max Planck Institute for Informatics, Saarbrücken, Germany  
{mohamed.khamis, florian.alt}@ifi.lmu.de, bulling@mpi-inf.mpg.de



**Figure 1.** We explore the use of text-based stimuli to enable gaze interaction with public displays using the Pursuits technique [47]. Motivated by the fact that much of the content on large displays is text, we investigate two use cases: (a) Users can spontaneously *interact* with text-based content without calibration. A sample application could be a survey where answers in the form of text are selected by reading them (left). (b) An eye tracker can be *calibrated* implicitly as users read text on the screen (right). After calibration, fine-grained information on the user's gaze point are obtained.

## ABSTRACT

In this paper we show how reading text on large display can be used to enable gaze interaction in public space. Our research is motivated by the fact that much of the content on public displays includes text. Hence, researchers and practitioners could greatly benefit from users being able to spontaneously *interact* as well as to implicitly *calibrate* an eye tracker while simply reading this text. In particular, we adapt Pursuits, a technique that correlates users' eye movements with moving on-screen targets. While prior work used abstract objects or dots as targets, we explore the use of Pursuits with text (read-and-pursue). Thereby we address the challenge that eye movements performed while reading interfere with the pursuit movements. Results from two user studies (N=37) show that Pursuits with text is feasible and can achieve similar accuracy as non text-based pursuit approaches. While calibration is less accurate, it integrates smoothly with reading and allows areas of the display the user is looking at to be identified.

## ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces — Input Devices and Strategies

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

UbiComp'16, September 12–16, 2016, Heidelberg, Germany

© 2016 ACM. ISBN 978-1-4503-4461-6/16/09 \$15.00

DOI: <http://dx.doi.org/10.1145/2971648.2971679>

## Author Keywords

Public Displays; Smooth Pursuit; Gaze Interaction; Text

## INTRODUCTION

As they are becoming ubiquitous and cheap to deploy, displays can be found in public spaces such as airports [4], shopping centers [9] and train stations [8]. At the same time, sensing technologies are becoming increasingly available for easy and low cost integration with public displays, supporting different ways of interaction. Common interaction modalities for displays include touch [10], smart phone interaction [3, 12], mid-air gestures [34], and recently also gaze [20, 48, 55].

Gaze holds particular promise for public displays [22]. It is intuitive [46], natural to use [47], indicates visual attention, and usually precedes action [31]. However, a drawback is that eye trackers require calibration, which is time-consuming and cumbersome [31]. While devoting time for calibration is acceptable for desktop settings, public displays require *immediate usability* [33] as interaction times are usually short [34]. Hence, calibration has been identified as one of core challenges of gaze-enabled public displays [21]. Prior work investigated alternative techniques [48, 55].

A popular approach is Pursuits [47, 49], which relies on correlating movements of dynamic objects on the display with the smooth pursuit eye movement performed when following a moving object. Pursuits was successfully deployed for multiple public display installations, where it was used for both gaze interaction [20, 48] and eye tracker calibration [6, 36].



Meanwhile, one of the most prominent content types on public displays is text. For example, displays are utilized for opinion gathering and sharing in public areas [17, 24]. In many applications passersby read and select from a set of text-based options [15, 16, 35, 42, 44, 50]. And also (pervasive) advertising on public displays often heavily relies on text [1].

Nevertheless, little is known about whether and how Pursuits can be used with text. To date, Pursuits has been studied with moving dot-like stimuli, for which the user gazes at a single, spatially clearly defined target. On the other hand, the use of Pursuits with textual stimuli is not straight forward: reading is not spatially confined and overlays the smooth pursuit movement, which could result in difficulty in correlating eye movements and the trajectory of text-based stimuli. Also, due to the Midas effect, gaze-based systems need to distinguish users reading textual content from interacting with it.

We investigate the use of text as stimulus for Pursuits. We see two main use cases for public displays: (1) It can be used for calibration-free gaze interaction [20, 48]; by displaying moving objects, users can “pursue” the object they want to select. The system then determines which object the user is looking at by correlating eye movements and movements of the objects. (2) Pursuits can be used for easier and less tedious calibration of eye trackers [6, 36]; by following a moving stimulus, mappings between movements of the user’s eyes and the stimulus can be collected and used for calibration.

To provide a proof-of-concept we implemented two applications: EyeVote is a survey application for public displays that enables a user to select an answer from a set of text-based options. Read2Calibrate shows animated text on the screen and feeds the gaze data to an algorithm that gradually enhances the calibration of the gaze tracking. We used both systems in two lab studies with the goal to assess accuracy based on different configurations (text inclination, text speed, text length, and trajectory) and to obtain early insights on the users’ view. The results show that text-based stimuli can be used for Pursuits-based gaze interaction but that designers need to be careful about the text trajectories and text length in order to minimize detection errors. Text-based calibration is in general less precise than state-of-the-art calibration procedures. However, the accuracy is sufficient to identify objects a user is looking at on the screen. We found text inclination to have a strong influence on the calibration quality. Both for interaction and calibration designers may need to make a tradeoff between the configuration leading to the most accurate results and the users’ preferred configuration.

The contribution of this work is threefold: (1) We present a prototype application, EyeVote, that allows text to be selected using Pursuits and report on a user study with 19 participants, assessing the accuracy and error rate based on different configurations (text length, trajectory). (2) We introduce Read2Calibrate, a system for calibrating eye trackers for displays by utilizing smooth pursuit eye movements and a text-based stimulus. Again, we evaluated the system with regard to accuracy based on different configurations (text inclination, text speed). (3) We derive a set of guidelines and recommendations for using text with Pursuits for both interaction and calibration.

### RELATED WORK

We build on two main strands of previous work: gaze interaction with public displays and interaction with text via gaze.

#### Gaze-based Interaction with Public Displays

Due to the benefits of gaze for public displays, two research directions have emerged to counter the problems associated with calibration on public displays: (1) enabling calibration-free gaze interaction for displays, and (2) making gaze calibration on public displays less tedious.

##### *Calibration-free Gaze Interaction with Public Displays*

Acknowledging the unsuitability of classic calibration for public displays, multiple systems were built to provide calibration-free gaze-based interaction. SideWays [53] and GazeHorizon [55, 56] use the pupil-canthen-ratio [54] to estimate horizontal gaze direction without calibration.

Pursuits [49, 47] can also be used for calibration-free gaze interaction. The technique requires displaying a dynamic interface [48], where “pursuitable” objects move. Eye movements are then correlated to the movements of the objects. The object whose movement correlates the most with that of the eyes is then assumed to be the one the user is looking at. Since its introduction, Pursuits has been used in a variety of applications including text entry [29], PIN code entry [11, 28] and entertainment applications [47, 49]. Pursuits has also been used for interaction with smart watches [13, 14] and interaction in smart environments [45]. The technique was shown to be intuitive and positively perceived by users [20].

##### *Eye Tracker Calibration for Public Displays*

Previous work aimed to reduce the effort needed for calibration. For example GazeProjector [27] allows gaze-based interaction across multiple displays using one time calibration and a mobile eye tracker. While mobile eye trackers have several advantages for interaction, public display users cannot be expected to wear them, unless trackers integrated with eye wear become commonplace. Hence remote eye trackers are currently more suited for that domain. Xiong et al. [52] used a remote RGB-D camera that requires one-time calibration.

Work by Pfeuffer et al. [36] uses the eye’s smooth pursuit movement to facilitate calibration. The approach relies on showing a moving object, which acts as a stimulus for the eyes to perform the smooth pursuit movement. Mappings between eye movements and positions of the stimulus are then collected and used to calibrate the eye tracker.

Pfeuffer et al. used a floating “Please wait” label to calibrate eye trackers. Rather than reading a label and keeping fixating it, our approach for calibration relies on gradually revealing text, which intrigues the user to fixate at the gradually revealed letters to understand the statement. Moreover, our work on interaction with text via Pursuits investigates a different aspect, namely we study how users can select from a set of text-based options using Pursuits.

#### Interacting with Text via Gaze

In gaze-based systems, the “Midas touch” effect [18] occurs when the system mistakes a user perceiving content for selecting content. This effect is amplified in the case of text

as reading requires time to perceive and understand the text. This challenge has been traditionally addressed by using dwell times – the system would require fixating the action element for a longer period of time (e.g. 1000 ms [30]).

Another approach to overcome the Midas touch is to use another modality in addition to gaze. Users of EyePoint [25] gaze at text, press and hold a keyboard button to magnify the area, refine the selection, and then release the button to select text. Stellmach et al. [40, 41] employed a similar approach by combining gaze and touch input. Although this approach was not used for text selection in particular, it is deemed suitable for the task. Kishi and Hayashi [23] combined gaze with on-screen buttons to enable users to select text. Chatterjee et al. [7] developed a text editor where users can move a text cursor by using gaze and pinch gestures.

A third approach is to use gaze gestures. In work by Toyama et al. [43], text selection was done either by repeatedly gazing at the beginning and the end of the text to be translated, or by gazing gradually from the beginning till the end of the text.

Sharmin et al. [38] introduced an automatic scrolling technique that is based on the user's gaze while reading text. Text 2.0 [5] exploits gaze by, for example, revealing content based on the words the user is currently reading.

Pursuits has the potential to cope with the Midas touch effect. Reading overlays the smooth pursuit eye movement, making false selections while reading less likely. Moreover, the Pursuits algorithm requires setting a window size, which is a time frame after which the correlation is calculated. This gives users the chance to perceive and read the text.

## INTERACTING WITH TEXT USING PURSUITS

The use of text for interaction via Pursuits has not been investigated in detail before. With our work we close this gap and support designers and developers when it comes to creating text-based content suitable for interaction using Pursuits. In particular, the following section introduces a prototype application that allowed us important aspects of using text for pursuit interaction to be investigated.

### Concept and Implementation

We implemented a voting system called EyeVote, that uses Pursuits as its only input mechanism for selecting one of several floating textual answers (see Figures 1A and 2). Once the system detects a selection, a confirmation message is shown on the screen, telling the user which answer was recognized. The message is kept for some seconds, followed by the next question and its options.

In the following we describe our implementation of Pursuits, and the experimental variables that we used in the study.

#### Text Selection via Pursuits

Pursuits works by correlating eye movements with those of the selectable options. Prior work utilized the Pearson's product-moment coefficient to calculate the correlation. Based on pilot experiments and previous work [14, 20, 47, 49], we used the same correlation function with a threshold of 0.9 and a window size of 2000 ms. This means that every 2 seconds, the

system computes Pearson's correlation. The floating answer whose movement correlates the most with the eye movement, is deemed to be the object the user is looking at, as long as the correlation is more than 90%.

To account for reading time and overcome the midas effect, the used window size value is higher than those used in other implementations (e.g. previous work used 500 ms [20, 47] and 1000 ms [14]).

#### Trajectories and Text Representations

We investigate how already established trajectory movements perform with respect to text selection. In particular, the following trajectories were used in our experiments.

1. Circular trajectory [13, 14, 20, 47] (Figure 2 top left).
2. Linear trajectory [20, 47, 49] (Figure 2 top right).
3. Rectangular trajectory [36] (Figure 2 bottom left).
4. Zigzag trajectory [47] (Figure 2 bottom right).
5. Mixed trajectory (each object follows one of the above trajectories).

We supported different text representations for the answers:

1. Short answers (<25 characters).
2. Two-lined answers.
3. Long answers (25+ characters).

### Evaluating Text Selection Using Pursuits

The main goal of this experiment was to understand the influence of different text characteristics on the accuracy of selection via Pursuits. In particular, we compared the effect of the aforementioned trajectory types and text lengths on detection errors. In addition, we assessed the effect of the different trajectory types on the perceived workload. To minimize any external influences, we conducted the study in the lab [2].

#### Design

The study was designed as a repeated measures experiment. Each participant performed five blocks with each block covering one of the five trajectory types. In every block, participants performed 4 selections  $\times$  3 text representations = 12 text selections using Pursuits. The order was counter-balanced across participants using a Latin-square.

The theme of the study was a voting application, where participants had to answer questions by selecting one of three possible floating text-based answers via Pursuits (see examples in Figure 2). In total, every participant answered 60 questions: 5 trajectory types  $\times$  3 text representations  $\times$  4 selections.

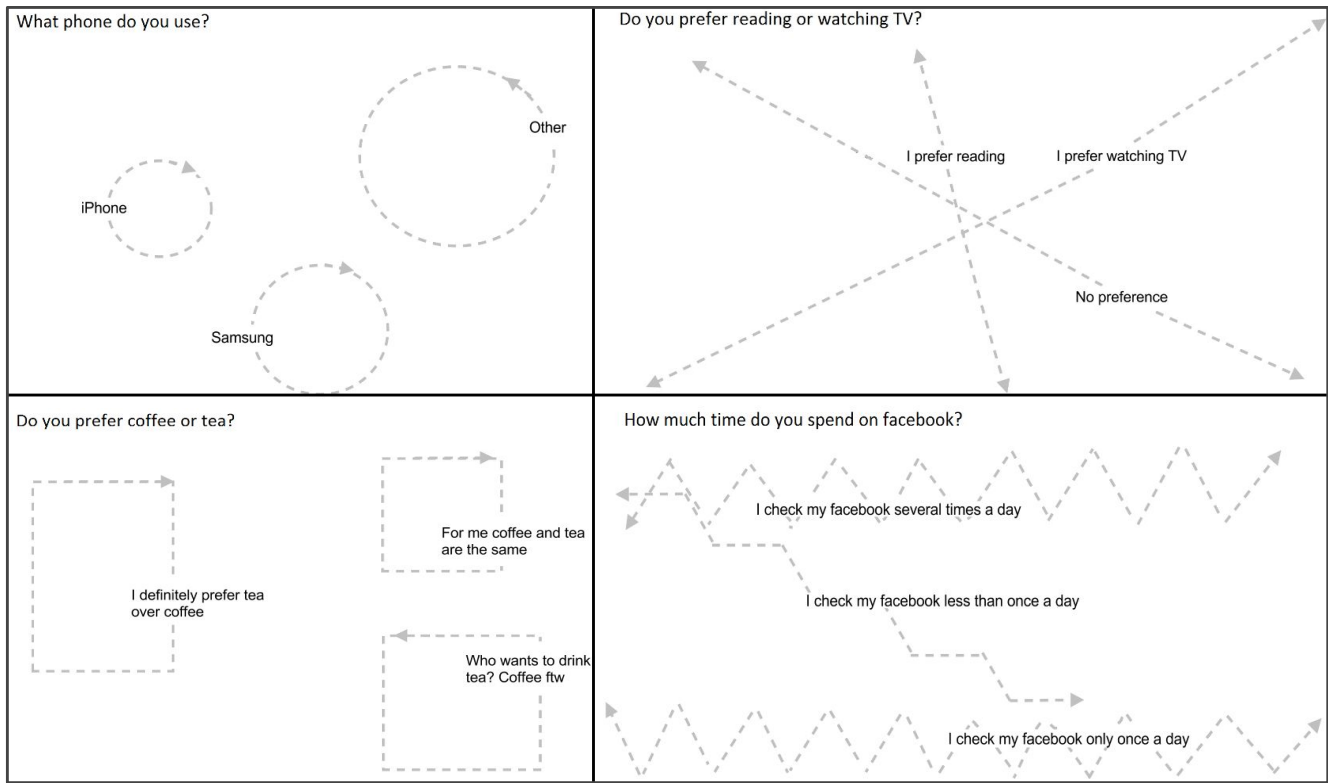
#### Apparatus

We deployed the EyeVote system on a 42-inch display (3810  $\times$  2160 pixels) in our lab (see Figure 1A). The display was equipped with a Tobii EyeX Controller (30Hz). Participants stood at a distance of roughly 60 cm from the eye tracker.

#### Participants

We recruited 19 participants (10 females) between 20 and 60 years through mailing lists and social networks. Four of them had previous experience with eye trackers but with Pursuits. All participants had normal or corrected-to-normal vision.

## 10 TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.



**Figure 2.** We manipulated two experimental variables: (1) Trajectory type: circular, linear, rectangular and zigzag trajectories shown above, in addition to a fifth condition “Mixed trajectory”, where each answer followed a different trajectory (arrows are only for illustration and were not shown to participants). (2) Text representation: short answers (top), two-lined answers (bottom left) and long answers (bottom right).

### Procedure

The experimenter began by explaining the study and asking the participant to sign a consent form. The experimenter then started the first block of 12 questions. After each successful Pursuit selection and before showing the following question, the system showed the user which answer was recognized. At that point the participant was asked to confirm whether or not the system detected the intended answer. In case of false detection, the participant was further asked to specify whether (a) the system detected a selection prematurely (i.e. the participant was still reading it) or (b) the participant was trying to select a different answer. To assess the perceived workload associated with text-based selection of every trajectory type, participants filled in a Nasa TLX questionnaire after each block.

### Results

We logged the time taken to answer each question as well as the false detections by the system. In total, we recorded 1140 selections (19 participants  $\times$  60 selections).

We classify errors as (a) *early detection errors*, that are errors due to a premature recognition of an option while the participant is still reading, and (b) *false detection errors*, that are errors due to the system recognizing a selection other than the one the participant intended. Out of 1140 selections, there were 124 errors ( $\approx 10.9\%$ ): 88 of them were early detections ( $\approx 7.7\%$ ), while 36 were false detections ( $\approx 3.2\%$ ).

Significantly different pair		p-value
Circular (10.2%)	Linear (36.4%)	$(p < 0.005)$
Circular (10.2%)	Zigzag (35.2%)	$(p < 0.05)$
Rectangular (8.0%)	Linear (36.4%)	$(p < 0.001)$
Rectangular (8.0%)	Zigzag (35.2%)	$(p < 0.005)$
Mixed (10.2%)	Linear (36.4%)	$(p < 0.001)$
Mixed (10.2%)	Zigzag (35.2%)	$(p < 0.05)$

**Table 1.** Trajectory type has a significant main effect on early detection errors, which are cases where the system recognized a selection while the participant is still reading. The table above summarizes the results of post-hoc analyses using Bonferroni correction, which revealed significant differences between multiple pairs. The numbers between brackets denote the percentage of early detection errors caused by the corresponding trajectory type out of all 88 early detection errors. The results indicate that circular, mixed and rectangular trajectories result in fewer early detection errors compared to linear and zigzag trajectories.

We classify errors as (a) *early detection errors*, that are errors due to a premature recognition of an option while the participant is still reading it, and (b) *false detection errors*, that are errors due to the system recognizing a selection other than the one the participant intended. Out of the 1140 selections, there were 124 errors ( $\approx 10.9\%$ ): 88 of them were early detections ( $\approx 7.7\%$ ), while the remaining 36 were false detections ( $\approx 3.2\%$ ).

#### Early Detection Errors

A repeated measures ANOVA showed significant effects for trajectory type on early detection errors  $F_{4,72} = 15.353$ ,  $p < 0.001$ . Table 1 summarizes the results of post-hoc anal-

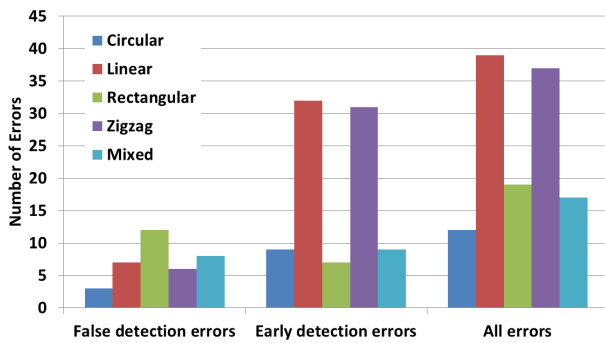


Figure 3. Circular, rectangular and mixed trajectories result in significantly fewer early detection errors compared to linear and zigzag ones. False detection errors are also fewer in the case of circular trajectories.

yses using Bonferroni correction, which revealed significant differences between multiple pairs. Circular, rectangular and mixed trajectories result in significantly fewer early detection errors compared to linear and zigzag trajectories.

No significant effects were found for text representation on early detection errors. However the results suggest that fewer early detections occurred in the case of short text (Figure 5).

#### False Detection Errors

No significant effects for trajectory type on false detection errors were found ( $p > 0.05$ ). Although a repeated measures ANOVA showed significant effects for text representation on false detection errors  $F_{4,36} = 3.916$ ,  $p < 0.05$ , no significant differences were found between pairs. This is likely due to the low number of false detection errors (36 out of 1140 selections). Figures 3 and 5 indicate a tendency for fewer false detections in the case of Circular trajectories and short text.

#### Perceived Workload

Figure 4 summarises the mean values for each subscale. A repeated measures ANOVA showed significant effects for trajectory type on *physical demand*  $F_{4,84} = 4.631$ ,  $p < 0.005$  and *effort*  $F_{4,84} = 4.334$ ,  $p < 0.005$ . Post-hoc analyses using Bonferroni correction showed that there are significant differences between *physical demand* ( $p < 0.05$ ) induced by

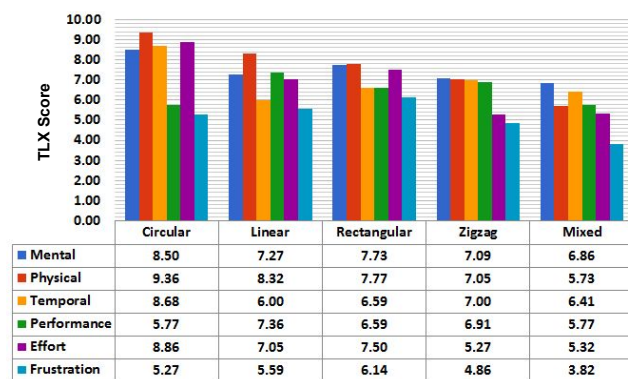


Figure 4. Although circular trajectories were perceived to have equal performance to mixed trajectories, participants perceived mixed trajectories to be less demanding in all other aspects.

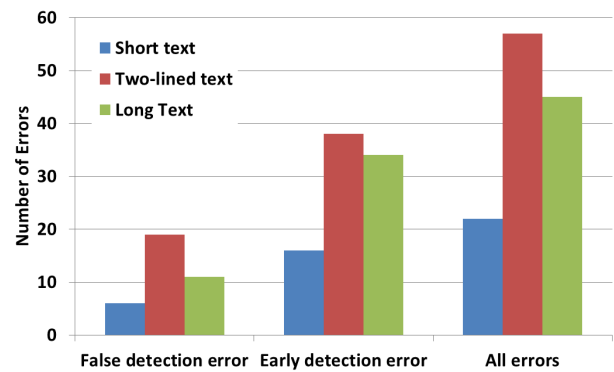


Figure 5. Fewer errors occurred when using short text compared to two-lined and long text.

circular ( $M = 9.4$ ,  $SD = 5.1$ ) and mixed trajectories ( $M = 5.7$ ,  $SD = 3.8$ ). Significant differences were also found between *effort* ( $p < 0.05$ ) to select circular ( $M = 8.9$ ,  $SD = 5.5$ ) and mixed trajectories ( $M = 5.3$ ,  $SD = 3.7$ ).

In summary, although circular trajectories were perceived to perform similar to mixed trajectories, participants perceived mixed trajectories as less demanding in all other aspects.

#### Summary

The previous study showed that it is feasible to use Pursuits for text selection, with the selectable text itself being the moving stimulus. Our results confirm that by using certain text representations and trajectory types, text can be used as a stimulus for one of the major uses of Pursuits, that is interaction. Next we investigate the second major usage of Pursuits, which is eye tracker calibration, with text used as stimulus.

#### PURSUIT CALIBRATION USING TEXTUAL STIMULI

One of the major strengths of using text for Pursuit calibration is that it allows for seamlessly calibrating an eye tracker on a public display as passersby simply read text. Previous work by Pfeuffer et al. [36] on Pursuit calibration also included a text label “Please wait”, floating from one corner of the display to another for 13 seconds. However, this required users to fixate the floating word for a long period even after reading, which might not be intuitive without prior instructions. To address this, we reveal text gradually to intrigue users to *pursue* the revealing text till its end (see Figure 1B).

In this section we present the implementation and evaluation of our prototype system called Read2Calibrate.

#### Concept and Implementation

In our implementation of Read2Calibrate, we developed a method that reveals the text gradually at an inclined angle. As new parts of the text gradually appear, the preceding parts disappear at the same rate (see Figure 1B). We opted for this representation in order to (1) ensure that users follow the stimulus (i.e. the text) till the end of the calibration session, (2) control the user’s reading speed, and (3) calibrate for as much area of the display as possible. In the following, we explain the rationale behind these three motives.

## 10 TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.



Figure 6. In the calibration session, participants read text gradually appearing across the display (as in Figure 1B). (A) and (B) show a sample text displayed at inclination angles of 15° and 315°. The angle and dotted lines are only shown for illustration and were not displayed to participants.

### Revealing Speed

To correlate eye movements to movements of a stimulus, the speed of the moving stimulus needs to be known beforehand. Reading speeds are different across users [26], which makes it difficult to predict which part of the text a user is looking at, in particular in a public display setting with no prior knowledge about the user<sup>1</sup>. However, revealing the text gradually ensures a controlled reading speed, that is, if the user is reading any part of the text, we can expect that the user is looking at the letters being revealed at that moment. To ensure the user is performing a smooth pursuit eye movement rather than a series of saccadic jumps, new parts of the text are gradually revealed. As newer parts of the text appear, preceding parts disappear gradually, to reduce the chances of backward saccadic jumps (see Figure 1B).

The speed of revealing text is an important variable. If the revealing speed is too fast, users might not have enough time to comprehend the text. Additionally, because eye trackers are typically limited to a certain sampling frequency, the faster text is revealed, the less mappings between eye movements and points on the display are collected. On the other hand, very slow revealing speed could also result in difficulty in understanding the text; as the time difference between revealing the first and last letters of a word is larger, the more difficult it becomes to understand the word. There is also a risk of the users losing interest and looking at other parts on the screen, which would negatively influence the calibration.

Based on prior work [36] and pilot tests, we introduced pauses of 500 ms, 350 ms and 200 ms in-between revealing each letter, which in visual angles equates to speeds of 1°/s, 2.1°/s, and 4°/s. Higher and lower revealing speeds were found to be very difficult to read. We refer to these speeds as the slow, medium and fast speed, respectively.

### Inclination Angle

Read2Calibrate needs to collect mappings between gaze points and points on the display. To cover as large an area of the display as possible, previous work used a stimulus that moved

across the display in diagonal, circular, or rectangular trajectories [36, 37]. Circular and rectangular trajectories are unnatural for gradually revealing text, while limiting stimuli to a horizontal line would calibrate with respect to the x-axis only. Hence, we chose to reveal the text in diagonal shapes.

However, since there has been no previous work about reading inclined text that is gradually appearing, we experimented with multiple inclination angles. Latin script is read from left to right, hence the text could be shown in two ways: starting from the upper-left part and ending in the lower-right part of the screen, or starting from the lower-left part and ending in the upper-right part of the screen. This translates to inclination angles between 270°–360° and between 0°–90 degrees°.

Taking into consideration the need to move the stimulus with respect to both axes, we experimented with six angles: 15°, 45°, 75°, 285°, 315°, and 345°. Figures 6A and 6B show sample text displayed at inclination angles of 15° and 315°.

### Calibration Correction

A prerequisite for calibration is to gaze at the stimulus. To exclude cases where users are not looking at the stimulus, our system calibrates after a certain correlation has been reached. We used the Pearson’s product-moment coefficient with a correlation threshold of 0.6, that is, the user is assumed to follow the stimulus if the correlation between its movement and the users’ eye-movements is  $\geq 60\%$ . We selected this value based on pilot testing and experience from prior work [36, 47].

In a calibration session, the letters are placed on the screen according to their angle. As the letters start to appear, pairs of gaze points and the revealed letter’s coordinates are collected.

To calculate the correction offset, for every gaze point ( $G$ ) recorded by the eye tracker, we measured the Euclidean distance between  $G$  and the center of the currently revealed letter ( $L$ ). After the calibration phase ends, the sum of these distances is divided by the total number of gaze points ( $N$ ) detected in that time frame. The resulting average distance value is then used as the correction offset (see Equation 1).

$$Offset = \frac{\sum_{k=1}^N L_k - G_k}{N} \quad (1)$$

<sup>1</sup>Note, that a future implementation could try to automatically assess the reading speed. This, however, would prolong the process.



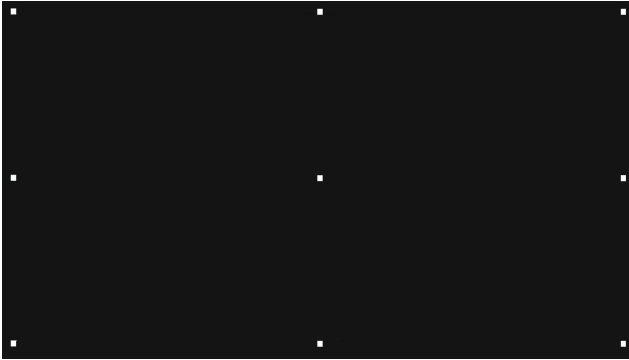


Figure 7. After each calibration session, participants proceeded to a testing session where they fixated at each shown point. During the study, one stimulus was shown at a time and blinked for 3 seconds, before transitioning to the following one.

### Evaluation of Pursuit Calibration Using Text

The goal of this study was to evaluate the effectiveness of Read2Calibrate, as well as to understand the influence of the different variables (inclination angle of the text and revealing speed). We studied the influence of these variables on the calibration quality, in addition to how users perceive them.

#### Apparatus

A 24 inch display (1920×1080 pixels) was equipped with a Tobii REX eye tracker (30Hz) and deployed in our lab (see Figure 1B). We invited 18 participants (10 females) aged 18 – 42 years ( $M = 26.2$ ,  $SD = 5.3$ ) through mailing lists. All participants had normal or corrected to normal vision.

#### Design

In a repeated measures experiment, every participant performed one calibration session per condition (6 angles × 3 speeds = 18 conditions). Each calibration session was followed by a testing session, where the participant was asked to gaze at a stationary point that appeared at nine positions on the screen (see Figure 7). The point blinked at each of the nine positions for 3 seconds. The order of the conditions was counter balanced across participants using a Latin-square.

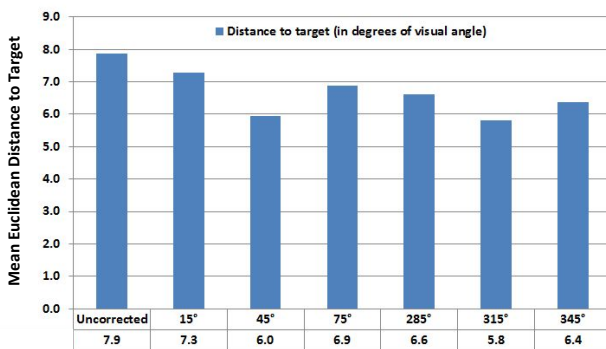


Figure 8. The graph shows the mean Euclidean distances between the gaze point and the target point after calibration by revealing text at different inclination angles, and in case of not applying any corrections. Lowest distance is achieved when text is revealed at an angle of 315°, followed by 45°.

#### Procedure

The experimenter started by explaining the study and the participant filled in a consent-form. The participant was then asked to stand at a distance of 60 cm from the eye tracker. The font-size was set to be easily readable at that distance. In each calibration session, the participant read a piece of a story that was gradually revealing across the display according to the condition's angle and speed. The participant then proceeded to a testing session, where we logged the gaze points, as detected by the eye tracker, the coordinates of the revealed text, as well as the corrections by our algorithm. After gazing at all nine stimuli, the calibration was reset and the participant proceeded to the next calibration session.

To reduce possible eye fatigue, participants were optionally allowed to take a break after every 3 calibration sessions. No visual feedback was shown during the whole experiment to avoid influencing the participant's gaze behavior.

We used a fable as a source of the revealing text. To ensure the validity of our analysis, it was crucial to make sure participants paid attention to the text. Hence, we asked participants three questions about the fable at the end. In addition to the compensation for participation, participants were encouraged to pay attention to the story by promising them an additional monetary incentive (50 Euro cents) for each correct answer they provide to the questions. All participants were aware of the rewarding mechanism before taking part in the study.

We concluded the study with a questionnaire and a semi-structured interview.

### Results

To evaluate the effectiveness of Read2Calibrate, we based our calculations on

1. the *target point* (the center of the stimulus which was gazed at during the testing session),
2. the *uncalibrated gaze point* (the gaze point as detected by the uncalibrated eye tracker), and
3. the *calibrated gaze point* (the gaze point after correction by Read2Calibrate).

For each stimuli shown in the testing sessions, we measured the *mean Euclidean distance* (1) between the uncalibrated gaze points and the target point and (2) between the calibrated points and the target point. Moreover, we measured the *positive correction rate*, which we define as the number of times the calibrated gaze point was closer to the target compared to the uncalibrated one.

#### Quantitative Results

Figure 8 shows that the mean Euclidean distance is shorter when text is revealed at angles of 315° and 45°. Thus, these angles result in better correction compared to others.

A repeated measures ANOVA showed significant main effects for angle  $F_{3,2,51.3} = 5.2$ ,  $p < 0.005$  and speed  $F_{2,34} = 4.8$ ,  $p < 0.05$  on positive correction rate. Post-hoc analysis using Bonferroni correction showed a significant difference ( $p < 0.05$ ) in positive correction rate for an inclination angle of 315° ( $M = 65\%$ ,  $SD = 0.06\%$ ) compared to 15° ( $M = 39.7\%$ ,

## 10 TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.

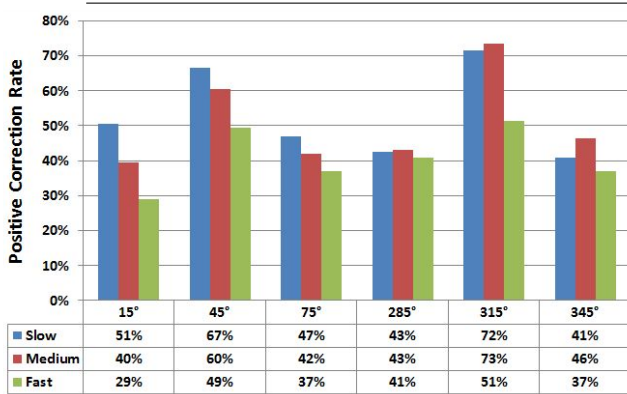


Figure 9. Revealing the text in a 315° inclination resulted in the highest number of positive corrections, e.g. 73% of the corrections by the Read-2Calibrate brought the gaze point closer to the target.

$SD = 0.05\%$ ), and also for 315° ( $M = 65\%$ ,  $SD = 0.06\%$ ) compared to 345° ( $M = 41.4\%$ ,  $SD = 0.06\%$ ). There were also significant differences in positive correction rate for fast revealing speed ( $M = 40.7\%$ ,  $SD = 0.05\%$ ) compared to slow revealing speed ( $M = 53.2\%$ ,  $SD = 0.045\%$ ). Figure 9 shows that angles of 315° and 45° resulted in more positive corrections compared to other angles. The figure also shows that fast revealing speeds result in less positive corrections.

### Qualitative Feedback

When asked how easy it is to read the text at the different angles (5-point Likert scale; 1=Strongly disagree; 5=Strongly agree), participants indicated that most angles were easy to read (see Table 2). As for the revealing speeds, participants found the medium speed ( $Med = 5$ ,  $SD = 0.6$ ) to be easier to follow compared to slow ( $Med = 4$ ,  $SD = 1.3$ ) and fast speeds ( $Med = 4$ ,  $SD = 1.2$ ).

When asked about their preference, participants pointed that they preferred angles that are closer to a horizontal line (i.e. 15° and 345°) as they felt more natural. However as indicated in the questionnaire, other angles are also easy to follow. On the other hand, multiple participants indicated that it felt unnatural to read the slow revealing text, P6 noted that “*I felt I was following the letters without really understanding the words*”. According to the participants, fast text is easy to follow, but difficult to comprehend.

## DISCUSSION & DESIGN RECOMMENDATIONS

Overall, the results of both studies suggest that text can be used as a stimulus to support interaction and calibration using Pursuits.

The text selection study showed that when using circular and mixed trajectories, shorter text can be selected with high accuracy using Pursuits while selecting longer pieces of text is more difficult. The text-based Pursuit calibration study showed that text can be an effective stimulus for seamlessly integrating eye tracker calibration as users read text. More specifically, gradually revealing text inclined at 45° or 315° at a speed of 2.1 visual degree angles per second highly improves the accuracy of the gaze point. The results also indicated that although

Text Inclination Angle	15°	45°	75°	285°	315°	345°
Median Score	5	4	4	4	4	4.5
Standard Deviation	0.43	0.67	0.92	0.96	0.83	0.70

Table 2. The table summarizes responses of participants when asked if it is easy to read text inclined at the corresponding angles (5-point Likert scale; 1=Strongly disagree; 5=Strongly agree). This indicates that the closer the text’s inclination to a horizontal line (i.e. 15° and 345°), the easier it is to read. However, the average scores of other angles as well as feedback from participants indicate that although other angles are less preferred, they are still readable.

participants preferred flat text, revealing text at inclined angles is easily readable and can be used for calibration.

### Text Selection via Pursuits

Overall, there was a low number of false detection errors – 36 false detection errors out of 1140 selections. Figure 3 shows that circular trajectories tend to be associated with less false detection errors. The results of the text selection study show that circular, rectangular, and mixed trajectories result in fewer early detection errors compared to linear and zigzag ones (see Figure 5 and Table 1). This means that reading text moving in linear and zigzag trajectories results in high correlation between eye movements and text movement, making the system confuse reading with selection.

Reading involves repeated saccades across a line of text as well as back to the beginning of the next line. Performing these saccades while pursuing text moving in a circular or rectangular trajectory distinguishes the eye movements from the text’s trajectory. This reading overlay makes the gaze trajectory less likely to correlate with that of the moving text, giving the user a chance to read and comprehend the text. On the other hand, reading text moving in linear and zigzag motion can be hardly distinguished from a Pursuit selection, resulting in a high correlation while reading, which in turn results in many early detection errors.

### Selecting Long Pieces of Text

Our motivation behind the use of different text representations and trajectories was to study how the Pursuits method can cope with a read-and-pursue eye movement. Our main finding is that Pursuits is indeed suitable for text selection, but only with shorter pieces of text. In cases where it is essential to show passersby longer pieces of text to select from, we recommend using a different stimulus.

**R1:** Use Pursuits for selection of short pieces of text; for longer pieces of text (25+ letters) use non-textual stimuli.

In case longer textual descriptions are needed, a display could be divided into two regions: a non-interactive region and an interactive region. In the case of the voting application, the non-interactive region could display the detailed answer options, each with a short but meaningful headline. The interactive region could then display the moving headlines only from which users can then make their selection. Alternatively, answers in the non-interactive region could be associated with colors or shapes. However, this may result in a higher cognitive load, since users need to associate colors or forms with the correct answer option.

### *Choosing the Right Trajectory Type*

Our analyses of the text selection experiment results show that circular trajectories are safer to use, as they result in significantly fewer errors. However circular trajectories are perceived to be highly demanding (see Figure 4). Mixed trajectories result in slightly more errors than circular ones. However, mixed trajectories were perceived to be significantly less demanding compared to other trajectories. This indicates a trade-off between user preference and accuracy of the system.

**R2:** *For peak performance when performing text selection using Pursuits, move text in circular trajectories. To increase user experience, a mixture of trajectories can be used together with an undo option.*

We conclude that if high accuracy when selecting text using Pursuits is required, designers should opt for circular trajectories. An example could be a situation where users are encountered in a passing-by situation or in a situation where they have only too little time to undo their selections, such as asking a question on customer satisfaction near a cashier. On the other hand, in cases where user experience should be maximized, mixed trajectories may be used that are slightly more error prone. In these cases, a floating “undo” label could be shown after the system detected a selection. An example could be users filling in a longer survey in return for an incentive, such as a gift voucher. Here it may be acceptable to occasionally ‘correct’ an answer while at the same time having a less demanding experience with the system.

### **Text-based Pursuit Calibration**

In general, participants’ feedback indicates that Read2Calibrate is positively perceived. The results of its evaluation show that inclination angles that result in diagonal-like orientation of the revealing text, such as 45° and 315°, significantly improve the accuracy of the gaze point. This is due to the fact that these angles result in the text covering larger areas of the screen. Inclination angles that bring the text closer to a horizontal line are preferred by users (15° and 345°), as they are more similar to flat text which users are acquainted to read. However, at these angles the text covers relatively less area with respect to the y-axis, resulting in poor calibration.

Slow revealing speeds result in users focusing on letters and losing track of the words they read. By analyzing the data, it was found that when using slower speeds participants were more likely to lose interest and look at other parts of the screen, presumably out of boredom. Fast speeds result in less data collected for the correlation check, which in turn results in lower calibration accuracy. Moreover, participants reported that revealing the text too fast makes it harder for them to understand what they read. Medium revealing speed turned out to be a good compromise: it is preferred by users and also results in a good calibration quality.

It should be noted, however, that the accuracy achieved by Read2Calibrate is lower than that of previous approaches that use Pursuits for calibration as well as of explicit calibration methods commonly known from eye tracking in desktop settings. At the same time, the major advantage of text-based Pursuits calibration is the seamless integration with users sim-

ply reading content on the public display. As a result, the calibration can be performed and used even in cases where the reader is not being consciously aware of it. Gaze information can then be used to enhance the user interface. For example, one may show a description of different sights next to a map of a city, like museums, churches, or historic buildings. As the system is calibrated, dwell time towards different sights could be used to determine what the reader is most interested in and additional information on how to reach a particular sight together with a discount coupon could be presented.

**R3:** *For moderate eye tracker calibration (accuracy of 5.8° of visual angle), text-based Pursuit calibration is recommended as it results in better user experience. If accuracy is crucial, classical Pursuit calibration should be used.*

The trade-off between accuracy and user experience can also be found when determining the angles at which the revealed text is inclined in Read2Calibrate. While bringing the text closer to a horizontal line makes reading feel more natural, revealing the text in a diagonal-like path results in the highest accuracy. Very steep text (e.g. 75° and 285°) result in both low accuracy and worse user experience and should hence be avoided.

**R4:** *Use diagonally-shaped paths, at an inclination of 315° or 45°, when revealing text to achieve highest accuracy with Read2Calibrate. For better user experience at the expense of calibration accuracy, reveal text in flatter shaped paths.*

A clear recommendation with regard to revealing speed can be provided. Here, accuracy is highest for medium revealing speed (2.1°/s) and this is in line with the users’ preference.

**R5:** *For text-based Pursuit calibration, an average revealing speeds of about 2.1° of visual angles per second should be used.*

The more comfortable a participant is with the revealing speed, the more accurate gaze-to-display mappings are collected and hence the more accurate the calibration is. Faster speeds result in fewer mappings, while slower ones distract the user.

### **Use Cases and Integration With Interactive Applications**

As a sample application that can be explicitly controlled using gaze, we implemented EyeVote, a voting system. Civic discourse is a popular use case for public displays [15, 16, 35, 39], where passersby select from a set of text-based options to express their opinions. Given the advantages of gaze for public displays [22], gaze-based voting systems can utilize Pursuits for selection of textual answers. Similarly, Pursuits can be used to answer on-screen quizzes. Selection of text via Pursuits can be useful in various other contexts, for example, users can select from a set of text-based options displayed at a museum to learn more about particular topics. In train-stations and airports, Pursuits can be employed to set fare preferences or pick products where possible options are displayed as text.

The second major use case is the implicit use of gaze data, either for analysis or for adaptive interfaces. Therefore, text-based Pursuit calibration can be integrated into public display applications in several ways. For example, a common practice



## 10 TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.

to tackle interaction blindness on public displays is to use call-to-action labels [32]. Such labels could serve as stimuli to calibrate an eye-tracker via Read2Calibrate. Further stimuli could be welcome messages or brief instructions on how to use a system or how to play a game. While the aforementioned examples utilize Read2Calibrate at the beginning of the interaction process, revealing text can also be shown amidst interaction. For example, a short text could hint at hidden or yet undiscovered features. Such a *calibration while interacting* may be useful for displays featuring multiple applications. Here, a game that is instantly usable may serve as an entry point. As the user finishes playing the game in the course of which the eye-tracker was calibrated using in-game textual content, the display could present further content that could benefit from knowledge about the user's gaze behavior. Note, that after the calibration, fine-grained gaze points can be collected and, hence, also other types of eye movements, such as fixations and saccades can be detected. As a result, an application may determine interest towards a particular content – this may be of particular interest for advertisers – as well as identify difficulties of users in perceiving content, for example, as they read text over and over again.

### Limitations and Future Work

Firstly, our evaluations so far were conducted in the lab. While this controlled setting was necessary to maximize internal validity and obtain comparable results, future work could employ text-based stimuli for Pursuits in an in-the-wild setting. Apart from verifying the results with regard to accuracy and errors, this may yield further insights on audience behavior and acceptance.

Secondly, participants of the text selection study answered 60 questions using Pursuits, whereas participants of the Read2-Calibrate study performed 18 calibration sessions. In a real-world situation, it is unlikely that users would perform such a high number of selections and users would not be required to verify the accuracy of the calibration. As a result, we expect the study to have caused a higher level of eye fatigue as an in-the-wild exposure to any of the systems would have done. Hence, participants may have been overly critical during their assessment of the system. Future work could capture in-situ feedback to verify the impact of our approach on the experience users have during use of our system.

Recent work explored feedback methods for Pursuit selections. Kangas et al. [19] compared different feedback modalities for Pursuits and found that haptic feedback is preferred by users compared to visual and auditory feedback. Špakov et al. [51] compared two smooth pursuit widgets to find that circular widgets exhibit higher performance. An additional direction for future work is to enable feedback methods to improve the user experience when using EyeVote and Read2-Calibrate. For example, in text-selection tasks, visual cues can be used to incrementally highlight the text whose trajectory correlates the most with eye movements, depending on the current correlation value.

Another interesting direction for future work would be to try different scripts. For example, Arabic and Hebrew are read from right to left, while Chinese, Japanese and Korean

can also be read vertically. In our implementation of Read2-Calibrate, text was revealed in a diagonal path. The flexibility of some east Asian scripts makes it possible to experiment with revealing text in different paths (e.g., a rectangular path).

### CONCLUSION

In this work we investigated the use of text as a stimulus for Pursuits, to enable gaze-based interaction and eye tracker calibration on public displays. Our results show that text can be a powerful stimulus for both tasks. Shorter pieces of text can be robustly selected using Pursuits, and text-based calibration improves gaze point accuracy. We found that Pursuits-based text selection is less error-prone when text follows circular trajectories. We also found that the use of different trajectories simultaneously (mixed trajectories) is better perceived by users and results in relatively few errors. Read2Calibrate was shown to improve the accuracy of the gaze point, in particular when using text that is gradually revealing at a speed of  $2.1^\circ/\text{s}$  and inclined at a  $315^\circ$  or  $45^\circ$  angle.

### REFERENCES

1. Florian Alt, Jörg Müller, and Albrecht Schmidt. 2012a. Advertising on Public Display Networks. *Computer* 45, 5 (2012), 50–56.
2. Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012b. How to Evaluate Public Displays. In *Proc. of PerDis '12*. ACM, New York, NY, USA.
3. Florian Alt, Alireza Sahami Shirazi, Thomas Kubitz, and Albrecht Schmidt. 2013. Interaction Techniques for Creating and Exchanging Content with Public Displays. In *Proc. of CHI '13*. ACM, New York, NY, USA, 1709–1718.
4. Florian Alt and Julia Vehns. 2016. Opportunistic Deployments: Challenges and Opportunities of Conducting Public Display Research at an Airport. In *Proc. of PerDis '16*. ACM, New York, NY, USA.
5. Ralf Biedert, Georg Buscher, Sven Schwarz, Jörn Hees, and Andreas Dengel. 2010. Text 2.0. In *Proc. of CHI EA '10*. ACM, New York, NY, USA, 4003–4008.
6. Feridun M. Celebi, Elizabeth S. Kim, Quan Wang, Carla A. Wall, and Frederick Shic. 2014. A Smooth Pursuit Calibration Technique. In *Proc. of ETRA '14*. ACM, New York, NY, USA, 377–378.
7. Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proc. of ICMI '15*. ACM, New York, NY, USA, 131–138.
8. Jorgos Coenen, Niels Wouters, and Andrew Vande Moere. 2016. Synchronized Signage on Multiple Consecutively Situated Public Displays. In *Proc. of PerDis '16*. ACM, New York, NY, USA.
9. Ashley Colley, Leena Ventä-Olkkonen, Florian Alt, and Jonna Häkkinen. 2015a. Insights from Deploying See-Through Augmented Reality Signage in the Wild. In *Proc. of PerDis '15*. ACM, New York, NY, USA, 179–185.

10. Ashley Colley, Lasse Virtanen, Jani Väyrynen, and Jonna Häkkinen. 2015b. Physically Guiding Touch Screen Interaction with Public Displays. In *Proc. of PerDis '15*. ACM, New York, NY, USA, 255–256.
11. Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz and Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7(4):1 (2014), 1–11.
12. Tilman Dingler, Tobias Bagg, Yves Grau, Niels Henze, and Albrecht Schmidt. 2015. uCanvas: A Web Framework for Spontaneous Smartphone Interaction with Ubiquitous Displays. In *Human-Computer Interaction–INTERACT 2015*. Springer, 402–409.
13. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015a. Orbits: Enabling Gaze Interaction in Smart Watches Using Moving Targets. In *Proc. of UbiComp/ISWC'15 Adjunct*. ACM, New York, NY, USA, 419–422.
14. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015b. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proc. of UIST '15*. ACM, New York, NY, USA, 457–466.
15. Luke Hespanhol, Martin Tomitsch, Ian McArthur, Joel Fredericks, Ronald Schroeter, and Marcus Foth. 2015a. Situated Interfaces for Engaging Citizens on the Go. *interactions* 23, 1 (Dec. 2015), 40–45.
16. Luke Hespanhol, Martin Tomitsch, Ian McArthur, Joel Fredericks, Ronald Schroeter, and Marcus Foth. 2015b. Vote as you go: Blending interfaces for community engagement into the urban space. In *7th International Conference on Communities and Technologies (C&T)*. ACM, University of Limerick, Limerick, Ireland, 29–37.
17. Simo Hosio, Jorge Goncalves, Vassilis Kostakos, and Jukka Riekk. 2014. Exploring Civic Engagement on Public Displays. In *User-Centric Technology Design for Nonprofit and Civic Engagements*. Springer, 91–111.
18. Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169.
19. Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Feedback for Smooth Pursuit Gaze Tracking Based Control. In *Proc. of AH '16*. ACM, New York, NY, USA, Article 6, 8 pages.
20. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2015. A Field Study on Spontaneous Gaze-based Interaction with a Public Display Using Pursuits. In *Adjunct Proc. of UbiComp '15*. ACM, New York, NY, USA, 863–872.
21. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2016. Challenges and Design Space of Gaze-enabled Public Displays. In *Adjunct Proc. of UbiComp '16*. ACM, New York, NY, USA, 10.
22. Mohamed Khamis, Andreas Bulling, and Florian Alt. 2015. Tackling Challenges of Interactive Public Displays Using Gaze. In *Adjunct Proc. of UbiComp '15*. ACM, New York, NY, USA, 763–766.
23. R. Kishi and T. Hayashi. 2015. Effective gazewriting with support of text copy and paste. In *Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on*. 125–130.
24. Lisa Koeman, Vaiva Kalnikaitė, and Yvonne Rogers. 2015. "Everyone Is Talking About It!": A Distributed Approach to Urban Voting Technology and Visualisations. In *Proc. of CHI '15*. ACM, New York, NY, USA, 3127–3136.
25. Manu Kumar, Andreas Paepcke, and Terry Winograd. 2007. EyePoint: Practical Pointing and Selection Using Gaze and Keyboard. In *Proc. of CHI '07*. ACM, New York, NY, USA, 421–430.
26. Kai Kunze, Katsutoshi Masai, Masahiko Inami, Ömer Sacakli, Marcus Liwicki, Andreas Dengel, Shoya Ishimaru, and Koichi Kise. 2015. Quantifying Reading Habits: Counting How Many Words You Read. In *Proc. of UbiComp '15*. ACM, New York, NY, USA, 87–96.
27. Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. 2015. GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays. In *Proc. of UIST '15*. ACM, New York, NY, USA, 395–404.
28. Dachuan Liu, Bo Dong, Xing Gao, and Haining Wang. 2015. Exploiting Eye Tracking for Smartphone Authentication. In *Proc. of ACNS '15*. 20.
29. Otto Hans-Martin Lutz, Antje Christine Venjakob, and Stefan Ruff. 2015. SMOOVs: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research* 8(1):2 (2015), 1–11.
30. Päivi Majaranta. 2012. Communication and Text Entry by Gaze. In *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*. IGI Global, 63–77.
31. Päivi Majaranta and Andreas Bulling. 2014. Eye Tracking and Eye-Based Human-Computer Interaction. In *Advances in Physiological Computing*. Springer, 39–65.
32. Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. 2011. Rethinking 'Multi-user': An In-the-wild Study of How Groups Approach a Walk-up-and-use Tabletop Interface. In *Proc. of CHI '11*. ACM, New York, NY, USA, 3033–3042.
33. Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In *Proc. of MM '10*. ACM, New York, NY, USA, 1285–1294.
34. Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proc. of CHI '12*. ACM, New York, NY, USA, 297–306.

## 10 TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays.

35. Erica C. Ostermann, Long Ma, Daniel Sussman, and Susan R. Fussell. 2015. CommunityConnect: An Interactive Display for Educational Residential Settings. In *Proc. of CSCW '15 Companion*. ACM, New York, NY, USA, 175–178.
36. Ken Pfeuffer, Mélodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proc. of UIST '13*. ACM, New York, NY, USA, 261–270.
37. Ken Pfeuffer, Yanxia Zhang, and Hans Gellersen. 2015. A Collaborative Gaze Aware Information Display. In *Adjunct Proc. of UbiComp '15*. ACM, New York, NY, USA, 389–391.
38. Selina Sharmin, Oleg Špakov, and Kari-Jouko Räihä. 2013. Reading On-screen Text with Gaze-based Auto-scrolling. In *Proc. of ETSA '13*. ACM, New York, NY, USA, 24–31.
39. Fabius Steinberger, Marcus Foth, and Florian Alt. 2014. Vote With Your Feet: Local Community Polling on Urban Screens. In *Proc. of PerDis '14*. ACM, New York, NY, USA, 44:44–44:49.
40. Sophie Stellmach and Raimund Dachsel. 2012. Look & Touch: Gaze-supported Target Acquisition. In *Proc. of CHI '12*. ACM, New York, NY, USA, 2981–2990.
41. Sophie Stellmach and Raimund Dachsel. 2013. Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In *Proc. of CHI '13*. ACM, New York, NY, USA, 285–294.
42. Nick Taylor, Justin Marshall, Alicia Blum-Ross, John Mills, Jon Rogers, Paul Egglestone, David M. Frohlich, Peter Wright, and Patrick Olivier. 2012. Viewpoint: Empowering Communities with Situated Voting Devices. In *Proc. of CHI '12*. ACM, New York, NY, USA, 1361–1370.
43. Takumi Toyama, Daniel Sonntag, Andreas Dengel, Takahiro Matsuda, Masakazu Iwamura, and Koichi Kise. 2014. A Mixed Reality Head-mounted Text Translation System Using Eye Gaze Input. In *Proc. of IUI '14*. ACM, New York, NY, USA, 329–334.
44. Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: Sparking Civic Discourse by a Public Interactive Poll Visualization. In *Proc. of CSCW '14*. ACM, New York, NY, USA, 1323–1332.
45. Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proc. DIS '16*. ACM, New York, NY, USA, 812–817.
46. Roel Vertegaal. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.
47. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proc. of UbiComp '13*. ACM, New York, NY, USA, 439–448.
48. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2015. Pursuits: Spontaneous Eye-Based Interaction for Dynamic Interfaces. *GetMobile: Mobile Comp. and Comm.* 18, 4 (Jan. 2015), 8–10.
49. Mélodie Vidal, Ken Pfeuffer, Andreas Bulling, and Hans W. Gellersen. 2013. Pursuits: Eye-based Interaction with Moving Targets. In *Proc. of CHI EA '13*. ACM, New York, NY, USA, 3147–3150.
50. Vasilis Vlachokyriakos, Rob Comber, Karim Ladha, Nick Taylor, Paul Dunphy, Patrick McCorry, and Patrick Olivier. 2014. PosterVote: Expanding the Action Repertoire for Local Political Activism. In *Proc. of DIS '14*. ACM, New York, NY, USA, 795–804.
51. Oleg Špakov, Poika Isokoski, Jari Kangas, Deepak Akkil, and Päivi Majaranta. 2016. PursuitAdjuster: An Exploration into the Design Space of Smooth Pursuit-based Widgets. In *Proc. of ETRA '16*. ACM, New York, NY, USA, 287–290.
52. Xuehan Xiong, Zicheng Liu, Qin Cai, and Zhengyou Zhang. 2014. Eye Gaze Tracking Using an RGBD Camera: A Comparison with a RGB Solution. In *Adjunct Proc. of UbiComp '14*. ACM, New York, NY, USA, 1113–1121.
53. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proc. of CHI '13*. ACM, New York, NY, USA, 851–860.
54. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2014. Pupil-canthy-ratio: a calibration-free method for tracking horizontal gaze direction. In *Proc. of AVI '14*. ACM, New York, NY, USA, 129–132.
55. Yanxia Zhang, MingKi Chong, Jörg Müller, Andreas Bulling, and Hans Gellersen. 2015. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing* 19, 5-6 (2015), 967–981.
56. Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proc. of UbiComp '14*. ACM, New York, NY, USA, 559–563.



# 11

**EyeVote in the Wild**

**Do Users bother Correcting System Errors on  
Public Displays?**

# EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays?

Mohamed Khamis<sup>1</sup>, Ludwig Trotter<sup>1</sup>, Markus Tessmann<sup>1</sup>, Christina Dannhart<sup>1</sup>, Andreas Bulling<sup>2</sup>, Florian Alt<sup>1</sup>

<sup>1</sup> LMU Munich, Germany

<sup>2</sup> Max Planck Institute for Informatics, Saarbrücken, Germany  
{mohamed.khamis, florian.alt}@ifi.lmu.de, bulling@mpi-inf.mpg.de

## ABSTRACT

Although recovering from errors is straightforward on most interfaces, public display systems pose very unique design challenges. Namely, public display users interact for very short amounts of times and are believed to abandon the display when interrupted or forced to deviate from the main task. To date, it is not well understood whether public display designers should enable users to correct errors (e.g. by asking users to confirm or giving them a chance correct their input), or aim for faster interaction and rely on other types of feedback to estimate errors. To close this gap, we conducted a field study where we investigated the users willingness to correct their input on public displays. We report on our findings from an in-the-wild deployment of a public gaze-based voting system where we intentionally evoked system errors to see if users correct them. We found that public display users are willing to correct system errors provided that the correction is fast and straightforward. We discuss how our findings influence the choice of interaction methods for public displays; interaction methods that are highly usable but suffer from low accuracy can still be effective if users can “undo” their interactions.

## ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces – Input Devices and Strategies

## Author Keywords

Public Displays; Smooth Pursuit; Gaze Interaction; Voting

## INTRODUCTION AND BACKGROUND

Designing interfaces for interactive public displays is often associated with unique challenges. Previous work has shown that public display users interact for short amounts of time [2, 15], and usually abandon displays if response time is slow [5] or interaction is interrupted [22].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

MUM '16, December 12 - 15, 2016, Rovaniemi, Finland

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4860-7/16/12... \$15.00

DOI: <http://dx.doi.org/10.1145/3012709.3012743>

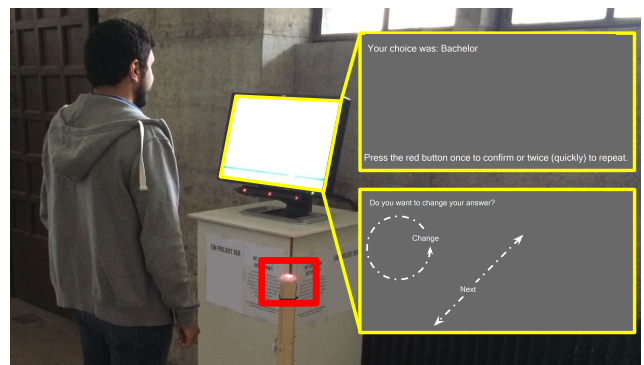


Figure 1. We investigate the willingness of public display users to correct system errors. We deployed a public gaze-based voting system, EyeVote, that occasionally shows intentionally falsified feedback and prompts users to correct their vote. In a two-days field study we experimented with two methods that allow users to correct cases where the system recognizes a selection other than the one the user intended on a situated public display. On the first day we deployed a button-based correction approach (top), while on the second day we deployed a gaze-based correction approach (bottom). Results show that most participants do correct system errors.

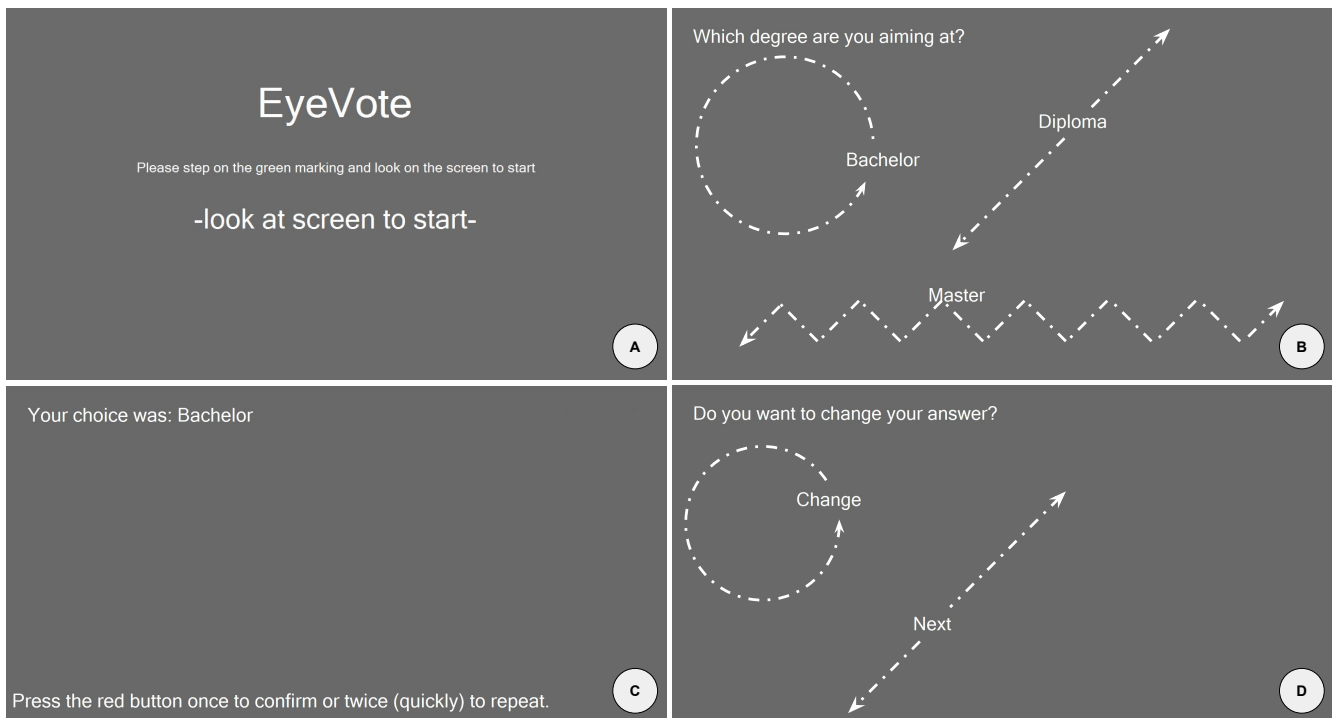
At the same time, users of public displays often need to deal with errors. This includes situations where users might want to correct typos when providing feedback or undo a selection.

This motivated us to study the willingness of users to correct their input on public displays. The question of whether to enable public display users to correct their input, or if they are not motivated to do so and hence designers should rely on other means for detecting errors, has not been addressed by prior research before.

Most relevant to our work is the work by Kukka et al. [13] who found that public display users are willing to dismiss error messages and continue interaction if the messages give users an active role (e.g. Press OK to continue). Our work builds over that by understanding if users bother correcting system errors rather than abandoning the display in frustration.

In this work we study how users behave when correcting errors and how the design of feedback mechanisms can assist in error correction. To do so, we deployed a gaze-based voting application on a public situated display (see Figure 1) in which we occasionally showed intentionally falsified confirmation messages to investigate whether users are willing to correct

## 11 EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays?



**Figure 2.** The figure above shows the different views of the EyeVote system. The white dashed arrows are to illustrate the trajectory of the options and were not shown to users. (A) The default view featured a call-to-action label asking participants to “look at the screen to start”. (B) After the eyes have been detected, the system shows the first question. The recap view shows the answer detected by the system, in addition to the option to undo the choice. On day 1 users could undo using a button (C). While on day 2 they could undo via gaze (D).

system errors, or if they will abandon the display instead. Using a voting app was motivated by recent research, showing that public displays are a promising medium for collecting votes and opinions [6, 7, 12], encouraging civic discourse [17, 19], and reflecting the population’s opinions [16]. Existing systems used touch [6], mid-air gestures [16, 19], feet [17] and recently also gaze [12] for voting in public.

We deemed gaze as interaction modality to be particularly interesting for our research [11], since it supports fast and natural interaction [23] and is hard to observe by others. The latter property is promising since making voters’ choices less obvious to surrounding users was shown to be a desired quality in public voting systems [19]. At the same time, gaze interaction is prone to input recognition errors, making it a suitable candidate for investigating users’ error correction behavior and the employment of suitable features for error correction.

To obtain a better understanding of how error correction can be implemented and how users react to it, we introduce an undo feature to our voting app that allows users to correct their input by means of different modalities, i.e., gaze and physical touch. We then deployed the system in the real world setting for two days. Findings show that users correct system errors given that the correction method is fast and straightforward.

This paper contributes an investigation of passersbys’ willingness to correct errors caused by faulty system detections.

### VOTING SYSTEM

EyeVote is a voting system for gaze-enabled public displays. The system displays questions to the public, with multiple floating textual options to choose from (see Figure 2B).

Collecting accurate gaze points requires time-consuming calibration, which is unaffordable on public displays [9]. It was necessary to either use a calibration-free technique (e.g. [21, 24, 18]) or blend the eye tracker calibration into interaction (e.g. [12]). EyeVote uses Pursuits, a calibration-free gaze-interaction technique that relies on correlating the user’s eye movements with movements of dynamic stimuli on the screen. The moving stimulus that has the highest correlation to the eye movements, is deemed to be what the user is looking at.

Previous work experimented with different types of stimuli for Pursuits, such as images [21], icons [4] abstract objects [20], game elements [8], letters [14] and digits [1]. Only recently, it was found that users are able to pursue moving text while reading it [12], which means that text strings can also be used as stimuli for the Pursuits approach.

### Pursuits Implementation

The system uses the Pearson’s product-moment coefficient to calculate the correlation between the eyes movements and the movements of the floating options. Based on pilot experiments and previous work [4, 8, 12, 21], we used the same correlation function with a threshold of 0.7 and a window size of 2000 ms. This means that every 2 seconds, the system computes Pearson’s correlation. The floating answer whose movement correlates the most with the eye movement, is deemed to be the

object the user is looking at, as long as the correlation is more than 70%. Note that previous systems used more conservative thresholds (e.g. Khamis et al. [12] used a threshold of 90% in their lab study of EyeVote). A larger threshold means that the system will not accept an answer until a high correlation is found, which increases accuracy but might result in spending longer times until the system responds. In our study, we opted for a lower value to increase the responsiveness of the system.

### System Walkthrough and Undo Feature

By default, a call-to-action label invites passersby to "Look at the screen to start" (Figure 2A). Once gaze is detected, the interface shows the first question (Figure 2B). After the system correlates the user's choice, it proceeds to a recap view, where the detected answer is shown to the user. At this point the user is presented with the choice of changing the detected answer, or proceeding to the next question. On the first day we showed the button-condition (Figure 2C), while the gaze-condition was shown on the second day (Figure 2D). Depending on the user's choice, the system either proceeds to the following question or repeats the last question. After completing 8 questions, the system resets to the default view. If the system loses track of the user's eyes it shows a warning message indicating that eyes are not detected. If eyes are not detected for 8 continuous seconds, the system restarts. The system showed straight forward questions about the favorite band, the user's study program, etc.

### Falsified answers

Due to the nature of field studies, it was not feasible to have an experimenter ask every participant whether or not the system detected their vote correctly. Hence, to investigate whether users undo system mistakes, we intentionally introduced *falsified* answers. This was done by showing an answer in the recap view that was not among the options the user had available to choose from. This way we are confident that the system is showing the user a wrong answer, and that the expected behavior is to undo the choice. In every set of 8 questions, fake answers were always shown the first time questions 3 and 6 are answered. For example, the first time a user answers question 3, the system shows "Asics" in the recap view even though it was not among the options. If that user decides to "Change" the answer, the system shows the question again, but this time the system shows the answer that was really detected.

### FIELD STUDY

We deployed a 27 inch display (1920×1200 pixels) in large university hall that is expected to be reached by many students, academics, and university staff members. The display was equipped with a Tobii EyeX Controller. A squared marker was placed on the floor with a distance of one meter to the display to guide passersby to standing in the eye tracker's range.

### Design

The study ran for two days and covered two conditions: (1) button-based undo, and (2) gaze-based undo.

On the first day, we deployed the button-based undo approach. A red button was placed next to the display, participants were

asked to press the button once to proceed to the next question, or twice to repeat the question (see Figures 1 and 2C).

On the second day, we deployed the gaze-based undo approach. After each question, participants were asked if they want to change their answer (Figure 2D), and were shown two floating textual options that can be selected via Pursuits. Following the circular trajectory of the word "Change" would repeat the last question again. While following the linear trajectory of the word "Next" would show the following question.

### Results

We logged the raw gaze data, all presented and selected answers, as well as button and gaze based undos. The system launched the first question 187 times in total during two days. This means that there were 187 instances where the system detected a user standing within the marked area and facing the screen. We refer to these instances as "interactions".

On Day 1 there were 106 interactions, that is, gaze was detected at 106 different instances. Out of which, there were 49 instances where at least one question was answered. In total, 220 questions were answered on the first day. On Day 2, at least one question was answered in 30 out of the 81 times in which a user interaction was detected. In total, 243 questions were answered on the second day.

### Undos

We distinguish two cases where the "undo" feature was used by users: (1) cases where users corrected falsified answers, and (2) cases where users corrected unaltered answers. The former are cases where we are confident that the system showed a wrong answer that is was not among the choices, while the latter are cases where the system might or might not have shown a wrong answer (see Figure 3).

Out of the 220 questions answered on Day 1, 37 were falsified answers. Out of those, 22 answers (59.5%) were changed using the undo feature. While on Day 2, there were 31 falsified answers out of 243 answered questions. Users corrected 27 out of 31 (87%) falsified answers.

Note that users did not have any motivator to correct their input apart from their intrinsic motivation. Therefore, we interpret the correction rates of 59.5% and 87% as indicators that the users are willing to correct input mistakes.

The undo rate for the gaze method is higher compared to the button method. We attribute this to the use of the same interaction method for selection and correction which is seemingly better accepted by the users.

### User dropouts

As this field study was not supervised, users could join and leave at any time. We relied on the number of dropouts as a measure of satisfaction and how users cope with the system. We define a dropout as a situation where the system lost eye contact for more than 8 continuous seconds.

Figure 4 illustrates the number of users who answered each question, as well as the number of users who dropped out after answering each question. It is noticeable that 20 out of 29 users (41%) dropped out after answering the first question in



## 11 EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays?

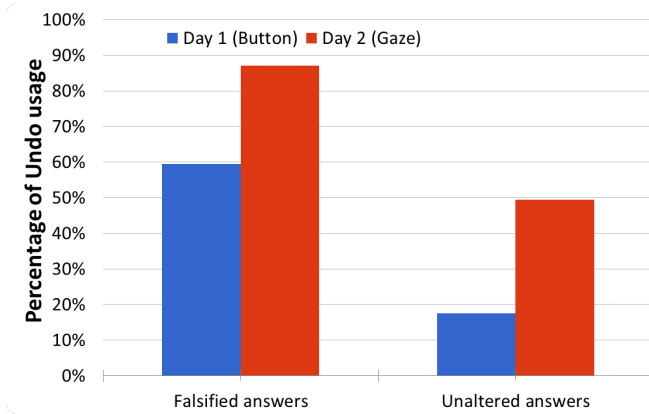


Figure 3. The figure illustrates the relative undos for falsified and unaltered questions for each day. The figure shows that users corrected the majority of the falsified answers on both days (59.5% and 87% respectively). Gaze-based undo was used more frequently than the button-based approach, we believe that this is due to the use of gaze for both selection and correction, which was better perceived by users.

the first day in which the button-based undo was deployed. While on the second day, relatively few users dropped out after the first question, but relatively more dropped out after questions 3 and 6, in which the falsified answers were shown.

Interestingly, the number of users who answered all eight questions is very similar on both days: 20.4% on Day 1, and 20% on Day 2.

### DISCUSSION

To our surprise, many users (59.5% and 87% on both days respectively) corrected their input when we displayed falsified answers. Although prior work reports that people are likely to abandon the display if they find it unresponsive or faulty [2]. Our results indicate that passersby are willing to fix system mistakes as long as the option is available and feasible.

By examining Figure 4, we expect that the high dropout rate for the first question on the first day is due to the use of two separate modalities for selection and confirmation; we used gaze for voting, and the button for confirming the answer. On the other hand, although there are less dropouts at the first question on the second day, there are more dropouts at the falsified answers. By the time users reached the falsified questions (questions 3 and 6) they had already answered and confirmed many questions via gaze. Hence we expect that the temporal demand led to fatigue, which discouraged some users from completing the 8 questions. Previous work had reported that performing multiple consecutive gaze-based selections is tiring to the users [12].

The results suggest that users are indeed willing to correct system errors. Figure 4 shows that although relatively more users dropout at system errors, the majority of users corrects them and continues interacting. With this conclusion we encourage the use of highly usable interaction techniques even if this leads to sacrificing some system accuracy, and relying on users to correct the occasional system errors. Taking EyeVote as an example, future systems can offer a *dynamic undo function*,

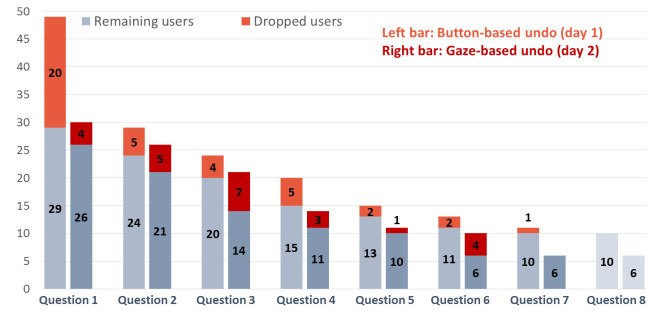


Figure 4. The figure reflects the number of dropped-out users; for example, on the first day 49 users answered question 1, 20 of which dropped out afterwards. We suspect the lower number of dropouts at the first question in day 2 is due to using the same modality for selection and confirmation. While the relatively higher dropout rates in the second day at the falsified questions (questions 3 and 6), is likely due to eye fatigue.

which can be realized by introducing an additional threshold; if the highest correlation between the eye movements and one of the trajectories is higher than the selection threshold (set to 70% in our implementation), a more conservative threshold is checked (e.g. 90%) and if the correlation is lower than that (i.e. between 70% and 90%) the user is presented with a confirmation message, otherwise the user proceeds to the following question.

Although the use of a button to confirm or revoke an interaction is intuitive, the results suggest that users were more likely to dropout in the case of the button-based undo compared to the gaze-based approach. We believe that the use of the same modality for interaction and confirmation maintains a straightforward flow and is less confusing to the passersby. Hence we recommend future systems to use the same modality for both interaction and confirmation.

While we investigated the willingness to correct input on a gaze-enabled display, the results are also applicable to other modalities. For example, systems can reduce dwell times when selecting via mid-air hand gestures; this increases the responsiveness of the system but makes it more error prone, hence the system should also allow users to undo their actions.

### CONCLUSION AND FUTURE WORK

Our results indicate that users will tolerate input errors and correct them if the system allows that. Therefore, we encourage the use of highly usable metrics even if they reduce accuracy provided that correction mechanisms are implemented. In our study, users corrected most of the falsified inputs (59.5% for the button-based undo and 87% for the gaze-based undo). Although the dropout rate was higher when falsified answers were shown during the gaze-based undo condition, the majority of the users corrected them before proceeding to the following question.

In this work we evaluated an undo feature for public displays using Pursuits. One direction for future work is to experiment with other gaze-input methods such as gaze gestures [3, 10]. Additionally, more modalities can be experimented with and compared such as mid-air hand gestures, touch and head gestures (e.g. nodding to confirm interactions).

## REFERENCES

1. Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz and Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7(4):1 (2014), 1–11.
2. Nigel Davies, Sarah Clinch, and Florian Alt. 2014. *Pervasive Displays: Understanding the Future of Digital Signage* (1st ed.). Morgan & Claypool Publishers.
3. Heiko Drewes and Albrecht Schmidt. 2007. *Interacting with the Computer Using Gaze Gestures*. Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488. DOI: [http://dx.doi.org/10.1007/978-3-540-74800-7\\_43](http://dx.doi.org/10.1007/978-3-540-74800-7_43)
4. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 457–466. DOI: <http://dx.doi.org/10.1145/2807442.2807499>
5. Miriam Greis, Florian Alt, Niels Henze, and Nemanja Memarovic. 2014. I Can Wait a Minute: Uncovering the Optimal Delay Time for Pre-moderated User-generated Content on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1435–1438. DOI: <http://dx.doi.org/10.1145/2556288.2557186>
6. Luke Hespanhol, Martin Tomitsch, Ian McArthur, Joel Fredericks, Ronald Schroeter, and Marcus Foth. 2015a. Situated Interfaces for Engaging Citizens on the Go. *interactions* 23, 1 (Dec. 2015), 40–45.
7. Luke Hespanhol, Martin Tomitsch, Ian McArthur, Joel Fredericks, Ronald Schroeter, and Marcus Foth. 2015b. Vote as you go: Blending interfaces for community engagement into the urban space. In *7th International Conference on Communities and Technologies (C&T)*. ACM, University of Limerick, Limerick, Ireland, 29–37.
8. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2015. A Field Study on Spontaneous Gaze-based Interaction with a Public Display Using Pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. ACM, New York, NY, USA, 863–872. DOI: <http://dx.doi.org/10.1145/2800835.2804335>
9. Mohamed Khamis, Florian Alt, and Andreas Bulling. 2016a. Challenges and Design Space of Gaze-enabled Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 1736–1745. DOI: <http://dx.doi.org/10.1145/2968219.2968342>
10. Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016b. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proceedings of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 6. DOI: <http://dx.doi.org/10.1145/2851581.2892314>
11. Mohamed Khamis, Andreas Bulling, and Florian Alt. 2015. Tackling Challenges of Interactive Public Displays Using Gaze. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. ACM, New York, NY, USA, 763–766. DOI: <http://dx.doi.org/10.1145/2800835.2807951>
12. Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 12. DOI: <http://dx.doi.org/10.1145/2971648.2971679>
13. Hannu Kukka, Jorge Goncalves, Tommi Heikkinen, Olli-Pekka Suua, Yifei Zuo, Hannu Raappana, Mohamed Abdellatif, Olli Okkonen, Raúl Jiménez, and Timo Ojala. 2015. Touch OK to Continue: Error Messages and Affective Response on Interactive Public Displays. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*. ACM, New York, NY, USA, 99–105. DOI: <http://dx.doi.org/10.1145/2757710.2757723>
14. Otto Hans-Martin Lutz, Antje Christine Venjakob, and Stefan Ruff. 2015. SMOOVS: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research* 8(1):2 (2015), 1–11.
15. Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 297–306. DOI: <http://dx.doi.org/10.1145/2207676.2207718>
16. Erica C. Ostermann, Long Ma, Daniel Sussman, and Susan R. Fussell. 2015. CommunityConnect: An Interactive Display for Educational Residential Settings. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW'15 Companion)*. ACM, New York, NY, USA, 175–178. DOI: <http://dx.doi.org/10.1145/2685553.2699000>
17. Fabius Steinberger, Marcus Foth, and Florian Alt. 2014. Vote With Your Feet: Local Community Polling on Urban Screens. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. ACM, New York, NY, USA, Article 44, 6 pages. DOI: <http://dx.doi.org/10.1145/2611009.2611015>

## 11 EyeVote in the Wild: Do Users bother Correcting System Errors on Public Displays?

18. Yusuke Sugano, Xucong Zhang, and Andreas Bulling. 2016. AggreGaze: Collective Estimation of Audience Attention on Public Displays. In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*.
19. Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: Sparking Civic Discourse by a Public Interactive Poll Visualization. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1323–1332. DOI : <http://dx.doi.org/10.1145/2531602.2531639>
20. Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 812–817. DOI : <http://dx.doi.org/10.1145/2901790.2901867>
21. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 439–448. DOI : <http://dx.doi.org/10.1145/2493432.2493477>
22. Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 841–850. DOI : <http://dx.doi.org/10.1145/2470654.2470774>
23. Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 246–253. DOI : <http://dx.doi.org/10.1145/302979.303053>
24. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 851–860. DOI : <http://dx.doi.org/10.1145/2470654.2470775>



## BIBLIOGRAPHY

- [1] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Presented in Denver, Colorado, USA. ACM, New York, NY, USA, 3751–3763. DOI:<http://dx.doi.org/10.1145/3025453.3025461>
- [2] Christopher Ackad, Andrew Clayphan, Martin Tomitsch, and Judy Kay. 2015. An In-the-wild Study of Learning Mid-air Gestures to Browse Hierarchical Information at a Large Interactive Public Display. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Presented in Osaka, Japan. ACM, New York, NY, USA, 1227–1238. DOI:<http://dx.doi.org/10.1145/2750858.2807532>
- [3] Christopher Ackad, Martin Tomitsch, and Judy Kay. 2016. Skeletons and Silhouettes: Comparing User Representations at a Gesture-based Large Display. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Presented in San Jose, California, USA. ACM, New York, NY, USA, 2343–2347. DOI:<http://dx.doi.org/10.1145/2858036.2858427>
- [4] Hazim Almuhiemedi, Florian Schaub, Norman Sadeh, Idris Adjeric, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location Has Been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Presented in Seoul, Republic of Korea. ACM, New York, NY, USA, 787–796. DOI:<http://dx.doi.org/10.1145/2702123.2702210>
- [5] Florian Alt, Andreas Bulling, Lukas Mecke, and Daniel Buschek. 2016. Attention, Please!: Comparing Features for Measuring Audience Attention Towards Pervasive Displays. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. Presented in Brisbane, QLD, Australia. ACM, New York, NY, USA, 823–828. DOI:<http://dx.doi.org/10.1145/2901790.2901897>

- 
- [6] Florian Alt, Jörg Müller, and Albrecht Schmidt. 2012. Advertising on Public Display Networks. *Computer* 45, 5 (May 2012), 50–56. DOI:<http://dx.doi.org/10.1109/MC.2012.150>
- [7] Florian Alt, Stefan Schneegass, Alireza Sahami Shirazi, Mariam Hassib, and Andreas Bulling. 2015. Graphical Passwords in the Wild: Understanding How Users Choose Pictures and Passwords in Image-based Authentication Schemes. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. Presented in Copenhagen, Denmark. ACM, New York, NY, USA, 316–322. DOI:<http://dx.doi.org/10.1145/2785830.2785882>
- [8] Florian Alt and Julia Vehns. 2016. Opportunistic Deployments: Challenges and Opportunities of Conducting Public Display Research at an Airport. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*. Presented in Oulu, Finland. ACM, New York, NY, USA, 106–117. DOI: <http://dx.doi.org/10.1145/2914920.2915020>
- [9] Adam J. Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. 2010. Smudge Attacks on Smartphone Touch Screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies (WOOT'10)*. Presented in Washington, DC. USENIX Association, Berkeley, CA, USA, 1–7. <http://dl.acm.org/citation.cfm?id=1925004.1925009>
- [10] Sileye O. Ba and Jean-Marc Odobez. 2011. Multiperson Visual Focus of Attention from Head Pose and Meeting Contextual Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 1 (Jan 2011), 101–116. DOI: <http://dx.doi.org/10.1109/TPAMI.2010.69>
- [11] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. 2013. "Little Brothers Watching You": Raising Awareness of Data Leaks on Smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS '13)*. Presented in Newcastle, United Kingdom. ACM, New York, NY, USA, Article 12, 11 pages. DOI:<http://dx.doi.org/10.1145/2501604.2501616>
- [12] Vikas Behal and Sania Sareen. 2014. Guerilla marketing: A low cost marketing strategy. *International Journal of Management Research and Business Strategy* 3, 1 (2014), 1–6.
- [13] Richard A. Bolt. 1981. Gaze-orchestrated Dynamic Windows. *SIGGRAPH Comput. Graph.* 15, 3 (Aug. 1981), 109–119. DOI:<http://dx.doi.org/10.1145/965161.806796>
- [14] Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT '03)*. 17–24.

- 
- [15] Frederik Brudy, David Ledo, Saul Greenberg, and Andreas Butz. 2014. Is Anyone Looking? Mitigating Shoulder Surfing on Public Displays Through Awareness and Protection. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. Presented in Copenhagen, Denmark. ACM, New York, NY, USA, Article 1, 6 pages. DOI:<http://dx.doi.org/10.1145/2611009.2611028>
- [16] Andreas Bulling and Kai Kunze. 2016. Eyewear Computers for Human-computer Interaction. *interactions* 23, 3 (April 2016), 70–73. DOI:<http://dx.doi.org/10.1145/2912886>
- [17] Sandy Claes and Andrew Vande Moere. 2013. Street Infographics: Raising Awareness of Local Issues Through a Situated Urban Visualization. In *Proceedings of the 2Nd ACM International Symposium on Pervasive Displays (PerDis '13)*. Presented in Mountain View, California. ACM, New York, NY, USA, 133–138. DOI:<http://dx.doi.org/10.1145/2491568.2491597>
- [18] Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7(4):1 (2014), 1–11.
- [19] Nicholas S. Dalton, Emily Collins, and Paul Marshall. 2015. Display Blindness?: Looking Again at the Visibility of Situated Displays Using Eye-tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Presented in Seoul, Republic of Korea. ACM, New York, NY, USA, 3889–3898. DOI:<http://dx.doi.org/10.1145/2702123.2702150>
- [20] Nigel Davies, Sarah Clinch, and Florian Alt. 2014a. *Pervasive Displays: Understanding the Future of Digital Signage* (1st ed.). Morgan & Claypool Publishers.
- [21] Nigel Davies, Marc Langheinrich, Sarah Clinch, Ivan Elhart, Adrian Friday, Thomas Kubitz, and Bholanathsingh Surajbali. 2014b. Personalisation and Privacy in Future Pervasive Display Networks. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. Presented in Toronto, Ontario, Canada. ACM, New York, NY, USA, 2357–2366. DOI:<http://dx.doi.org/10.1145/2556288.2557287>
- [22] Connor Dickie, Roel Vertegaal, Changuk Sohn, and Daniel Cheng. 2005. eye-Look: Using Attention to Facilitate Mobile Media Consumption. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. Presented in Seattle, WA, USA. ACM, New York, NY, USA, 103–106. DOI:<http://dx.doi.org/10.1145/1095034.1095050>



- 
- [23] Tilman Dingler, Tobias Bagg, Yves Grau, Niels Henze, and Albrecht Schmidt. 2015. uCanvas: A Web Framework for Spontaneous Smartphone Interaction with Ubiquitous Displays. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Lecture Notes in Computer Science, Vol. 9298. Springer International Publishing, 402–409. DOI:[http://dx.doi.org/10.1007/978-3-319-22698-9\\_27](http://dx.doi.org/10.1007/978-3-319-22698-9_27)
- [24] Michael D. Dodd, John R. Hibbing, and Kevin B. Smith. 2011. The politics of attention: gaze-cuing effects are moderated by political temperament. *Attention, Perception, & Psychophysics* 73, 1 (01 Jan 2011), 24–29. DOI:<http://dx.doi.org/10.3758/s13414-010-0001-x>
- [25] Heiko Drewes and Albrecht Schmidt. 2007. *Interacting with the Computer Using Gaze Gestures*. Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488. DOI: [http://dx.doi.org/10.1007/978-3-540-74800-7\\_43](http://dx.doi.org/10.1007/978-3-540-74800-7_43)
- [26] Andrew T. Duchowski. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4 (01 Nov 2002), 455–470. DOI:<http://dx.doi.org/10.3758/BF03195475>
- [27] Andrew T. Duchowski. 2017. *Diversity and Types of Eye Tracking Applications*. Springer International Publishing, Cham, 247–248. DOI:[http://dx.doi.org/10.1007/978-3-319-57883-5\\_20](http://dx.doi.org/10.1007/978-3-319-57883-5_20)
- [28] Marc Eaddy, Gabor Blasko, Jason Babcock, and Steven Feiner. 2004. My Own Private Kiosk: Privacy-Preserving Public Displays. In *Proceedings of the Eighth International Symposium on Wearable Computers (ISWC '04)*. IEEE Computer Society, Washington, DC, USA, 132–135. DOI:<http://dx.doi.org/10.1109/ISWC.2004.32>
- [29] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Presented in Denver, Colorado, USA. ACM, New York, NY, USA, 11.
- [30] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. Presented in Charlotte, NC, USA. ACM, New York, NY, USA, 457–466. DOI:<http://dx.doi.org/10.1145/2807442.2807499>
- [31] Augusto Esteves, David Verweij, Liza Suraiya, Rasel Islam, Youryang Lee, and Ian Oakley. 2017. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the 30th Annual ACM Symposium on*



- User Interface Software and Technology (UIST '17)*. Presented in Quebec City, QC, Canada. ACM, New York, NY, USA, 167–178. DOI:<http://dx.doi.org/10.1145/3126594.3126616>
- [32] Kit Galloway and Sherrie Rabinowitz. 1980. Hole-In-Space. Website. (1980). Retrieved March 7, 2017 from <http://www.ecafe.com/getty/HIS/>.
- [33] Vito Gentile, Mohamed Khamis, Salvatore Sorce, and Florian Alt. 2017. They Are Looking at Me!: Understanding How Audience Presence Impacts on Public Display Users. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*. Presented in Lugano, Switzerland. ACM, New York, NY, USA, Article 11, 7 pages. DOI:<http://dx.doi.org/10.1145/3078810.3078822>
- [34] Jorge Goncalves, Denzil Ferreira, Simo Hosio, Yong Liu, Jakob Rogstadius, Hannu Kukka, and Vassilis Kostakos. 2013. Crowdsourcing on the Spot: Altruistic Use of Public Displays, Feasibility, Performance, and Behaviours. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. Presented in Zurich, Switzerland. ACM, New York, NY, USA, 753–762. DOI:<http://dx.doi.org/10.1145/2493432.2493481>
- [35] Miriam Greis, Florian Alt, Niels Henze, and Nemanja Memarovic. 2014. I Can Wait a Minute: Uncovering the Optimal Delay Time for Pre-moderated User-generated Content on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Presented in Toronto, Ontario, Canada. ACM, New York, NY, USA, 1435–1438. DOI:<http://dx.doi.org/10.1145/2556288.2557186>
- [36] Jonna Häkkinä, Olli Koskenranta, Maaret Posti, Leena Ventä-Olkkonen, and Ashley Colley. 2013. Clearing the Virtual Window: Connecting Two Locations with Interactive Public Displays. In *Proceedings of the 2Nd ACM International Symposium on Pervasive Displays (PerDis '13)*. Presented in Mountain View, California. ACM, New York, NY, USA, 85–90. DOI:<http://dx.doi.org/10.1145/2491568.2491587>
- [37] Benjamin Hatscher, Maria Luz, Lennart E. Nacke, Norbert Elkmann, Veit Müller, and Christian Hansen. 2017. GazeTap: Towards Hands-free Interaction in the Operating Room. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*. Presented in Glasgow, UK. ACM, New York, NY, USA, 243–251. DOI:<http://dx.doi.org/10.1145/3136755.3136759>
- [38] Luke Hespanhol, Martin Tomitsch, Ian McArthur, Joel Fredericks, Ronald Schroeter, and Marcus Foth. 2015. Situated Interfaces for Engaging Citizens on the Go. *interactions* 23, 1 (Dec. 2015), 40–45. DOI:<http://dx.doi.org/10.1145/2851200>

- 
- [39] Norman Höller, Johann Schrammel, Manfred Tscheligi, and Lucas Paletta. 2009. The Perception of Information and Advertisement Screens Mounted in Public Transportation Vehicles-Results from a Mobile Eye-tracking Study. In *Proceedings of the 2nd Workshop on Pervasive Advertising*.
- [40] Thomas E. Hutchinson, K. Preston White, Worthy N. Martin, Kelly C. Reichert, and Lisa A. Frey. 1989. Human-computer interaction using eye-gaze input. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 6 (Nov 1989), 1527–1534. DOI:<http://dx.doi.org/10.1109/21.44068>
- [41] Tina In-Albon and Silvia Schneider. 2010. *Using Eye Tracking Methodology in Children with Anxiety Disorders*. John Wiley & Sons, Ltd, 129–149. DOI:<http://dx.doi.org/10.1002/9780470661468.ch6>
- [42] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. Presented in Seattle, Washington, USA. ACM, New York, NY, USA, 11–18. DOI:<http://dx.doi.org/10.1145/97243.97246>
- [43] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169. DOI:<http://dx.doi.org/10.1145/123078.128728>
- [44] Shahram Jalaliniya and Diako Mardanbegi. 2016. EyeGrip: Detecting Targets in a Series of Uni-directional Moving Objects Using Optokinetic Nystagmus Eye Movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Presented in San Jose, California, USA. ACM, New York, NY, USA, 5801–5811. DOI:<http://dx.doi.org/10.1145/2858036.2858584>
- [45] Louis Émile Javal. 1878. Essai sur la physiologie de la lecture. *Annales d'oculistique* 79, 97 (1878).
- [46] Warren Jones and Ami Klin. 2013. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature* 504, 7480 (Dec 2013), 427–431.
- [47] Wendy Ju and David Sirkin. 2010. Animate Objects: How Physical Motion Encourages Public Interaction. In *Persuasive Technology*, Thomas Ploug, Per Hasle, and Harri Oinas-Kukkonen (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 40–51.
- [48] Christopher Kaiser and Michael Schmitz. 2015. PaperFlow: A Gamified Multitouch Application to Increase Motivation and Participation in Scientific Research. In *Proceedings of the 4th International Symposium on Pervasive Displays*

- (PerDis '15). Presented in Saarbruecken, Germany. ACM, New York, NY, USA, 265–266. DOI:<http://dx.doi.org/10.1145/2757710.2776814>
- [49] Rudolf Kajan, Adam Herout, Roman Bednarik, and Filip Povolný. 2016. PeepList: Adapting ex-post interaction with pervasive display content using eye tracking. *Pervasive and Mobile Computing* 30, C (Aug. 2016), 71–83. DOI: <http://dx.doi.org/10.1016/j.pmcj.2015.12.004>
- [50] Jakob Karolus, Paweł W. Wozniak, Lewis L. Chuang, and Albrecht Schmidt. 2017. Robust Gaze Features for Enabling Language Proficiency Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Presented in Denver, Colorado, USA. ACM, New York, NY, USA, 2998–3010. DOI:<http://dx.doi.org/10.1145/3025453.3025601>
- [51] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. Presented in Seattle, Washington. ACM, New York, NY, USA, 1151–1160. DOI: <http://dx.doi.org/10.1145/2638728.2641695>
- [52] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2015. A Field Study on Spontaneous Gaze-based Interaction with a Public Display Using Pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. Presented in Osaka, Japan. ACM, New York, NY, USA, 863–872. DOI:<http://dx.doi.org/10.1145/2800835.2804335>
- [53] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2016. Challenges and Design Space of Gaze-enabled Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. Presented in Heidelberg, Germany. ACM, New York, NY, USA, 1736–1745. DOI:<http://dx.doi.org/10.1145/2968219.2968342>
- [54] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2018a. The Past, Present, and Future of Gaze-enabled Handheld Mobile Devices: Survey and Lessons Learned. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. Presented in Barcelona, Spain. ACM, New York, NY, USA.
- [55] Mohamed Khamis, Christian Becker, Andreas Bulling, and Florian Alt. 2018b. Which One is Me?: Identifying Oneself on Public Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Presented in Montreal QC, Canada. ACM, New York, NY, USA, Article 287, 12 pages. DOI:<http://dx.doi.org/10.1145/3173574.3173861>

- 
- [56] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018c. VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI '18)*. Presented in Grosseto, Italy. ACM, New York, NY, USA, 7.
- [57] Lisa Koeman, Vaiva Kalnikaitė, and Yvonne Rogers. 2015. "Everyone is Talking About It!": A Distributed Approach to Urban Voting Technology and Visualisations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Presented in Seoul, Republic of Korea. ACM, New York, NY, USA, 3127–3136. DOI:<http://dx.doi.org/10.1145/2702123.2702263>
- [58] Lisa Koeman, Vaiva Kalnikaitė, Yvonne Rogers, and Jon Bird. 2014. What Chalk and Tape Can Tell Us: Lessons Learnt for Next Generation Urban Displays. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. Presented in Copenhagen, Denmark. ACM, New York, NY, USA, Article 130, 6 pages. DOI:<http://dx.doi.org/10.1145/2611009.2611018>
- [59] Thomas Kubitz, Sarah Clinch, Nigel Davies, and Marc Langheinrich. 2013. Using Mobile Devices to Personalize Pervasive Displays. *SIGMOBILE Mob. Comput. Commun. Rev.* 16, 4 (Feb. 2013), 26–27. DOI:<http://dx.doi.org/10.1145/2436196.2436211>
- [60] Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Gonçalves, and Timo Ojala. 2013. What Makes You Click: Exploring Visual Signals to Entice Interaction on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Presented in Paris, France. ACM, New York, NY, USA, 1699–1708. DOI:<http://dx.doi.org/10.1145/2470654.2466225>
- [61] Michael F. Land and Sophie Furneaux. 1997. The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 352, 1358 (1997), 1231–1239. DOI:<http://dx.doi.org/10.1098/rstb.1997.0105>
- [62] Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. 2015a. GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. Presented in Charlotte, NC, USA. ACM, New York, NY, USA, 395–404. DOI:<http://dx.doi.org/10.1145/2807442.2807479>
- [63] Christian Lander, Marco Speicher, Denise Paradowski, Norine Coenen, Sebastian Biewer, and Antonio Krüger. 2015b. Collaborative Newspaper Demo: Exploring an Adaptive Scrolling Algorithm in a Multi-user Reading Scenario. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*.

- Presented in Saarbruecken, Germany. ACM, New York, NY, USA, 271–272. DOI:<http://dx.doi.org/10.1145/2757710.2776817>
- [64] Christian Lander, Marco Speicher, Denise Paradowski, Norine Coenen, Sebastian Biewer, and Antonio Krüger. 2015c. Collaborative Newspaper: Exploring an Adaptive Scrolling Algorithm in a Multi-user Reading Scenario. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*. Presented in Saarbruecken, Germany. ACM, New York, NY, USA, 163–169. DOI:<http://dx.doi.org/10.1145/2757710.2757734>
- [65] James L. Levine. 1984. Performance of an eyetracker for office use. *Computers in Biology and Medicine* 14, 1 (1984), 77 – 89. DOI:[http://dx.doi.org/https://doi.org/10.1016/0010-4825\(84\)90022-2](http://dx.doi.org/https://doi.org/10.1016/0010-4825(84)90022-2)
- [66] Hong Liu. 2017. *Chinese business: Landscapes and strategies*. Taylor & Francis.
- [67] Otto Hans-Martin Lutz, Antje Christine Venjakob, and Stefan Ruff. 2015. SMOOVs: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research* 8, 1 (2015). <https://bop.unibe.ch/index.php/JEMR/article/view/2394>
- [68] Paul P. Maglio, Teenie Matlock, Christopher S. Campbell, Shumin Zhai, and Barton A. Smith. 2000. Gaze and Speech in Attentive User Interfaces. In *Advances in Multimodal Interfaces — ICMI 2000*, Tieniu Tan, Yuanchun Shi, and Wen Gao (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–7.
- [69] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human–Computer Interaction*. Springer London, 39–65. DOI:[http://dx.doi.org/10.1007/978-1-4471-6392-3\\_3](http://dx.doi.org/10.1007/978-1-4471-6392-3_3)
- [70] Päivi Majaranta and Kari-Jouko Räihä. 2002. Twenty Years of Eye Typing: Systems and Design Issues. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications (ETRA '02)*. Presented in New Orleans, Louisiana. ACM, New York, NY, USA, 15–22. DOI:<http://dx.doi.org/10.1145/507072.507076>
- [71] Ville Mäkelä, Jobin James, Tuuli Keskinen, Jaakko Hakulinen, and Markku Turunen. 2017. “It’s Natural to Grab and Pull”: Retrieving Content from Large Displays Using Mid-Air Gestures. *IEEE Pervasive Computing* 16, 3 (2017), 70–77. DOI:<http://dx.doi.org/10.1109/MPRV.2017.2940966>
- [72] Ville Mäkelä, Mohamed Khamis, Lukas Mecke, Jobin James, Markku Turunen, and Florian Alt. 2018. Pocket Transfers: Interaction Techniques for Transferring Content from Situated Displays to Mobile Devices. (2018), 13. DOI:<http://dx.doi.org/10.1145/3173574.3173709>



- 
- [73] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Presented in Seoul, Republic of Korea. ACM, New York, NY, USA, 2953–2962. DOI:<http://dx.doi.org/10.1145/2702123.2702539>
- [74] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. 2011. Rethinking 'Multi-user': An In-the-wild Study of How Groups Approach a Walk-up-and-use Tabletop Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Presented in Vancouver, BC, Canada. ACM, New York, NY, USA, 3033–3042. DOI:<http://dx.doi.org/10.1145/1978942.1979392>
- [75] Nemanja Memarovic. 2015. Public Photos, Private Concerns: Uncovering Privacy Concerns of User Generated Content Created Through Networked Public Displays. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*. Presented in Saarbruecken, Germany. ACM, New York, NY, USA, 171–177. DOI:<http://dx.doi.org/10.1145/2757710.2757739>
- [76] Nemanja Memarovic, Marc Langheinrich, and Florian Alt. 2012. Interacting places – A framework for promoting community interaction and place awareness through public displays. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. 327–430. DOI:<http://dx.doi.org/10.1109/PerComW.2012.6197526>
- [77] Alexander Meschtscherjakov, Wolfgang Reitberger, Michael Lankes, and Manfred Tscheligi. 2008. Enhanced Shopping: A Dynamic Map in a Retail Store. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*. Presented in Seoul, Korea. ACM, New York, NY, USA, 336–339. DOI: <http://dx.doi.org/10.1145/1409635.1409680>
- [78] Daniel Michelis and Jörg Müller. 2011. The Audience Funnel: Observations of Gesture Based Interaction With Multiple Large Displays in a City Center. *International Journal of Human–Computer Interaction* 27, 6 (2011), 562–579. DOI: <http://dx.doi.org/10.1080/10447318.2011.555299>
- [79] Keaton Mowery, Sarah Meiklejohn, and Stefan Savage. 2011. Heat of the Moment: Characterizing the Efficacy of Thermal Camera-based Attacks. In *Proceedings of the 5th USENIX Conference on Offensive Technologies (WOOT'11)*. Presented in San Francisco, CA. USENIX Association, Berkeley, CA, USA, 6–6. <http://dl.acm.org/citation.cfm?id=2028052>.2028058
- [80] Omar Mubin, Tatiana Lashina, and Evert van Loenen. 2009. How Not to Become a Buffoon in Front of a Shop Window: A Solution Allowing Nat-

- ural Head Movement for Interaction with a Public Display. In *Proc. of INTERACT '09*. Springer Berlin Heidelberg, Berlin, Heidelberg, 250–263. DOI: [http://dx.doi.org/10.1007/978-3-642-03658-3\\_32](http://dx.doi.org/10.1007/978-3-642-03658-3_32)
- [81] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. Presented in Firenze, Italy. ACM, New York, NY, USA, 1285–1294. DOI:<http://dx.doi.org/10.1145/1873951.1874203>
- [82] Jörg Müller, Juliane Exeler, Markus Buzeck, and Antonio Krüger. 2009. ReflectiveSigns: Digital Signs That Adapt to Audience Attention. In *Pervasive Computing*, Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. J. Bernheim Brush, and Yoshito Tobe (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 17–24.
- [83] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Presented in Austin, Texas, USA. ACM, New York, NY, USA, 297–306. DOI:<http://dx.doi.org/10.1145/2207676.2207718>
- [84] Pai Chet Ng, James She, Kang Eun Jeon, and Matthias Baldauf. 2017. When Smart Devices Interact With Pervasive Screens: A Survey. *ACM Trans. Multimedia Comput. Commun. Appl.* 13, 4, Article 55 (Aug. 2017), 23 pages.
- [85] Timo Ojala, Hannu Kukka, Tomas Lindén, Tommi Heikkinen, Marko Jurmu, Simo Hosio, and Fabio Kruger. 2010. UBI-Hotspot 1.0: Large-Scale Long-Term Deployment of Interactive Public Displays in a City Center. In *2010 Fifth International Conference on Internet and Web Applications and Services*. 285–294. DOI: <http://dx.doi.org/10.1109/ICIW.2010.49>
- [86] Erica C. Ostermann, Long Ma, Daniel Sussman, and Susan R. Fussell. 2015. CommunityConnect: An Interactive Display for Educational Residential Settings. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW'15 Companion)*. Presented in Vancouver, BC, Canada. ACM, New York, NY, USA, 175–178. DOI:<http://dx.doi.org/10.1145/2685553.2699000>
- [87] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. 2008. It's Mine, Don't Touch!: Interactions at a Large Multi-touch Display in a City Centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Presented in Florence, Italy. ACM, New York, NY, USA, 1285–1294. DOI: <http://dx.doi.org/10.1145/1357054.1357255>

- 
- [88] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. Presented in Honolulu, Hawaii, USA. ACM, New York, NY, USA, 509–518. DOI:<http://dx.doi.org/10.1145/2642918.2647397>
- [89] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. Presented in St. Andrews, Scotland, United Kingdom. ACM, New York, NY, USA, 261–270. DOI:<http://dx.doi.org/10.1145/2501988.2501998>
- [90] Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. 2016. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Presented in San Jose, California, USA. ACM, New York, NY, USA, 5776–5788. DOI:<http://dx.doi.org/10.1145/2858036.2858117>
- [91] Alex Poole and Linden J Ball. 2006. Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction* 1 (2006), 211–219.
- [92] Robert Ravník and Franc Solina. 2013. Audience Measurement of Digital Signage: Quantitative Study in Real-World Environment Using Computer Vision. *Interacting with Computers* 25, 3 (2013), 218–228. DOI:<http://dx.doi.org/10.1093/iwc/iws023>
- [93] Gabriel Reyes, Jason Wu, Nikita Juneja, Maxim Goldshtein, W. Keith Edwards, Gregory D. Abowd, and Thad Starner. 2018. SynchroWatch: One-Handed Synchronous Smartwatch Gestures Using Correlation and Magnetic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 158 (Jan. 2018), 26 pages. DOI:<http://dx.doi.org/10.1145/3161162>
- [94] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton, and Susanna Hertrich. 2010. Ambient Influence: Can Twinkly Lights Lure and Abstract Representations Trigger Behavioral Change?. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10)*. Presented in Copenhagen, Denmark. ACM, New York, NY, USA, 261–270. DOI:<http://dx.doi.org/10.1145/1864349.1864372>
- [95] Mondville Rosen and William Durfee. 1978. An Eye Gaze Controlled Keyboard. In *Proceedings of the 5th Conference on Systems and Devices for the Disabled, Houston*.



- [96] Constantin Schmidt, Jörg Müller, and Gilles Bailly. 2013. Screenfinity: Extending the Perception Area of Content on Very Large Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Presented in Paris, France. ACM, New York, NY, USA, 1719–1728. DOI: <http://dx.doi.org/10.1145/2470654.2466227>
- [97] Johann Schrammel, Elke Mattheiss, Susen Döbelt, Lucas Paletta, Alexander Almer, and Manfred Tscheligi. 2011. *Attentional Behavior of Users on the Move Towards Pervasive Advertising Media*. Springer London, London, 287–307. DOI: [http://dx.doi.org/10.1007/978-0-85729-352-7\\_14](http://dx.doi.org/10.1007/978-0-85729-352-7_14)
- [98] Jeffrey S. Shell, Roel Vertegaal, Daniel Cheng, Alexander W. Skaburskis, Changuk Sohn, A. James Stewart, Omar Aoudeh, and Connor Dickie. 2004. ECSGlasses and EyePliances: Using Attention to Open Sociable Windows of Interaction. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA '04)*. Presented in San Antonio, Texas. ACM, New York, NY, USA, 93–100. DOI: <http://dx.doi.org/10.1145/968363.968384>
- [99] Jeffrey S. Shell, Roel Vertegaal, and Alexander W. Skaburskis. 2003. EyePliances: Attention-seeking Devices That Respond to Visual Attention. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. Presented in Ft. Lauderdale, Florida, USA. ACM, New York, NY, USA, 770–771. DOI: <http://dx.doi.org/10.1145/765891.765981>
- [100] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. Presented in The Hague, The Netherlands. ACM, New York, NY, USA, 281–288. DOI: <http://dx.doi.org/10.1145/332040.332445>
- [101] Andreas Sippl, Clemens Holzmann, Doris Zachhuber, and Alois Ferscha. 2010. Real-Time Gaze Tracking for Public Displays. In *Ambient Intelligence*, Boris de Ruyter, Reiner Wichert, David V. Keyson, Panos Markopoulos, Norbert Streitz, Monica Divitini, Nikolaos Georgantas, and Antonio Mana Gomez (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 167–176. DOI: [http://dx.doi.org/10.1007/978-3-642-16917-5\\_17](http://dx.doi.org/10.1007/978-3-642-16917-5_17)
- [102] Brian A. Smith, Qi Yin, Steven K. Feiner, and Shree K. Nayar. 2013. Gaze Locking: Passive Eye Contact Detection for Human-object Interaction. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. Presented in St. Andrews, Scotland, United Kingdom. ACM, New York, NY, USA, 271–280. DOI: <http://dx.doi.org/10.1145/2501988.2501994>
- [103] Fabius Steinberger, Marcus Foth, and Florian Alt. 2014. Vote With Your Feet: Local Community Polling on Urban Screens. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. Presented in Copen-

- 
- hagen, Denmark. ACM, New York, NY, USA, Article 44, 6 pages. DOI:<http://dx.doi.org/10.1145/2611009.2611015>
- [104] Sophie Stellmach and Raimund Dachzelt. 2012. Look & Touch: Gaze-supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Presented in Austin, Texas, USA. ACM, New York, NY, USA, 2981–2990. DOI:<http://dx.doi.org/10.1145/2207676.2208709>
- [105] Sophie Stellmach and Raimund Dachzelt. 2013. Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Presented in Paris, France. ACM, New York, NY, USA, 285–294. DOI:<http://dx.doi.org/10.1145/2470654.2470695>
- [106] Yusuke Sugano, Xucong Zhang, and Andreas Bulling. 2016. AggreGaze: Collective Estimation of Audience Attention on Public Displays. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Presented in Tokyo, Japan. ACM, New York, NY, USA, 821–831. DOI: <http://dx.doi.org/10.1145/2984511.2984536>
- [107] Nick Taylor, Justin Marshall, Alicia Blum-Ross, John Mills, Jon Rogers, Paul Egglestone, David M. Frohlich, Peter Wright, and Patrick Olivier. 2012. Viewpoint: Empowering Communities with Situated Voting Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Presented in Austin, Texas, USA. ACM, New York, NY, USA, 1361–1370. DOI: <http://dx.doi.org/10.1145/2207676.2208594>
- [108] Peter Trillenber, Rebekka Lencer, and Wolfgang Heide. 2004. Eye movements and psychiatric disease. *Curr. Opin. Neurol.* 17, 1 (Feb 2004), 43–47.
- [109] Beverley Ann Tudor. 1986. Retail Trade Advertising in the 'Leicester Journal' and the 'Leicester Chronicle' 1855-71. *European Journal of Marketing* 20, 9 (1986), 41–56. DOI:<http://dx.doi.org/10.1108/EUM00000000004666>
- [110] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Presented in Seoul, Republic of Korea. ACM, New York, NY, USA, 4179–4188. DOI:<http://dx.doi.org/10.1145/2702123.2702355>
- [111] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. 2013a. Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet

- Wesson, and Marco Winckler (Eds.). Lecture Notes in Computer Science, Vol. 8118. Springer Berlin Heidelberg, 170–186. DOI:[http://dx.doi.org/10.1007/978-3-642-40480-1\\_11](http://dx.doi.org/10.1007/978-3-642-40480-1_11)
- [112] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2013b. Eye Drop: An Interaction Concept for Gaze-supported Point-to-point Content Transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. Presented in Luleå, Sweden. ACM, New York, NY, USA, Article 37, 4 pages. DOI:<http://dx.doi.org/10.1145/2541831.2541868>
- [113] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-device Gaze-supported Point-to-point Content Transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. Presented in Safety Harbor, Florida. ACM, New York, NY, USA, 19–26. DOI:<http://dx.doi.org/10.1145/2578153.2578155>
- [114] Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: Sparking Civic Discourse by a Public Interactive Poll Visualization. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Presented in Baltimore, Maryland, USA. ACM, New York, NY, USA, 1323–1332. DOI:<http://dx.doi.org/10.1145/2531602.2531639>
- [115] Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. Presented in Brisbane, QLD, Australia. ACM, New York, NY, USA, 812–817. DOI:<http://dx.doi.org/10.1145/2901790.2901867>
- [116] Roel Vertegaal. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.
- [117] Roel Vertegaal, Connor Dickie, Changuk Sohn, and Myron Flickner. 2002. Designing Attentive Cell Phone Using Wearable Eyecontact Sensors. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. Presented in Minneapolis, Minnesota, USA. ACM, New York, NY, USA, 646–647. DOI:<http://dx.doi.org/10.1145/506443.506526>
- [118] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. Presented in Zurich, Switzerland. ACM, New York, NY, USA, 439–448. DOI:<http://dx.doi.org/10.1145/2493432.2493477>

- 
- [119] Vasilis Vlachokyriakos, Rob Comber, Karim Ladha, Nick Taylor, Paul Dunphy, Patrick McCorry, and Patrick Olivier. 2014. PosterVote: Expanding the Action Repertoire for Local Political Activism. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. Presented in Vancouver, BC, Canada. ACM, New York, NY, USA, 795–804. DOI:<http://dx.doi.org/10.1145/2598510.2598523>
- [120] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. Presented in Santa Fe, NM, USA. ACM, New York, NY, USA, 137–146. DOI:<http://dx.doi.org/10.1145/1029632.1029656>
- [121] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Presented in Paris, France. ACM, New York, NY, USA, 841–850. DOI:<http://dx.doi.org/10.1145/2470654.2470774>
- [122] Miaosen Wang, Sebastian Boring, and Saul Greenberg. 2012. Proxemic Peddler: A Public Advertising Display That Captures and Preserves the Attention of a Passerby. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*. Presented in Porto, Portugal. ACM, New York, NY, USA, Article 3, 6 pages. DOI:<http://dx.doi.org/10.1145/2307798.2307801>
- [123] Colin Ware and Harutune H. Mikaelian. 1987. An Evaluation of an Eye Tracker As a Device for Computer Input2. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface (CHI '87)*. Presented in Toronto, Ontario, Canada. ACM, New York, NY, USA, 183–188. DOI:<http://dx.doi.org/10.1145/29933.275627>
- [124] Mark Weiser. 1991. The Computer for the 21 st Century. *Scientific American* 265, 3 (1991), 94–105. <http://www.jstor.org/stable/24938718>
- [125] Erroll Wood, Tadas Baltrušaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling. 2016. Learning an Appearance-based Gaze Estimator from One Million Synthesised Images. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. Presented in Charleston, South Carolina. ACM, New York, NY, USA, 131–138. DOI:<http://dx.doi.org/10.1145/2857491.2857492>
- [126] Meredith Ringel Morris Xiaoyi Zhang, Harish S. Kulkarni. 2017. Smartphone-Based Gaze Gesture Communication for People with Motor Disabilities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI*

- '17). Presented in Denver, Colorado, USA. ACM, New York, NY, USA. DOI: <http://dx.doi.org/10.1145/3025453.3025790>
- [127] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. 2017. Everyday Eye Contact Detection Using Unsupervised Gaze Target Discovery. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. Presented in Quebec City, QC, Canada. ACM, New York, NY, USA, 193–203. DOI:<http://dx.doi.org/10.1145/3126594.3126614>
- [128] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Presented in Paris, France. ACM, New York, NY, USA, 851–860. DOI: <http://dx.doi.org/10.1145/2470654.2470775>
- [129] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2014. Pupil-canthi-ratio: a calibration-free method for tracking horizontal gaze direction. In *Proc. of the 2014 International Working Conference on Advanced Visual Interfaces (AVI 14)* (2014-05-27). ACM, New York, NY, USA, 129–132. <http://dx.doi.org/10.1145/2598153.2598186>
- [130] Yanxia Zhang, MingKi Chong, Jörg Müller, Andreas Bulling, and Hans Gellersen. 2015. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing* 19, 5-6 (2015), 967–981. DOI:<http://dx.doi.org/10.1007/s00779-015-0866-8>
- [131] Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. Presented in Seattle, Washington. ACM, New York, NY, USA, 559–563. DOI:<http://dx.doi.org/10.1145/2632048.2636071>



## **CURRICULUM VITAE**



---

# Mohamed Khamis CV

Address:

Date of Birth: September 4<sup>th</sup>, 1989  
Nationality: Egyptian

📞 Mobile (DE):

✉ Email: [me@mkhamis.com](mailto:me@mkhamis.com)  
[Mohamed.Khamis@ifi.lmu.de](mailto:Mohamed.Khamis@ifi.lmu.de)  
🌐 Website: <http://www.mkhamis.com/>

## Education

08/2018 PhD in Media informatics, Ludwig Maximilian University of Munich (LMU).  
(expected) Dissertation title: Designing Gaze-based Interaction for Public Displays  
First Supervisor: Florian Alt  
Second Supervisor: Andreas Bulling  
External Examiner: Marc Langheinrich  
02/2012 M.Sc. in Computer Science & Engineering, German University in Cairo.  
07/2011 B.Sc. **with Honors** in Media Engineering & Technology, German University in Cairo.

## Personal profile

I am a final year PhD student at Ludwig Maximilian University of Munich (LMU) expected to graduate in August 2018. I have worked at **four academic institutions** throughout my career, in which I participated in **teaching** and **research**, as well as **administrative** tasks (e.g., ERASMUS coordinator).

I have significant research contributions in Human-Computer Interaction (HCI). I **published 38 peer reviewed papers** (24 since 2017), 5 of which at CHI—the highest ranked conference in HCI. I received 3 **Honorable Mention** awards (top 5% of submissions) at CHI'18, CHI'17 and MUM'17. I also published at other top HCI venues such as UIST, UbiComp, IMWUT, ICMI, and ETRA. Many of these accomplishments resulted from **collaborations** with scholars in Germany, UK, USA, Japan, Finland, Italy, and Egypt.

I worked in a large **diversity of topics** within HCI, privacy and ubiquitous computing. In particular, I am active in the fields of eye tracking, pervasive displays, usable security, and user privacy. I am mainly known for my major contributions that are at the crossroads of user privacy and ubiquitous computing:

- **Designing Gaze-based Systems:** I contributed to comprehensively understanding and addressing challenges of gaze interaction on ubiquitous devices, such as mobile devices (CHI'18) and public displays (UIST'17, UbiComp'16, MUM'16), as well as proposing novel concepts that employ gaze to address problems on public displays (CHI'18, IMWUT, PerDis'17), and mobile devices (ICMI'17).
- **Usable Security and Privacy:** I contributed to both: (1) understanding threats to user privacy that are caused/facilitated by ubiquitous technologies, such as thermal attacks (CHI'17, Honorable Mention), and shoulder surfing (CHI'17), and (2) inventing novel ubiquitous systems for protecting user privacy on mobile devices (ICMI'17), public displays (PerDis'17), and in VR (USEC'17).

I am actively involved in the relevant research communities. I am an **associate chair** for CHI LBW'18, **demo chair** for MUM'18, member of the **program committees** of MUM and PerDis, and **organizer** of PETMEI'16 and ArabHCI'18. I **reviewed 50+ papers** for esteemed conferences, and received an **Excellent Review** recognition at CHI'17. I also reviewed for **top journals** like ACM Computing Surveys (CSUR), IMWUT, and **ACM Transactions on Privacy and Security (TOPS)**.

I taught introductory and advanced courses in HCI, computer science, and security. I significantly contributed to the development of courses such as **“Introduction to Usable Privacy and Security”** at the LMU. I **supervised 25+ bachelor and master theses**. I also helped several students kick off their PhDs.



## Employment

- Since 10/2014 Research Associate at Ludwig Maximilian University of Munich (LMU) in Munich, Germany  
Duties include (1) research in HCI, particularly in gaze-based interaction and pervasive displays, (2) teaching courses, supervising seminars, and supervising thesis projects, (3) Erasmus coordination, and (4) internal organization tasks such as internal doctoral colloquiums and open lab days.
- 04/2014 - 10/2014 PhD Student at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern, Germany
- 03/2012 - 07/2013 Teaching and Research Assistant at the German University in Cairo, Egypt
- 09/2011 - 02/2012 Student research assistant at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern, Germany
- 07/2011 - 08/2011 Interaction Designer at G-Beehive in Cairo, Egypt
- 07/2010 - 08/2010 Software development intern at Etisalat in Cairo, Egypt
- 03/2010 - 05/2010 Student research assistant at University of Würzburg in Würzburg, Germany
- 08/2009 - 09/2009 Software development intern at Etisalat in Cairo, Egypt
- 09/2007 - 06/2011 Student teaching assistant at the German University in Cairo, Egypt

## Publications List

I published a total of **38 peer reviewed papers**. A full list of publications as well as **recordings of my talks** can be found on my personal website <http://www.mkhamis.com/publications.php>

### Journal papers (3)

38. **Mohamed Khamis**, Daniel Buschek, Tobias Thieron, Florian Alt, Andreas Bulling. EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **IMWUT** *1*, *4*, (December 2017). DOI: <https://doi.org/10.1145/3161168>
37. **Mohamed Khamis**, J. Henderson, Guiying Du. PerDis 2017. *IEEE Pervasive Computing* (Volume: 16, Issue: 4, October-December 2017). DOI: <https://doi.org/10.1109/MPRV.2017.3971126>
36. Dominic Gorecky, **Mohamed Khamis**, Katharina Mura. Introduction and establishment of virtual training in the factory of the future. *International Journal of Computer Integrated Manufacturing.* *30*, *1*, (May 2015), 182 – 190. DOI: <http://dx.doi.org/10.1080/0951192X.2015.1067918>

### Conference papers (20)

2018

35. **Mohamed Khamis**, Florian Alt, Andreas Bulling. The Past, Present, and Future of Gaze-enabled Handheld Mobile Devices: Survey and Lessons Learned. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile-HCI 2018)*. ACM, New York, NY, USA.
34. Thomas Mattusch, Mahsa Mirzamohammad, **Mohamed Khamis**, Andreas Bulling, Florian Alt. Hidden Pursuits: Evaluating Gaze-selection via Pursuits when the Stimulus Trajectory is Partially Hidden. In *Proceedings of the ACM International Symposium on Eye Tracking Research & Applications (ETRA 2018)*. ACM, New York, NY, USA.

33. Christian Mai, **Mohamed Khamis**. Public HMDs: Modeling and Understanding User Behavior around Public Head-Mounted Displays. In *Proceedings of the 7th International Symposium on Pervasive Displays (PerDis 2018)*. ACM, New York, NY, USA.
  32. **Mohamed Khamis**, Carl Oechsner, Florian Alt, Andreas Bulling. 2018. VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '18)*. ACM, New York, NY, USA.
  31. **Mohamed Khamis**, Christian Becker, Andreas Bulling, Florian Alt. 2018. Which one is me? Identifying Oneself on Public Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA. **HONORABLE MENTION AWARD**
  30. **Mohamed Khamis**, Anita Baier, Niels Henze, Florian Alt, Andreas Bulling. Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones used in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA.
  29. Ville Mäkelä, **Mohamed Khamis**, Lukas Mecke, Jobin James, Markku Turunen, Florian Alt. Pocket Transfers: Interaction Techniques for Transferring Content from Situated Displays to Mobile Devices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA.
- 2017
28. **Mohamed Khamis**, Linda Bandelow, Stina Schick, Dario Casadevall, Andreas Bulling and Florian Alt. 2017. They are all after you: Investigating the Viability of a Threat Model that involves Multiple Shoulder Surfers. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3152832.3152851> **HONORABLE MENTION AWARD**
  27. Mariam Hassib, **Mohamed Khamis**, Susanne Friedl, Stefan Schneegass and Florian Alt. 2017. BrainAtWork: Logging Cognitive Engagement and Tasks in the Workplace using Electroencephalography. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3152832.3152865>
  26. **Mohamed Khamis**, Mariam Hassib, Emanuel von Zezschwitz, Andreas Bulling, Florian Alt. 2017. GazeTouchPIN: Protecting Sensitive Data on Mobile Devices using Secure Multimodal Authentication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. ACM, New York, NY, USA. DOI: <https://doi.org/10.1145/3136755.3136809>
  25. **Mohamed Khamis**, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, Andreas Bulling. 2017. EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *Proceedings of the 30th Annual Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, USA. DOI: <https://doi.org/10.1145/3126594.3126630>
  24. **Mohamed Khamis**, Regina Hasholzner, Andreas Bulling, and Florian Alt. 2017. GTmoPass: two-factor authentication on public displays using gaze-touch passwords and personal mobile devices. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*. ACM, New York, NY, USA, Article 8. DOI: <https://doi.org/10.1145/3078810.3078815>
  23. Vito Gentile, **Mohamed Khamis**, Salvatore Sorce, and Florian Alt. 2017. They are looking at me!: understanding how audience presence impacts on public display users. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*. ACM, New York, NY, USA, Article 11. DOI: <https://doi.org/10.1145/3078810.3078822>
  22. Yomna Abdelrahman, **Mohamed Khamis**, Stefan Schneegass, and Florian Alt. 2017. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3751-3763. DOI: <https://doi.org/10.1145/3025453.3025461> **HONORABLE MENTION AWARD**

21. Malin Eiband, **Mohamed Khamis**, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4254-4265. DOI: <https://doi.org/10.1145/3025453.3025636>
20. Ceenu George, **Mohamed Khamis**, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt Florian Alt, Heinrich Hussmann. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *USEC conference in conjunction with (NDSS '17)*.

2016

19. **Mohamed Khamis**, Ludwig Trotter, Markus Tessmann, Christina Dannhart, Andreas Bulling, and Florian Alt. 2016. EyeVote in the wild: do users bother correcting system errors on public displays?. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 57-62. DOI: <https://doi.org/10.1145/3012709.3012743>
18. **Mohamed Khamis**, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: using text for pursuits-based interaction and calibration on public displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 274-285. DOI: <https://doi.org/10.1145/2971648.2971679>

2015

17. Omar Mowafi, **Mohamed Khamis** and Wael Abouelsaadat. AirDisplay: Experimenting with Air Flow as a Communication Medium. 2015. In *15th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT '15)*. Springer International Publishing, 316-323. DOI: [https://doi.org/10.1007/978-3-319-22701-6\\_23](https://doi.org/10.1007/978-3-319-22701-6_23)

2013

16. Jörn Hees, **Mohamed Khamis**, Ralf Biedert, Slim Abdennadher and Andreas Dengel. Collecting Links between Entities Ranked by Human Association Strengths. In *The 10<sup>th</sup> Extended Semantic Web Conference (ESWC '13)*. Springer Berlin Heidelberg, 517-531. DOI: [https://doi.org/10.1007/978-3-642-38288-8\\_35](https://doi.org/10.1007/978-3-642-38288-8_35)

#### Other Peer-reviewed papers (15)

15. Ebtisam Alabdulqader, Shaimaa Lazem, **Mohamed Khamis**, Susan Dray. Exploring Participatory Design Methods to Engage with Arab Communities. In *Proc. of the 36th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2018)*, Montréal, QC, Canada 2018.
14. Mennatallah Saleh, **Mohamed Khamis**, Christian Sturm. Privacy Invasion Experiences and Perceptions: A comparison between Germany and the Arab World. In *Proc. of the 36th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2018)*, Montréal, QC, Canada 2018.
13. Nađa Terzimehić, Mohamed Khamis, Florian Bemmman, Heinrich Hussmann. Lunchocracy: Improving Eating Dynamics in the Workplace Using a Bot-Based Anonymous Voting System. In *Proc. of the 36th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2018)*, Montréal, QC, Canada 2018.
12. Christian Mai, Lukas Rambold, **Mohamed Khamis**. TransparentHMD: Revealing the HMD User's Face to Bystanders. In *Adj. Proc. of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM 2017)*, Stuttgart, Germany. 2017.

11. Mohammed Al Sada, **Mohamed Khamis**, Akira Kato, Shigeki Sugano, Tatsuo Nakajima, Florian Alt. Challenges and Opportunities of Supernumerary Robotic Limbs. In *Proc. of the CHI 2017 Workshop on Amplification and Augmentation of Human Perception (Amplify 2017)*.
10. Oliver Duerr, **Mohamed Khamis**, Daniel Buschek, Florian Alt. HelpMe: Assisting Older Adults in Performing Tasks on Mobile Devices. In *Proc. of the CHI 2017 Workshop on Designing Mobile Interactions for the Ageing Populations*.
9. **Mohamed Khamis**, Florian Alt, Andreas Bulling. Challenges and Design Space of Gaze-enabled Public Displays. In *Adj. Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016)*, Heidelberg, Germany 2016.
8. Diako Mardanbegi, Shahram Jalaliniya, **Mohamed Khamis**, Päivi Majaranta. The 6th International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction. In *Adj. Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016)*, Heidelberg, Germany 2016.
7. **Mohamed Khamis**, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, Andreas Bulling. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proc. of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI '16)*, San Jose, CA, USA 2016.
6. Mariam Hassib, **Mohamed Khamis**, Stefan Schneegass, Alireza Sahami Shirazi, Florian Alt. Investigating User Needs for Bio-sensing and Affective Wearables. In *Proc. of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI '16)*, San Jose, CA, USA 2016.
5. **Mohamed Khamis**, Andreas Bulling, Florian Alt. TA Field Study on Spontaneous Gaze-based Interaction with a Public Display using Pursuits. In *In Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 865-874, 2015. (**UbiComp '15**), Osaka, Japan 2015.
4. **Mohamed Khamis**, Andreas Bulling, Florian Alt. Tackling Challenges of Interactive Public Displays using Gaze. In *In Adj. Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 763-766, 2015. (**UbiComp '15**), Osaka, Japan 2015.
3. Menna Bakry, **Mohamed Khamis** and Slim Abdennadher. AreCAPTCHA: Outsourcing Arabic Text Digitization to Native Speakers. In *The 11<sup>th</sup> IAPR International Workshop on Document Analysis Systems (DAS '14)*, Tours, France, 2014.
2. Sara Nassar, Nada Sharaf, **Mohamed Khamis**, Slim Abdennadher and Caroline Sabty. Collecting Arabic Dialect Variations using Games With A Purpose: A Case Study Targeting the Egyptian Dialect. In *The 2<sup>nd</sup> Workshop on Games and NLP (GAMNLP '13)*, Istanbul, Turkey, 2013.
1. Christian Schneider, **Mohamed Khamis**, Dietmar Seipel: Prolog Server Faces - A Declarative Framework for Dynamic Web Pages. In *The 24<sup>th</sup> Workshop on Constraint Logic Programming (WLP'10)*, Cairo, Egypt, 2010.

## Relevant Honors & Awards

Honorable mention best paper award (top 5% submissions) at CHI 2018,	February 2018
Honorable mention best paper award (top 5% submissions) at MUM 2017,	November 2017
Honorable mention best paper award (top 5% submissions) at CHI 2017,	May 2017
Excellent Review recognition at CHI 2017 and CHI 2018	
Google IoT Technology Research Award	March 2016

LMU travel grant for active participation in scientific conferences	July 2015
GERLS scholarship for PhD studies offered by the DAAD and the MHESR	April 2013
Excellent with Honors <i>Bachelor of Science</i>	July 2011
Outstanding Performance in Software Engineering <i>Academic Award</i>	June 2009
Software Engineering Cup <i>Academic Award</i>	June 2009
5-year Study Scholarship at the German University in Cairo	September 2006

## Invited Talks

- 11/2016 Guest Lecture at the "Wave, Touch, Look! - Novel interaction modalities for large public displays" Workshop – University of Palermo, Italy.
- 09/2016 Guest Talk at the "Current state and future of gaze interaction" Symposium – Technical University of Berlin, Germany.
- 01/2016 Guest Lecture at the Ubiquitous Interactive Systems workshop – University of Stuttgart, Germany.
- 12/2015 Invited Talk at the IoT research colloquium – German University in Cairo, Egypt.

## Professional Services

Erasmus program coordinator at Ludwig Maximilian University of Munich (LMU).

Organizing Committee member:

**Proceedings Chair:** PerDis 2017

**Workshop Organizer:** PETMEI 2016

**Demos chair:** MUM 2018

Program Committee member:

CHI LBW 2018

PerDis 2017, 2018

MUM 2016, 2017

PETMEI 2016

ET4S 2018

Mensch und Computer 2015, 2017

AutoUI 2015

Journal Reviewing:

IMWUT: 2017

ACM Transactions on Privacy and Security (TOPS): 2017

International Journal of Geographical Information: 2017

ACM Computing Surveys: 2016

International Journal of Computer Integrated Manufacturing: 2015

Conference Reviewing:

CHI (Papers): 2016, 2017, 2018  
 CHI (Late Breaking Work): 2015, 2016, 2017  
 CHI (alt.chi): 2016, 2017  
 UIST (Papers): 2017  
 HRI (Papers): 2017, 2018  
 ACM MM (Papers): 2017  
 CHI PLAY (Papers): 2017  
 ETRA (Papers): 2018  
 ICMI (Papers): 2017  
 DIS (Papers): 2016, 2017  
 MobileHCI (Papers): 2016, 2017  
 UbiComp (Papers): 2016  
 ISWC (Papers): 2016  
 NordiCHI (Papers): 2016

## Teaching Experience and Departmental Talks

Supervised 25+ bachelor and master theses.

Guest lecture at “Introduction to Usable Security” course at the LMU about **Research Methods in HCI**.

Guest lecture at the “Human Computer Interaction 2” course at the LMU about **Eye Tracking**.

Guest lecture at the “Advanced Topics in HCI” course at the LMU about **Eye Tracking**.

Guest lecture at the “Open Games Workshop” at the LMU about **Eye Tracking**.

Taught (in English) introductory and advanced courses in HCI, cryptography, and computer science at Ludwig Maximilian University of Munich (LMU), and the German University in Cairo (GUC), . These included:

1. Usable Privacy and Security (WS17/18, SS18) ~30 students.
2. Multimedia Programming (SS18).
3. Workshop on Persuasive Technologies (SS17) ~15 students.
4. Advanced Topics in HCI (SS15, SS16, SS17, SS18) ~30 students.
5. Advanced Seminar in Media Informatics (WS15/16, SS16—WS17/18, SS18) ~20 students.
6. Web Technologies and Usability (SS13, SS12) 2 groups × ~25 students
7. Computer and Network Security (SS13) 5 groups × ~25 students
8. Introduction to Media Engineering (WS12/13) 3 groups × ~25 students
9. Introduction to Networks (SS12) 3 groups × ~25 students
10. Information Security (WS12/13) 2 groups × ~25 students

Significantly contributed to the development of several courses at the LMU and the GUC, including “An introduction to Usable Security”, “Computer and Network Security”, and “Information Security”.

I am interested in teaching HCI, computer science, and security courses, as well as introducing new modules (e.g., Usable Security) targeted at interdisciplinary students.

## Languages

Arabic	<i>Native proficiency</i>
English	<i>Full professional proficiency</i>
German	<i>Intermediate working proficiency (Level C1)</i>
French	<i>Elementary proficiency</i>

## Referees

### **Florian Alt**

Professor and head of the Ubiquitous Interactive Systems Group at Ludwig Maximilian University of Munich (LMU).

✉ Email: [florian.alt@ifi.lmu.de](mailto:florian.alt@ifi.lmu.de)

✉ Address: Universität München, Frauenlobstr. 7a, 80337, Munich, Germany.

### **Andreas Bulling**

Head of the Perceptual User Interfaces Group at the Max Planck Institute for Informatics and the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University, Germany

✉ Email: [bulling@mpi-inf.mpg.de](mailto:bulling@mpi-inf.mpg.de)

✉ Address: Max Planck Institute for Informatics, Campus E1 4, 66123, Saarbrücken, Germany.

### **Albrecht Schmidt**

Professor and head of the Human Centered Ubiquitous Media Group at Ludwig Maximilian University of Munich.

✉ Email: [albrecht.schmidt@ifi.lmu.de](mailto:albrecht.schmidt@ifi.lmu.de)

✉ Address: Universität München, Frauenlobstr. 7a, 80337, Munich, Germany.



## Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

14<sup>th</sup> of May, 2018

Mohamed Khamis