

Deepak Akkil

Gaze Awareness in Computer-Mediated Collaborative Physical Tasks

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Information Technology and Communication Sciences of the Tampere University, for public discussion in the Pinni A

Paavo Koli Auditorium
on August 30th, 2019, at noon.

Faculty of Information Technology and Communication Sciences
Tampere University

Dissertations in Interactive Technology, Number 31
Tampere 2019

ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY

Responsible Supervisor and Custos: Docent Poika Isokoski
Faculty of Information Technology and Communication Sciences,
Tampere University, Finland

Opponent: Associate Prof. Roman Bednarik
School of Computing,
University of Eastern Finland, Finland

Reviewers: Prof. Susan Fussell
Departments of Communication and Information Science,
Cornell University, USA

Prof. Sebastian Pannasch
Institute of Psychology, Engineering Psychology and Applied
Cognitive Research,
Technische Universität Dresden, Germany

ISBN 978-952-03-1178-0 (pdf)
<http://urn.fi/URN:ISBN:978-952-03-1178-0>

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of Tampere University.

Dissertations in Interactive Technology, Number 31

Faculty of Information Technology and Communication Sciences
FIN-33014 Tampere University
FINLAND

ISBN 978-952-03-1177-3 (print)
ISSN 1795-9489

Juvenes Print – Suomen Yliopistopaino Oy
Tampere 2019

Abstract

Human eyes play an important role in everyday social interactions. However, the cues provided by eye movements are often missing or difficult to interpret in computer-mediated remote collaboration. Motivated by the increasing availability of gaze-tracking devices in the consumer market and the growing need for improved remote-collaboration systems, this thesis evaluated the value of gaze awareness in a number of video-based remote-collaboration situations.

This thesis comprises six publications which enhance our understanding of the everyday use of gaze-tracking technology and the value of shared gaze to remote collaborations in the physical world. The studies focused on a variety of collaborative scenarios involving different camera configurations (*stationary, handheld, and head-mounted cameras*), display setups (*screen-based and projection displays*), mobility requirements (*stationary and mobile tasks*), and task characteristics (*pointing and procedural tasks*). The aim was to understand the costs and benefits of shared gaze in video-based collaborative physical tasks.

The findings suggest that gaze awareness is useful in remote collaboration for physical tasks. Shared gaze enables efficient communication of spatial information, helps viewers to predict task-relevant intentions, and enables improved situational awareness. However, different contextual factors can influence the utility of shared gaze. Shared gaze was more useful when the collaborative task involved communicating pointing information instead of procedural information, the collaborators were mutually aware of the shared gaze, and the quality of gaze-tracking was accurate enough to meet the task requirements. In addition, the results suggest that the collaborators' roles can also affect the perceived utility of shared gaze.

Methodologically, this thesis sets a precedent in shared gaze research by reporting the objective gaze data quality achieved in the studies and also provides tools for other researchers to objectively view gaze data quality in different research phases.

The findings of this thesis can contribute towards designing future remote-collaboration systems; towards the vision of pervasive gaze-based interaction; and towards improved validity, repeatability, and comparability of research involving gaze trackers.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Dr. Poika Isokoski. Poika has played an instrumental role not just in this thesis but also in shaping the researcher I am today. He has always been available when I needed his time, mentored me on the right path, encouraged me to stay focused when I sometimes diverged, and motivated me when I occasionally felt lost in the publication process. This thesis would not have been possible without his guidance.

I take this opportunity to thank the pre-examiners Prof. Sebastian Pannasch and Prof. Susan Fussell for their insightful comments. Thank you, Dr. Roman Bednarik for agreeing to be the opponent during the public examination of the dissertation.

My early experience in the Haptic and Gaze Interaction (HAGI) project encouraged me to pursue a Ph.D. I owe a great deal to the vastly knowledgeable colleagues I had the privilege to work with in the HAGI project. Dr. Jari Kangas, Dr. Päivi Majaranta, Dr. Jussi Rantala, and Dr. Oleg Špakov, I have learnt so much about gaze tracking, haptics, and research methods by interacting with you. I also thank my colleagues in the Visual Interaction Research Group (VIRG) for their support, encouragement, and guidance and my co-authors from Nokia Labs for their contribution to the related publication. Thank you very much.

The work presented in this dissertation was conducted in the facilities of Tampere Unit for Computer-Human Interaction (TAUCHI). I would like to express my gratitude to Prof. Roope Raisamo, director of TAUCHI, for developing and promoting such a friendly and constructive research environment. It has been a privilege to be part of TAUCHI.

I was fortunate to have received a funded doctoral position in the School of Information Sciences (SIS) at the beginning of my thesis and in the School of Communication Sciences (COMS) towards the end. This thesis would not have seen the light of day without the support from these schools.

Mom, Dad, and Brother, thank you for encouraging me in my educational pursuits. My dear Oona, thank you for your understanding during this time. Biju, Jobin, and Bobin, your support have been invaluable.

Deepak Akkil

Contents

1	INTRODUCTION	1
1.1	Research Context and Focus	2
1.2	Method	6
1.3	Contribution	7
1.4	Structure.....	7
2	COMMUNICATIVE FUNCTIONS OF HUMAN EYES.....	9
2.1	Human Eyes and Eye Movements	9
2.2	Types of Eye Movements.....	10
2.3	Perception of Gaze Direction	12
2.4	Role of the Eyes in Social Interactions	14
2.5	Role of Eyes in Collocated Collaborative Physical Tasks	15
3	GAZE TRACKING AND GAZE-BASED HUMAN COMPUTER INTERACTION.....	19
3.1	Basic Concepts of Gaze Tracking.....	19
3.2	Gaze-Based Human Computer Interaction.....	21
3.3	Gaze Tracking Data Quality	23
4	GAZE SHARING IN COMPUTER-MEDIATED COMMUNICATION	27
4.1	Introduction to the Literature Review	27
4.2	Methodology	30
4.3	Classification Based on Johansen’s Time-Space Matrix.....	31
5	SHARED GAZE IN REAL-TIME REMOTE COLLABORATION	45
5.1	Classification of Previous Studies on Shared Gaze.....	45
5.2	Task Coupling	57
5.3	Benefits of Shared Gaze in Remote Collaboration	58
5.4	Limitations of Shared Gaze in Remote Collaboration	60
5.5	How Have Previous Studies Addressed Gaze-Data Quality?	64
6	METHODOLOGY	67
6.1	Constructive and Iterative Approach	67
6.2	Early Pilot Evaluations.....	68
6.3	Research Ethics.....	73
7	INTRODUCTION TO THE STUDIES	78
7.1	Study I: Measuring and Reporting Gaze-Tracking Quality.....	78
7.2	Study II: User Expectations of Everyday Gaze Interaction.....	80
7.3	Study III: Gaze Augmentation and Awareness of Intention	82
7.4	Study IV: Shared Gaze for Spatial Referencing.....	83
7.5	Study V: Shared Gaze for Stationary Collaborative Physical Tasks.....	84
7.6	Study VI: Shared Gaze for Mobile Collaborative Physical Tasks	86
8	DISCUSSION.....	89
9	CONCLUSION	103
10	REFERENCES	107

List of Figures

Figure 1. The three themes of the thesis	2
Figure 2. Conceptual illustration of the difference in acuity of foveal and peripheral vision in humans.....	10
Figure 3. Different types of VOG gaze trackers	21
Figure 4. Visualisation of gaze data in terms of spatial quality.....	24
Figure 5. Distribution of publications included in the literature review.....	31
Figure 6. Factors used in the classification of the literature on shared gaze in real-time remote collaboration	45
Figure 7. Shared display collaboration setup involving two users.....	46
Figure 8. Example gaze visualisations used in previous studies.....	51
Figure 9. Three gaze-sharing configurations based on the level of awareness of the gaze producer	55
Figure 10. Different real-world collaborative use cases of gaze-augmented egocentric video explored as part of the thesis	69
Figure 11. The university cafeteria environment where the pilot evaluations were conducted.....	71
Figure 12. Publications and the research themes	77
Figure 13. An example visualisation presented in TraQuMe showing good and bad gaze data quality	79

List of Publications

This thesis consists of a summary and the following six original publications, reproduced here by permission.

- I. **Deepak Akkil**, Poika Isokoski, Jari Kangas, Jussi Rantala, and Roope Raisamo. 2014. TraQuMe: A Tool for Measuring the Gaze Tracking Quality. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 327-330. DOI: <https://doi.org/10.1145/2578153.2578192> 127
- II. **Deepak Akkil**, Andrés Lucero, Jari Kangas, Tero Jokela, Marja Salmimaa, and Roope Raisamo. 2016. User Expectations of Everyday Gaze Interaction on Smartglasses. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 24, 10 pages. DOI: <https://doi.org/10.1145/2971485.2971496> 133
- III. **Deepak Akkil** and Poika Isokoski. 2016. Gaze Augmentation in Egocentric Video Improves Awareness of Intention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1573-1584. DOI: <https://doi.org/10.1145/2858036.2858127> 145
- IV. **Deepak Akkil** and Poika Isokoski. 2016. Accuracy of Interpreting Pointing Gestures in Egocentric View. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 262-273. DOI: <https://doi.org/10.1145/2971648.2971687> 159
- V. **Deepak Akkil** and Poika Isokoski. 2019. Comparison of Gaze and Mouse Pointers for Video-Based Collaborative Physical Task. In *Interacting with Computers*. Article iwc026, 19 pages. DOI: <https://dx.doi.org/10.1093/iwc/iwy026> 173
- VI. **Deepak Akkil**, Biju Thankachan, and Poika Isokoski. 2018. I See What You See: Gaze Awareness in Mobile Video Collaboration. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 32, 9 pages. DOI: <https://doi.org/10.1145/3204493.3204542> 195

Author's Contribution to the Publications

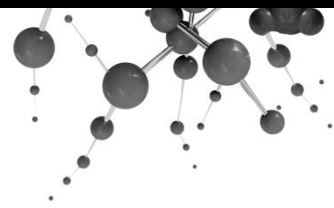
All six publications included in this thesis were co-authored and involved collaborative effort from all of the co-authors.

Dr. Poika Isokoski initiated the basic idea for developing a gaze data quality-measurement tool (Study I). The author of this thesis put forward the idea of gaze tracker calibration based on mouse clicks as well as the design of the calibration mechanism, which was used in the experiment to showcase the gaze tracker quality-measurement tool. Furthermore, the author of this thesis developed all of the required software solutions, conducted the user study, analysed the results, and was the main author of the manuscript.

Study II originated from a joint project with Nokia Labs, Finland. The research involved collaboration between the publication's co-authors at every stage, starting from the conceptualisation of the study. I was the main contributor to the design of the empirical work, the conducting of the focus group study, and the manuscript's writing. The study's analysis phase involved the author of this thesis and three other co-authors of the publication, to reduce the potential subjective bias in the qualitative data analysis.

Studies III, IV, V, and VI came to fruition after a series of pilot tests conducted by the author, to understand the value of sharing gaze information between collaboration partners. The author came up with the initial research idea and direction, which was further polished by extensive discussion with Dr. Poika Isokoski. The discussions and insights from the initial pilot study led to finalising several details of the experiment (experiment setup, task and design of the study, etc.).

For Studies III, IV, V, and VI, I developed the required hardware and software prototypes (using off-the-shelf gaze-trackers, cameras, projectors, and mobile phones). Furthermore, the author moderated the experiments, conducted the relevant descriptive and inferential data analysis, and was the lead author in the different stages of the manuscript preparation.



1 Introduction

The human eye performs dual roles in our lives; an organ for perception as well as communication. In addition to their role in visual perception, by virtue of their unique morphology, eyes communicate our current point of visual attention to an observer (Kobayashi & Kohshima, 2001). The awareness of where a person is looking, how long the person has been looking at a detail, and the temporal changes in their visual attention can communicate a wealth of information in our everyday social interactions. It is thus unsurprising that humans show a preferential bias towards attending to the eyes of other people to gather these important social cues (Jack, Scheepers, Fiset, Caldara, & Blais, 2008; Walker-Smith, Gale, & Findlay, 1977).

In our everyday lives, we infer much more than gaze direction from a person's eye movements. For instance, people naturally look at objects in an environment that they prefer or find attractive (Shimojo, Simion, Shimojo, & Scheier, 2003). Similarly, an onlooker's gaze directed at a person could be interpreted as a sign of general interest in the person, romantic attraction, intention to talk, or sometimes even as a threat. Every shift in the gaze direction encodes a meaning, defined by the cultural, social, and environmental context in which it is made. Humans are not only attuned to interpreting the direction of gaze but also the meaning it communicates.

There is growing interest in the field of Human-Computer Interaction (HCI) to use eye gaze as a means to interact with computing devices. Gaze tracking is the process of measuring the movement of the eyes or estimating a person's current focus of visual attention with the aid of

technology. Computing devices equipped with gaze trackers can use this information as an input channel in HCI.

Gaze-tracking devices traditionally have been used as an assistive technology or research tool in controlled environments. However, recent advancements in software and hardware technology have made gaze tracking cheaper, more accurate, and more ergonomic to use. The technology is increasingly seen as a viable and beneficial input technique to interact with computers (Kumar, Paepcke, & Winograd, 2007), mobile phones (Drewes, De Luca, & Schmidt, 2007), public displays (Melodie Vidal, Bulling, & Gellersen, 2013), wearables such as smartwatches (Akkil et al., 2015), and head-mounted devices (Baldauf, Fröhlich, & Hutter, 2010; Duchowski et al., 2004).

1.1 RESEARCH CONTEXT AND FOCUS

This thesis focuses on three interlinked themes (see Figure 1). In the following section, I briefly introduce the three themes.

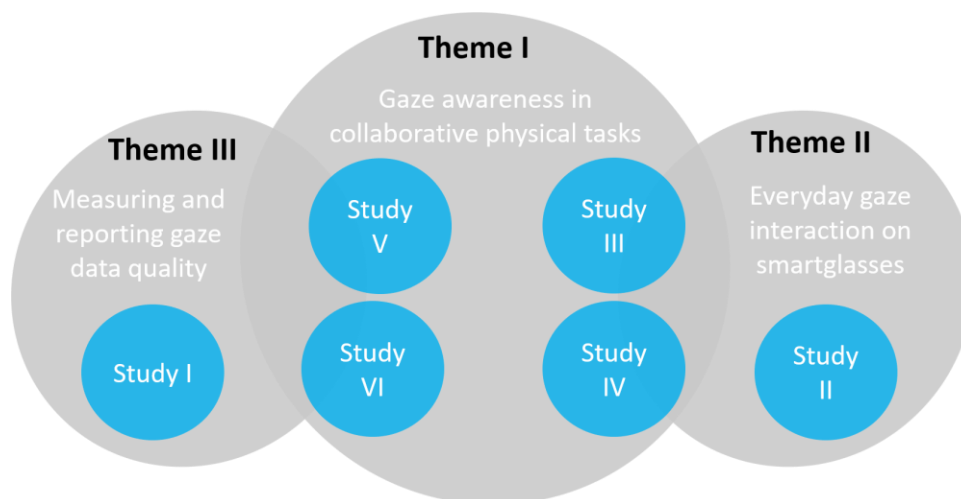


Figure 1. The three themes of the thesis

Theme I: Gaze Awareness in Collaborative Physical Tasks

The distributed nature of our current work and social networks has increased the need for technological tools and services that support collaboration between geographically separated individuals. Remote-collaboration technologies such as e-mail, instant messaging, and audio/video telephony are already integral parts of our personal and professional lives.

Within the scope of this thesis, *collaboration* is broadly defined as two or more individuals “working together with a shared goal” (Mattessich, Monsey, & Murray-Close, 2001).

Video-based collaboration technologies are particularly interesting since they provide a rich medium for communication between remote partners. A popular use of video-based collaboration is to facilitate remote meetings via video conferencing. Another growing use of video-based communication is to use *video-as-data* (Nardi, Kuchinsky, Whittaker, Leichner, & Schwarz, 1996). Here, video is not used to show “talking heads” but to provide images of the physical world to help remote participants support joint activities and experiences. Imagine young parents sharing the video of their son’s football game to the rest of the family in different locations, an industrial field worker video-calling an indoor expert to seek guidance on troubleshooting a piece of industrial equipment, an elder parent seeking the help of their child in another city to help operate a new microwave oven, a young adult video-calling a remote friend to seek suggestions while shopping, or a traveller in a new city video-calling a friend to show the interesting tourist attractions.

Current video-based remote-collaboration technologies are far from ideal, especially for tightly coupled tasks that requires frequent, complex, and real-time communication between collaborators (G. M. Olson & Olson, 2000; J. S. Olson & Olson, 2006). Such complex scenarios require cues, in addition to the shared visual information, to enable more efficient communication and improved awareness between collaborators (S. R. Fussell, Setlock, & Kraut, 2003; Gergle, Kraut, & Fussell, 2013).

Given the strong utility of eyes in everyday social interactions, previous research has explored the value of gaze awareness in video-based remote collaborations. Techniques that enable gaze awareness between remote partners involved in a discussion (Vertegaal, 1999) as well as shared display collaboration such as remote pair programming (D’Angelo & Begel, 2017) have been investigated. However, we still have limited knowledge regarding the costs and benefits of sharing gaze information between remote partners involved in collaborative physical tasks. The primary focus of this thesis is to explore **the costs and benefits of gaze awareness in video-based remote collaboration for physical tasks and to understand the different contextual factors that influence the usefulness of shared gaze.**

I used the ISO 9241-11:2018 (ISO, n.d.) definition of *context*. Within the scope of this thesis, *context* is defined as the combination of users, goals and tasks, resources, and environments (technical, physical, and social) within which a collaborative activity takes place.

Theme II: Everyday Gaze Interactions on Smartglasses

The second theme of the thesis focuses on everyday gaze-based interaction in augmented-reality (AR) smartglasses. Smartglasses are a device form factor that research community, technology enthusiasts and device manufacturers envision will revolutionise how we interact with our

environment and embedded computing devices. From the perspective of the primary theme of this thesis, smartglasses are also a relevant device form factor for remote collaborative physical tasks since they are often equipped with a world-facing camera that provides a first-person view of the world and is also hands free to use.

Gaze tracking is considered to be a viable and potentially beneficially input modality in such devices (Bulling & Gellersen, 2010). Numerous previous publications exist on leveraging gaze tracking in smartglasses (e.g. Baldauf et al. 2010; Lee et al. 2011), but all of them have focused on either core technology development or evaluating gaze as an interaction mechanism in specific use cases. Thus, we lack a holistic understanding of users' concerns and preferences when using gaze as an interaction technique in such devices. This is especially important since earlier versions of consumer smartglasses (e.g. Google Glass) faced severe social acceptability issues.

The second focus of this thesis is to enable a holistic understanding of potential users' expectations of everyday use of gaze tracking in smartglasses.

Theme III: Measuring and Reporting Gaze Data Quality

The third focus of my work is relevant to the broader gaze-tracking research community. Gaze tracking is used as a research tool in a variety of fields such as psychology, human behavioural science, marketing research, education, sports, and performance research. As the applicability of gaze tracking in research is increasing, a critical aspect that researchers often overlook is the quality of the gaze-tracking data achieved in the study and how it influences the research findings. Research that makes offline use of gaze data for analysis can perform post-calibration to reduce the influence of tracking errors. However, this is not possible in research fields that use real-time use of gaze tracking data (e.g. HCI).

Gaze data quality is critical to the validity, repeatability, and comparability of research findings. For example, in the field of HCI, two recent studies—Qian and Teather (2017) and Blattgerste et al. (2018)—independently compared gaze and head-based pointing in virtual reality. Despite the comparable research contexts, the two publications arrived at opposite results. Based on the authors' qualitative descriptions, the biggest differentiator between the two studies appeared to be in the achieved gaze data quality. However, since neither of the studies reported any objective measures of gaze data quality (note that Blattgerste et al. (2018) report a realistic accuracy measure based on a measurement conducted on 10 users separately from the user study), it is difficult to conclusively say how large the differences were or how much those differences may have affected the results.

One of the reasons why most researchers overlook gaze-tracking data quality is due to the lack of flexible tools to help easily measure, analyse, and report the quality metrics. Most tracker manufacturers do not include such flexible tools in their software offerings. For example, the Tobii Pro Lab (Tobii Technology, 2017) software outputs a numerical value for gaze-tracker calibration quality for individual users. However, the quality evaluation is coupled with the calibration routine and cannot be performed independently. Further, while tracker manufacturers may make such tools available in the future, they may still have different implementations; thus, the results across eye trackers from different manufacturers may not be comparable.

The third focus of this thesis is to develop an open-source, flexible, and gaze tracker independent tool to measure gaze data quality. In addition to developing and distributing the tool to the community, this thesis also presents examples on how to practically use the tool in research involving gaze trackers and how to report the quality measures in publications.

Research Questions and Objectives

This thesis takes inspiration from previous works in the cognitive and behavioural sciences on the value on gaze awareness in everyday social interactions and in collocated task-based collaborations as well as previous work in the HCI on gaze-based interaction in general and specifically gaze awareness in shared display collaboration.

The high-level focus of this thesis is to understand the costs and benefits of gaze awareness in real-world, collaborative physical tasks. At a lower level, this thesis focuses on answering the following research questions. The research questions were motivated and influenced by previous literature. They were arrived at after synthesising the previous work in the area and identifying gaps in our knowledge regarding the value of gaze awareness in remote collaboration. Each study was then designed to fill those gaps.

RQ1: *Can sharing gaze between collaborators lead to any measurable benefits in video-based collaborative physical tasks? If yes, what benefits does it provide? (Studies III, IV, and V)*

RQ2: *Do contextual factors influence the usability of shared gaze for collaboration? (Studies II, III, IV, and V)*

RQ3: *How does shared gaze compare against more explicit remote gesturing mechanisms such as shared mouse for collaborative physical tasks? (Studies IV and V)*

The tertiary theme of the work was motivated by the understanding that gaze-tracking-data quality can influence the usability of shared gaze for

remote collaboration and, more generally, the use of gaze in HCI. The tertiary theme of this thesis is not defined by a research question per se. Rather, it is motivated by a larger research objective.

To support and encourage gaze-tracking researchers to take a more objective view regarding gaze data quality. Furthermore, to facilitate research to easily record, analyse, and report the gaze data quality achieved in user studies. (Studies I, IV, V, and VI)

1.2 METHOD

The research reported in this thesis utilises both quantitative and qualitative approaches to gather a holistic understanding of the themes in this thesis.

Study II used focus groups as the study methodology. Furthermore, we analysed the unstructured qualitative data using affinity diagramming (Holtzblatt, Wendell, & Wood, 2005), an inductive/bottom-up thematic approach, to gather key insights regarding the end-user expectations of everyday gaze interaction using smartglasses.

Studies I, III, IV, V, and VI follow experimental research methods and were conducted in a controlled lab environment. The studies began with prototyping the relevant software and hardware systems for conducting the experiment. This was followed by a series of short pilot studies to define several key parameters and experimental design choices. From a quantitative viewpoint, we used both subjective (captured using questionnaire data) and objective measures. The main objective measures were the success rate of communicating information with and without gaze awareness (in Studies II and III) and collaboration performance measures, such as task completion times and the number of utterances required to complete the task (in Study V and VI). The subjective measures included users' confidence in interpreting the information from the video (in Studies III and IV) and a series of questions to evaluate the perceived quality of collaboration (in Studies V and VI).

Studies III and IV were conducted in two phases. We deconstructed a potentially collaborative scenario to understand the subtasks to which gaze awareness between partners could add value, in terms of intention prediction and spatial pointing, respectively. In contrast, Studies V and VI focused on real-world collaboration and involved real-time communication between participants in separate physical locations.

1.3 CONTRIBUTION

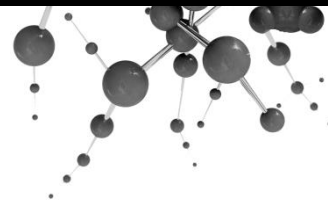
This thesis makes three contributions to gaze-based HCI. First, this thesis evaluates the value of gaze awareness in the context of video-based remote collaboration. The studies presented in this thesis will enable understanding of the contextual factors that influence the usefulness of shared gaze. More generally, the research findings will help with designing future video-based remote collaboration systems. A wider adoption of remote collaboration systems in our professional and personal lives will have many significant implications. Systems that can overcome the need for traditional collocated collaboration will help to reduce the frequent travelling needed to accomplish tasks and thus help reduce our carbon footprint. In addition, such collaboration systems will enable flexibility of work location for employees, potentially leading to improved well-being. Furthermore, they can result in time and cost savings for organisations and lead to optimised workflows.

Second, this thesis explores the users' expectations, preferences, and concerns in using everyday gaze-tracking technology and will contribute to the user-centric design of pervasive gaze-based interaction technologies.

Third, the work done in this thesis brings to the forefront the importance of gaze data quality in research involving gaze trackers. The software tool presented in this thesis will enable other researchers to record, analyse, and report the gaze data quality achieved at different research phases, thus contributing towards improving the validity, comparability, and repeatability of research involving gaze trackers.

1.4 STRUCTURE

This thesis consists of a summary of the work undertaken along with the six peer-reviewed publications. The structure of the thesis is as follows: In Chapter 2, I briefly present the communicative functions of eye movements. Chapter 3 introduces the gaze-tracking technology and its use in HCI. In particular, the chapter introduces the different factors that influence the gaze tracking data quality. In Chapters 4 and 5, I present a literature review of shared gaze interfaces for computer-mediated collaboration. The chapters also present a classification of the previous literature focusing on shared gaze. In Chapter 6, I describe the methodology used in the study and contextualise the work done as part of this thesis. Chapter 7 introduces the six publications in more detail, in terms of the methodology used and the key results. In Chapter 8, I discuss the key findings of this thesis in light of the initial research questions and objectives. The chapter also presents the limitations of this research as well as directions for future research. I conclude the thesis in Chapter 8 by highlighting the key contributions of the work.



2 Communicative Functions of Human Eyes

2.1 HUMAN EYES AND EYE MOVEMENTS

The human eye is a complicated organ that has evolved over millions of years, from a simple light-sensitive structure to the complex organ responsible for binocular vision. Our eyes work roughly like a camera, collecting incoming light through the tiny opening of the pupil and focusing it with specialised lens arrangements to the retina, the inner photosensitive layer of the eye. However, unlike a camera, which uses photographic film or digital sensors to form the image, the light striking the retina's photoreceptors causes a series of chemical and electrical events that ultimately convert the light into electrical impulses for the brain.

In addition to their working principles, what makes our eyes even more fascinating is the fact that our eyes are mobile. Eyes move both voluntarily and involuntarily to enable and enhance our sense of vision. The primitive eyes evolved to move as a means of stabilizing the image on the retina, in the presence of head movements (Walls, 1962). Gradually, the evolutionary need for higher vision resolution led to development of a specialised area in the retina, called the fovea, with relatively higher visual acuity and colour sensitivity. It also led to the complementary development of mechanisms that enable eye movements responsible for "aiming" the incoming light to this area of highest visual acuity in the retina (Walls, 1962).

The human retina is covered with two types of photoreceptors: rods and cones. These two types of photoreceptors complement each other in their

functionality and placement in the retina. Rods are useful for low-light vision but have low visual acuity and colour sensitivity. On the other hand, cones are responsible for vision in well-lit conditions. Cones have higher visual acuity and are responsible for colour vision. The fovea, the region of highest visual acuity, is almost exclusively composed of cone photoreceptors, while the rest of the retina is scattered with rods. The high concentration of cones in the fovea is the reason for its colour sensitivity and high resolution of vision. Our visual acuity degrades dramatically beyond the foveal region. The size of the fovea in humans is approximately 400 μm , which translates to 1.3 degrees of visual angle (Duchowski, 2007). To put this number into perspective, it is roughly the size of the thumbnail when held at arm's length (O'Shea, 1991). Figure 2 shows a representative image showing the difference in acuity between foveal and peripheral vision in humans.

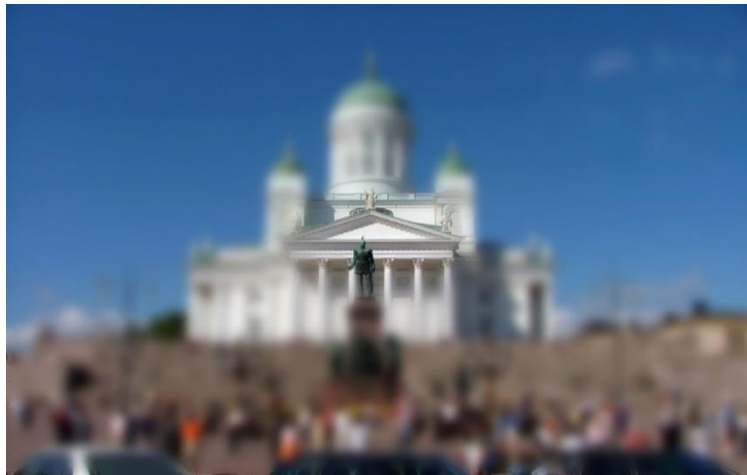


Figure 2. Conceptual illustration of the difference in acuity of foveal and peripheral vision in humans.

2.2 TYPES OF EYE MOVEMENTS

Eye movements can be classified into three categories based on functionality: gaze-shifting eye movements, gaze-stabilising eye movements, and fixational eye movements. Our eyes move to bring the region of our current interest to the foveal region for detailed inspection (gaze-shifting movements), to stabilise the image on the retina in the presence of movement (gaze-stabilising movements), or to maintain the object of interest in the foveal region and to prevent the sensory adaptation of the retina by refreshing the visual information on the retina (fixational eye movements).

There are two types of gaze-shifting eye movements: saccades and smooth pursuits.

Saccades are rapid, discrete, ballistic eye movements performed to bring an object of interest in the environment to the region of foveal vision. A saccade can last between 10 and 100 ms, during which the eyes move at a peak velocity of up to 800 degrees per second (A. T. Bahill, Clark, & Stark, 1975). Our visual perception is blinded during saccades, a phenomenon known as saccadic suppression (Matin, 1974). Fixations are the intervals between two saccades, when our eyes stay relatively stationary. During this relatively stationary period, the image on the retina is stable, and we perceive the visual information.

Smooth pursuits are smooth eye movements that enable clear vision of moving targets by visually following the target. Our environment is filled with a variety of moving stimuli, such as flying birds, vehicles in transit, sprinting animals, floating clouds, and even virtual objects moving on a computer display. Smooth pursuits are eye movements produced as a result of maintaining attention on moving targets. Pursuits are unique in the sense that one cannot produce such eye movements at will, as they require perceived motion. In comparison with saccades, which are rapid and discrete in nature, smooth pursuits are usually slower than saccades (the eye velocity depends on target velocity, typically <50 deg/sec) and are continuous in nature (C. H. Meyer, Lasker, & Robinson, 1985).

Gaze-stabilizing movements are involuntary eye movements responsible for counteracting self-motion and are required to stabilise the image of the visual world on the retina. Vestibulo-ocular reflexes (VORs) and optokinetic reflexes (OKR) are the two gaze-stabilizing eye movements in humans.

Fixational eye movements include a variety of involuntary “micro” eye movements, such as microsaccades, tremors, and drifts. Fixational eye movements play a key role in refreshing the visual information on the retina, when the eyes are relatively still. This helps to keep the objects in the visual field from perceptual fading and corrects any offsets in eye position (Martinez-Conde, Macknik, & Hubel, 2004).

Yet another type of eye movement is the vergence eye movement, characterised by simultaneous movement of the eyes in opposite directions. When looking at a nearby target, the eyes rotate inwards along the horizontal axis (i.e. convergence). Similarly, while looking at a target far away, the eyes rotate outwards, away from each other, until roughly parallel (i.e. divergence).

A detailed review of gaze-stabilising and fixational eye movements is out of the scope of this thesis. Gaze-shifting eye movements are of the most importance in terms of their communicative role and are thus also most relevant from the perspective of this thesis.

2.3 PERCEPTION OF GAZE DIRECTION

The unique morphological characteristics differentiate the human eye from the eyes of the rest of the primates. The human eye has the largest exposed sclera region, devoid of any pigmentation, surrounding the darker iris. The white sclera region and the darker iris provide a high degree of contrast, enabling an onlooker to easily perceive one's direction of gaze. The light sclera in humans is believed to be an evolutionary adaptation to enable signalling and communication using the eyes, in contrast to the gaze-camouflaging eyes found in most other primates (Kobayashi & Kohshima, 2001).

Notably, while eye movements contribute significantly to our shifts in gaze direction, we do not shift our gaze direction exclusively using eye movement. Our body position and head and eye orientation jointly modulate our gaze direction. Generally, small shifts in visual attention are almost exclusively performed using eye movement (Land, 2006). In contrast, during larger shifts in attention, eye movements are accompanied by head and body re-orientation (Land, 2006).

Head orientation is a coarse indicator of our gaze direction, while eye position combined with head orientation provides a refined interpretation of one's gaze direction. Loomis et al. (2008) note that an onlooker can accurately observe a person's head orientation through their peripheral vision, as head orientation is a large visual stimulus. In contrast, eye movements are a relatively subtle visual stimulus, and accurately interpreting them requires onlookers to fixate near the person's eyes.

Dyadic Gaze and Triadic Gaze

Our social communications contain two fundamentally different types of gaze signals: dyadic gaze and triadic gaze (George & Conty, 2008; Symons, Lee, Cedrone, & Nishimura, 2008). Dyadic gaze concerns cues provided by eye contact, while triadic gaze concerns information provided by the eyes while attending to a third party (i.e. objects or people in the environment).

Dyadic and triadic gaze show differences in their information-processing requirements, function, and underlying neurological mechanisms (Symons et al., 2008). Perceiving eye contact or dyadic gaze involves relatively simpler information processing (i.e. *are you looking at me?*). In contrast, triadic gaze requires more complex analysis (i.e. *what are you looking at?*). Similarly, one of the main functions of dyadic gaze is to regulate face-to-face social interaction, while triadic gaze has a role in regulating social interactions, revealing one's interests to the onlooker, and establishing joint attention. Developmental studies suggest that infants as young as 2 to 3 months old are sensitive to dyadic gaze (Hains & Muir, 1996), while sensitivity to triadic gaze emerges as late as 18 months (Corkum & Moore, 1998).

The perception of being looked at is a special case of a more general gazing behaviour, which seems to have unique neurological judgement mechanisms and cognitive processes associated specifically with it (George & Conty, 2008).

Previous works have investigated the acuity of perceiving dyadic and triadic gaze. In their classic study, Gibson and Pick (1963) reported that a human observer can accurately discriminate between an onlooker's gaze directed at them and one that is directed 1 cm horizontally away from them, from a distance of 200 cm. Jenkin et al. (2003) reported similar discrimination thresholds. Furthermore, the sensitivity to perceiving dyadic gaze seems to be higher along the horizontal direction than the vertical (Cline, 1967).

Similarly, humans also show remarkable sensitivity to triadic gaze. Symons et al. (2008) noted that observers can analyse another person's eye movement, derive directional information from it, and triangulate it to 3D space from the onlooker's perspective with relative ease. However, the acuity of triadic gaze changes as the target moves away from the observer and the onlooker. People are best at judging where another person is looking, when the target is between the looker and themselves, whereas acuity degrades further away (Symons et al., 2008). Bock et al. (2008) demonstrated that the overall threshold of interpreting triadic gaze varied between 1.8 degrees to 3.9 degrees of visual angle, based on the target's location. They also reported a systematic upward bias, with all of the gaze target interpretations skewed by an average of 1.2 degrees of visual angle upwards.

Generally, head orientation affects the accuracy of both dyadic and triadic gaze perception (Cline, 1967). A divergence between head and eye position introduces a constant error in gaze judgement. Furthermore, observers subconsciously integrate the information derived from the individual eyes. The information from one eye corrects the positional bias introduced from the other eye (Symons et al., 2008). This also suggests that the onlooker's relative positioning (e.g. frontal compared to sideways) can influence the accuracy of gaze perception. Overall, for most head angles and observer positions, gaze directed at the observer is discriminated with greater accuracy than other lines of regard are (Bock et al., 2008).

To summarise, humans are incredibly good at perceiving both the dyadic and triadic gaze of an onlooker. The dyadic and triadic gaze discrimination thresholds have direct implications on shared gaze in remote collaboration. In shared gaze interfaces, the gaze-tracking accuracy needs to be comparable to 1.8 degrees to 3.9 degrees to match the accuracies of gaze awareness available in naturalistic collaboration scenarios in which the collaborators are facing each other. Modern-day gaze trackers can estimate a person's point of regard at a much higher

accuracy (≤ 0.5 degrees). This improved accuracy of gaze awareness can potentially enable improved utility of shared gaze in remote collaboration, as compared to naturalistic collocated collaboration scenarios.

2.4 ROLE OF THE EYES IN SOCIAL INTERACTIONS

Many previous studies have shown that individuals are biased towards attending to the eyes of others. In an image of a face, viewers disproportionately fixate on the eye region (Jack et al., 2008; Walker-Smith et al., 1977). Other studies have found that following the gaze of others is at least partially automatic (Itier & Batty, 2009). However, most of such studies are conducted in the lab environment, using unnatural stimuli such as images and videos and avoiding the social context associated with the interaction. In real-world situations, factors such as sociocultural norms and personality traits of the individuals involved have a profound influence on when, how frequently, and how long the gaze of another person is perceived and followed.

Gallup et al. (2012) demonstrated that walkers are less likely to gaze at other pedestrians and follow their gaze cues when the pedestrians are facing them. Foulsham et al. (2010) noted that walkers gaze at other approaching pedestrians less often in the real world than when watching a first-person video of a similar situation. Similarly, Laidlaw et al. (2011) showed that when people are seated in a waiting area with strangers, they are more likely to look at non-social objects in the environment than other people. Zuckerman et al. (1983) demonstrated that when in an elevator with a stranger, people initially show brief eye contact followed by prolonged gaze aversion. Taken together, these results suggest that the implicit bias humans exhibit towards looking at others, and specifically fixating at eye regions of others, is malleable.

On the other hand, gaze is a potent stimulus to initiate (Cary, 2006) and maintain conversations (Gullberg, 2003). In a live conversation, people actively perceive and follow the gaze cues provided by their conversation partners. The face of the conversation partner is one of the most fixated-upon areas in a face-to-face conversation (Gullberg, 2003).

Research on gaze patterns during face-to-face communication shows that speakers frequently look at their conversation partners (presumably to monitor the listener's state of attention and understanding). In comparison, listeners spend more time looking at the speaker (presumably to extract gaze signals and facial expression of the speaker; Cook 1977). Gaze cues towards and away from the partner also correlate with turn transition (Kendon, 1967). Speakers tend to gaze away from the partner when they start to speak and gaze back at the partner at the end of their utterance as a means to enable smooth turn transition. Also, speakers avert their gazes

during times when they are not ready for turn transition (e.g. during hesitation, or when conveying complex or emotional details) (Kendon, 1967). While there exist many generalisable gaze patterns in face-to-face conversations, there are also large differences in gaze behaviour based on individual gaze allocation characteristics (Kendon, 1967), personality traits (Cook, 1977), familiarity between participants (Broz, Lehmann, Nehaniv, & Dautenhahn, 2012), the type of interaction (Foddy, 1978), gender (Cook, 1977; Foddy, 1978), and culture (H. Z. Li, 2004).

Gaze allocation is closely linked with the semantics of speech. Griffin and Bock (2000) note that people look at things in the environment when speaking about them. In their study, participants speaking extemporaneously consistently fixated at objects for one second before naming them in their spoken description, providing evidence for a systematic temporal linkage between eye movements and spoken utterances. In a complementary line of research, Cooper (1974) showed that people tend to look at objects in the visual field when they hear a semantic reference to them, or a related word in the speech. Cooper (1974) presented participants with a variety of pictures on a computer display simultaneously with spoken language. They observed that participants spontaneously fixated at elements on the screen which are closely related to the meaning of the speech (e.g. looking at a picture of lion upon hearing the words "lion" or "Africa"). Their results suggest that eye movements are influenced by interpretation of the language heard.

2.5 ROLE OF EYES IN COLLOCATED COLLABORATIVE PHYSICAL TASKS

Two or more collocated individuals working together on a physical task that requires frequent referring and manipulating objects in the environment is very different from a typical conversational interaction. Such collaborations may often involve several non-verbal elements in addition to spoken language, such as pointing at objects in the environment using the hand, interpreting the partner's pointing target, monitoring the objects in the task space, manipulating the objects etc. The gaze allocation strategy and the functional role played by gaze could be influenced by these additional requirements, which are imposed by the characteristics of the physical task and the common goals of the collaboration.

Clark and Krych (2004) observed collocated collaboration in a LEGO building task. They found that, in such situations, people generally communicate with a variety of non-verbal cues, such as pointing, nodding, shaking the head and eye gaze. Macdonald and Tatler (2012) conducted a study to understand how people use the gaze cues of their partner in real-world collaborative tasks. The experimental task for the participants was to work in pairs to make a cake. Half of the pairs were assigned to

predefined roles of *chef* and *gatherer*, while the other half did not have any roles. They found that, across all conditions, participants spent more time mutually fixating at the objects required for the task and spent less time looking at each other. However, the results showed an interesting difference when the participants had predefined roles. *Gatherers* sought *chefs'* gaze cues more often than when no roles were defined (Macdonald & Tatler, 2012). Gaze cues provided by the chef may be more informative to the gatherer than in the situation in which no roles are defined – that is, people may seek the gaze cues of others depending on the perceived informativeness of the cue.

In an important follow-up study, Macdonald and Tatler (2017) found that when verbal instructions are ambiguous, people tend to seek, follow, and benefit from spatial cues provided by the gaze of the collaboration partner. In their study, an instructor had to use speech to convey the identity of one of the many objects arranged on the table, which the collaboration partner, seated frontally, had to select. In the gaze-cued condition, the instructor fixated at the object being referred to, while in the condition without gaze, the instructor read the speech from a paper. When gaze cues were available, participants actively sought the cues by initially fixating at the face of the instructor and made more correct selections when the verbal instructions could not uniquely identify the objects. In contrast, when the verbal instructions were unambiguous, participants seldom sought gaze cues provided by the instructor and no difference in task performance was found. These results suggest that people in naturalistic situations follow a flexible approach to seeking the cues provided by gaze. When speech is unambiguous, gaze provides little additional value and is hence ignored. In contrast, the value provided by gaze becomes greater when the language used is ambiguous.

Hanna and Brennan (2007) note that when communication partners have to convey spatial information, gaze cues, available through the head orientation and eye position, help disambiguate referring expressions much earlier than the linguistic point of disambiguation. In a follow-up study designed to tease out the role of head orientation and gaze direction (S. S. E. Brennan, Hanna, Zelinsky, & Savietta, 2012), instructors wearing mirrored sunglasses that would prevent an onlooker from perceiving accurate gaze cues but allow perceiving head orientation provided verbal instructions required to identify spatially arranged objects. They found that head orientation information alone was less informative and incurred a cost in accuracy when other competitor objects were located near the referred object. Boucher et al. (2012) extended this work and found that when eyes are visible, as opposed to situations when the instructor is wearing sunglasses, the efficiency of collaboration was improved by reducing the time needed for the participants to select the objects.

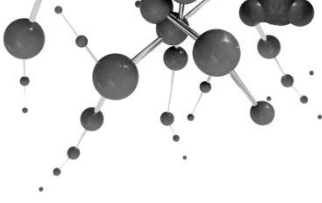
More recently, Garcia et al. (2017a) conducted an empirical study to understand the value of gaze for multimodal referentiality in naturalistic collaborative tasks. In their study, participants worked in pairs, standing face-to-face to each other across a table, to arrange different objects in predefined abstract shapes. The arrangement was known to one collaborator, while the other could only manipulate the task space. The researchers manipulated the availability of gaze cue as an independent variable. In one of the experimental conditions, both collaborators wore goggles that prevented the visibility of eyes to the partner; in the other condition, the participants collaborated without the goggles. The results suggest that the availability of gaze cues not only enabled higher joint task performance, but also led to higher frequency of deictic references and reduced frequency of conversational repair.

Interestingly, Garcia et al. (2017a) reported that familiarity between participants modulates their communication strategy. Pairs who were familiar with each other more actively used each other's gaze cues. When the pairs were not familiar with each other, they were also reluctant to engage in direct eye contact and to infer the spatial information encoded in the eye position, despite its utility in the task. Similar observations were also reported by Macdonald and Tatler (2017).

Knowing where one's collaboration partner is looking is useful in collocated collaborative physical tasks, even though social norms and personality traits can modulate the usefulness. How do these benefits translate to scenarios involving video-mediated remote collaborative physical tasks? Answering this question is the focus of this thesis.

Summary of the chapter

- *Humans are remarkably good at perceiving both dyadic and triadic gazes of an onlooker.*
- *Gaze cues provide multiple benefits in our social interactions and collocated collaborative physical tasks*
- *The reluctance of people to engage in eye contact reduces the benefits of gaze cues in collocated collaborative physical tasks*



3 Gaze tracking and Gaze-Based Human Computer Interaction

3.1 BASIC CONCEPTS OF GAZE TRACKING

Understanding how our eyes move has been instrumental in gathering intricate details regarding our sense of vision and how we perform visuo-cognitive tasks such as reading. The earliest studies in this area used direct physical observation of the eyes to collect information about eye movements. It wasn't until the late 19th century that devices to assist in measurement of eye movements were developed. The earliest such devices were mechanical in nature and invasive to use, requiring the device to be directly attached to the eyes of the user. However, with years of technological advancements, a variety of non-invasive solutions have been developed which assist in measuring movement of the eyes.

Based on the underlying technology, contemporary gaze-tracking systems can be classified into three broad categories: electro-oculography (EOG), scleral search coil, and video oculography (VOG).

EOG relies on the electrostatic charge difference between the cornea and the retina of the eye (Mowrer, Ruch, & Miller, 2017). The cornea is 0.40 mV to 1.0 mV positively charged relative to the retina (Young & Sheena, 1970). As the eyes move, the electric dipole moves with them, causing a variation in electric potential around the eyes. Skin electrodes strategically placed around the eyes can detect this variation in electrical potential to measure the movement of the eyes in relation to the head. The recorded potentials are small, in the range of 20 to 200 μ V, with a sensitivity of the order of 20 μ V/deg of eye movement (Young & Sheena, 1970). EOG-based gaze-

tracking devices have the advantage of simpler information processing and power requirements. On the downside, they need sensors to be attached on the skin around the eyes and suffer from problems of accuracy due to interference from other bio-signals (e.g. due to muscle activity) and external electrical interferences (Young & Sheena, 1970).

The scleral search coil method requires the user to wear a contact lens embedded with a thin wire coil, which is also connected to an external voltage measurement unit. When the coil is subjected to a known alternating magnetic field, a voltage is induced in the coil according to Faraday's law of induction. The induced voltage depends on the orientation of the coil, and hence on the orientation of the eye. Normally, multiple orthogonal magnetic fields operating at different quadratures or frequencies are used to measure the eye position along its multiple degrees of freedom (Robinson, 1963). The scleral search coil method for gaze tracking, although invasive, is very accurate and can track eyes at a very high sampling rate. It is used in the medical field for research, as well as for diagnosis of neurologic, ophthalmologic, and vestibular disorders (Houben, Goumans, & Van Der Steen, 2006). More recently, the scleral search coil contact lens method was proposed as a feasible gaze-tracking technique to interact with virtual-reality headsets (Whitmire et al., 2016).

VOG, on the other hand, is a camera-based technique that relies on advanced computer vision to landmark characteristic points of the eye area (e.g. center of pupil, limbus, corner of the eye, etc.). The direction of gaze is calculated based on the position of these landmark points. In general, two or more landmark points are required to estimate the point of gaze, with at least one point that is independent of eye movement (e.g. corner of the eye) and one point that is dependent on the eye position (e.g. pupil centre). The specific landmark points tracked depend on the algorithm used. There are two broad categories of VOG gaze-tracking techniques based on the illumination source used. Active illumination techniques use near-infrared illumination to track the gaze, while passive illumination approaches rely on visible light (Hansen & Ji, 2010). In active illumination trackers, the infrared light source is either placed on or off the optical axis of the video camera, rendering the pupil of the eye bright (when IR light source is placed on axis) or dark (when IR light source is placed off axis) in contrast to the iris. This high contrast enables robust tracking of the pupil. In addition, the infrared light is reflected at the surface of the cornea, creating a glint in the camera image. The position of the glint remains static and is invariant of the eye orientation. Active illumination tracking relies on the position of the glint and the pupil as landmark points to estimate the gaze vector. For a detailed review of the different eye landmark detection and gaze estimation techniques, see Hansen and Ji (2010).

In terms of the gaze-tracking technology used, this thesis will focus on VOG-based active illumination gaze tracking. Currently, this is the most commonly used gaze-tracking device setup for the purpose of interacting with computers. Further, all the research work done as part of this thesis employed active illumination VOG-based gaze trackers. Figure 3 shows the two most common form factors for VOG-based gaze trackers: (a) head-mounted and (b) remote.

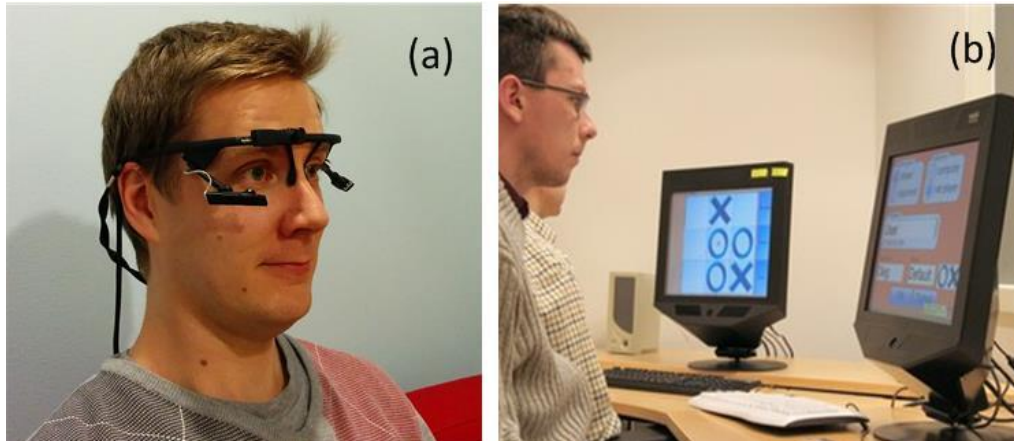


Figure 3. Different types of VOG gaze trackers. (a) User wearing a PUPIL 120Hz binocular head-mounted gaze tracker¹. (b) User interacting with a Tobii remote gaze tracker²

3.2 GAZE-BASED HUMAN COMPUTER INTERACTION

Gaze tracking is gradually transitioning from being a niche technology towards the mainstream consumer market. Microsoft Windows now supports gaze trackers as a standard input device. Gaze trackers comes integrated in gaming laptops (e.g. Alienware 17³), VR headsets (e.g. HTC Vive Pro⁴) and AR devices (e.g. Microsoft HoloLens II⁵). These recent developments in the consumer market stands as an evidence for the maturity of the technology and its promise in HCI.

There are two key challenges in using gaze for HCI. First is coping with the Midas-touch problem – that is, the difficulty in distinguishing between gaze shifts that are part of perception and those that are directed as commands to the computer (P Majaranta & Rähkä, 2002). The second challenge is overcoming the issues related to gaze data quality.

There are different gaze-based interaction techniques that use various strategies to overcome these challenges. Dwell-based interaction relies on

¹ <https://pupil-labs.com/products/invisible/> (accessed 7 July 2019)

² Source: <http://www.uta.fi/sis/tauchi/virg/laboratory.html> (accessed 7 July 2019)

³ <https://gaming.tobii.com/onboarding/> (Accessed 1 July 2019)

⁴ <https://enterprise.vive.com/ca/product/vive-pro-eye/> (Accessed 1 July 2019)

⁵ <https://www.microsoft.com/en-us/hololens/> (Accessed 1 July 2019)

prolonged staring (or dwelling) to distinguish gaze commands from regular eye movements. For example, disabled users type on an on-screen keyboard by staring at each key for a defined duration of time (P Majaranta & Riih , 2002). A shortcoming of dwell-based interaction is that it is sensitive to gaze-tracking accuracy. A typical way to deal with low gaze-tracking accuracy in dwell-based interaction is to make the screen area of the interactive elements larger. This approach is obviously not sustainable in devices with smaller displays, or applications with large numbers of interactive on-screen elements (e.g. an on-screen keyboard).

Gaze gesture is another gaze-based interaction technique. Gaze gesture requires the user to perform a sequence of saccades in a predefined pattern such as making a Z gesture with the eyes, normally within a limited time period. These predefined movements are considered as commands to a computer to perform a predetermined action. The gaze gesture pattern needs to be simple, so that the user can remember and perform the gesture with ease. At the same time, the gesture needs to be unique, so that it does not occur as part of normal gaze behaviour. An advantage of a gaze gesture is that it relies on relative eye movements and is less sensitive to gaze-tracking accuracy.

A more recent gaze-based interaction technique is to use smooth-pursuit eye movements. Smooth-pursuit gaze interaction requires a display with targets moving on predefined trajectories. When the user visually follows any specific target, the trajectories of the object and gaze are matched and the command associated with the followed object is performed (M lodie Vidal, Pfeuffer, Bulling, & Gellersen, 2013). An advantage of smooth-pursuit gaze interaction is that, like gaze gestures, it is less sensitive to accuracy of tracking and can be performed using even an uncalibrated gaze tracker. On the downside, this technique requires visualising a moving target to produce the corresponding pursuit eye movements (i.e. it needs a display). Further, the user is required to follow the moving target long enough to differentiate between natural eye movement and the intentional target following meant as an input to the computing device.

Dwell-based interaction, gaze gestures and smooth-pursuit interactions all have two things in common: They all rely on gaze as the sole interaction modality, and all of them require explicit use of the eyes to interact. Such gaze-only explicit interactions may be suitable for specific situations (e.g. when the user's hands are occupied), functionality (e.g. to calibrate the gaze tracker), or user groups (e.g. disabled user group). However, its utility outside the niche usage context may be limited.

On the other hand, a person's gaze, even if produced without the intention to communicate, is naturally informative. HCI researchers have long argued that attentive computing systems using this implicit information contained in our natural eye movements, as opposed to requiring explicit

use of the eyes to interact, may have much wider applicability among mainstream users (Päivi Majaranta & Bulling, 2014).

In addition to implicit gaze-based interactions, the strength of gaze as an interaction technique can be harnessed in multimodal interfaces that combine the wealth of information provided by gaze with the explicitness and flexibility of conventional interaction techniques. There is a growing amount of research in multi-modal interfaces where gaze is used as a complementary input modality. For example, gaze can be used for pointing and touch for selection while interacting with on-screen or physical objects (Kumar et al., 2007; Stellmach & Dachsel, 2012), or computer games can be played with a gamepad, where a player's visual attention is used as a complementary input to augment the social interactions inside the game (Melodi Vidal, Bulling, & Gellersen, 2015).

3.3 GAZE TRACKING DATA QUALITY

The data returned by a gaze tracker, in addition to other parameters, includes the (x, y, z) coordinates of the point the user is currently looking at. However, this data contains both a noise component and systematic error. The quality of gaze data could potentially influence the validity of research results when a gaze tracker is used as a research tool, as well as influencing quality of interaction when gaze is used as an input mechanism in HCI (Holmqvist, Nyström, & Mulvey, 2012). The utility of a gaze tracker is dependent on the quality of the gaze data it can generate.

Gaze-tracking-data quality can be broadly divided into its spatial quality, robustness, and temporal delay. Spatial quality includes two different aspects: accuracy and precision of tracking. Accuracy of gaze data is the measure of the difference between the true point of gaze and the point of gaze estimated by the tracker. On the other hand, precision of gaze data indicates how consistent the gaze samples are when the true point of gaze is constant (Holmqvist et al., 2012). Figure 4 shows a visualisation of gaze data in terms of its spatial quality, generated using TraQuMe, a tool for measuring gaze data quality developed as part of this thesis (Akkil, Isokoski, Kangas, Rantala, & Raisamo, 2014).

Robustness of gaze data or trackability indicates the extent to which the tracker can deliver valid gaze data. Sometimes, a VOG-based gaze tracker with a fixed sampling rate returns invalid data that indicates that eyes cannot be tracked. This would be the expected behaviour when no user is present, when the user is looking away from the gaze tracker, or when the

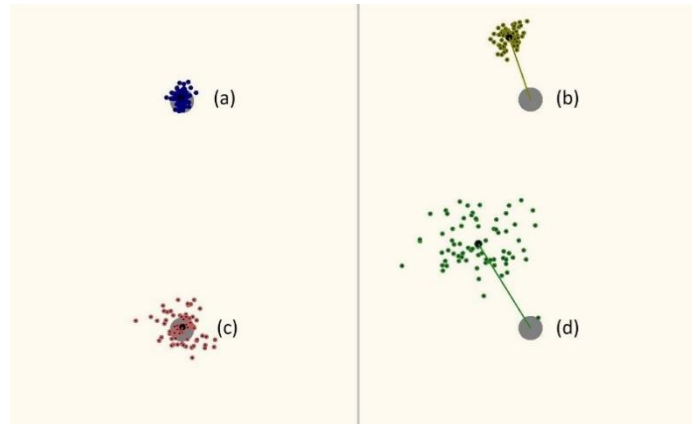


Figure 4. Visualisation of gaze data in terms of spatial quality. The figure illustrates the point of gaze generated by a gaze tracker when the user is fixating at a specific point (shown as a grey dot). (a) Gaze data with high accuracy and precision, (b) data with good precision, but low accuracy, (c) data with high accuracy, but low precision, and (d) data with low accuracy and low precision.

user blinks while interacting with the system. However, in some cases, the tracker may fail to report the gaze data even when a user is present. For example, this could be due to wrong positioning of the device relative to the user, such that eyes are not visible clearly in the image, or failure of the tracking algorithm to detect or landmark the eyes in the image.

Most commercial gaze-tracking manufacturers state an ideal condition accuracy of 0.5 degrees and precision of 0.1 degrees in their product specifications (Tobii Technology, 2016). This data quality is often measured in an ideal tracking environment (stable lighting, stable screen luminance, no other IR interference, strategically placed light sources to avoid unwanted reflections, etc.), on either artificial eyes or “best” participants using a head rest. For example, The gaze tracker manufacturer Tobii AB developed a gaze-tracking data quality measurement methodology (2011). In their method, 90 participants are first selected from a test pool of 200 participants based on the criteria of normal vision, no history of eye surgery or other eye conditions, and no droopy eyelids or narrow eye shape. From the 90 participants who take part in the study, 40 participants with the best gaze-tracking accuracy and precision are selected for further analysis. In short, the manufacturer-specified quality measure indicates the ideal system performance in optimal conditions for “best” participants.

Holmqvist et al. (2012) note that characteristics of the user, the gaze tracker, the test environment, and the task may influence the accuracy, precision, and robustness of gaze data.

- *Characteristics of the user:* Some users may wear eyeglasses or contact lenses, or have long eyelashes or droopy eyelids. All these personal characteristics of the user may influence how clearly the camera can see the eyes and track the characteristic points. Blignaut

et al. (2013) compared gaze data quality for users from different ethnic backgrounds and found that gaze trackers provide more accurate, precise, and robust data for African and European users relative to East Asian users. East Asian eyes appear narrow externally, and this could influence the gaze data quality.

- *Characteristics of the gaze tracker:* The number and resolution of the camera(s) used in the gaze tracker, relative positioning of the camera affording a good view of the eyes, the algorithm used to track characteristic points of the eye and estimate point of regard, the calibration procedure used, and whether the tracker is monocular or binocular are some of the gaze tracker characteristics that influences gaze data quality.
- *Characteristics of the environment and task:* The presence of other infrared light sources and vibrations in the environment are some of the environmental characteristics that adversely influence the quality of gaze data. Characteristics of the task, such as those that require frequent movements or require the user to be at too small or large distances from a remote gaze tracker or make large gaze angles, could also influence the quality of gaze data.

The temporal delay of gaze data indicates the latency between a gaze event taking place and the corresponding gaze data being delivered by the tracker. In a VOG-based gaze tracker, the temporal delay is the sum of latencies incurred in acquisition of the image from the camera, processing of the image to estimate the gaze point, and delivering the gaze data to the application. The temporal delay is influenced by the processing power of the computer and sampling rate of the tracker. The delay in commercial gaze trackers working on dedicated computers is less than 55ms (Gibaldi, Vanegas, Bex, & Maiello, 2017), and this likely goes unnoticed in gaze-based interaction. However, the delay may be an issue when gaze trackers are integrated with wearable devices with lower computation power, or when gaze-tracking systems relying on client-server models emerge. Also, unlike other gaze data quality measures, the temporal delay of gaze data is mostly a system characteristic and less dependent on characteristics of the user or the environment.

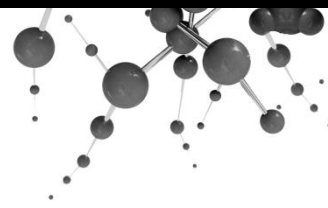
Many research studies have indicated the importance of gaze data quality in ensuring research validity (Blignaut & Wium, 2014; Holmqvist et al., 2012; Nyström, Andersson, Holmqvist, & van de Weijer, 2013). However, gaze data quality is an aspect that is still not given enough attention in the gaze research community. This is evident from the fact that vast majority of gaze-tracking research does not measure or report the gaze data quality metrics in the paper. The papers that do report on accuracy mostly rely on accuracy and precision values provided in manufacturer specifications, which can be very misleading. In the very few papers that do measure and

report the data quality, no standard and comparable way of measuring and reporting gaze data measures exists.

One of the goals of this thesis is the development of an open-source gaze data quality measurement system, called TraQuMe (Study I). TraQuMe is a light-weight and tracker-independent data quality measurement software that enables easy recording, analysis, and reporting of gaze data quality.

Summary of the chapter

- *Gaze-tracking technology is more affordable and ergonomic to use than ever before and is increasingly available in the mainstream consumer market.*
- *Characteristics of the gaze tracker, user, and environment influence gaze data quality.*
- *There is a need for a tracker-independent and flexible gaze data quality measurement tool.*



4 Gaze Sharing in Computer-Mediated Communication

Research on shared gaze in computer-mediated communication is not new. In this chapter, I introduce the research domain and present a literature review.

4.1 INTRODUCTION TO THE LITERATURE REVIEW

The earliest research on sharing gaze information between remote collaborators focused on conversational video conferencing. For example, in 1987, Acker and Levitt developed GazeCam, a video-conferencing system that uses mirror arrangements to remove the parallax associated with camera positioning. GazeCam facilitated eye contact between remote video conferencing participants.

Ishi and Kobayashi (1992) extended the concept from conversational video conferencing scenarios to task-oriented remote collaborations. They developed Clearboard, a platform that enables collaborative drawing whilst also providing awareness of where the collaboration partner is looking. Clearboard used the metaphor of “talking through and drawing on a transparent glass window”. Users saw the video feed of the partner and the drawing area overlaid on each other without the need to shift their attention between the two. Later, Monk and Gale (2002) devised a system based on the Clearboard architecture but using two separate displays. A semi-transparent display showed the shared workspace and a separate display behind it, to present the video of the collaborator. By decoupling

the two displays, more accurate gaze awareness in joint drawing tasks was provided.

All of the studies discussed thus far in this chapter have one aspect in common. They provide gaze awareness by presenting real-time video of the collaboration partner's face. This approach is natural but has several shortcomings. First, these methods require presenting the face of the partner along with a view of the activity space, which may not be desirable in certain situations (O'Hara, Black, & Lipson, 2006). Second, the accuracy of interpreting the point of regard is influenced by the limits of accuracy of human gaze perception and characteristics of the technological setup. Third, ascertaining the point of regard requires looking near the region of the eyes of the partner. Several factors, such as personality traits of the user and social norms pertaining to eye contact, may modulate how often the users seek, follow, and benefit from the gaze cues (García et al., 2017).

Another option is to use a gaze tracker to estimate the collaborator's point of regard and to present this information to the partner- for instance, as an abstract visual cue overlaid on the video.

Such an approach simplifies the gaze interpretation for the viewer and can present more accurate gaze information. On the other hand, communicating the gaze of a person artificially as an abstract element raises the question of whether such a presentation involves the same cognitive processes involved in natural gaze perception and gaze following. Our perception and interpretation of another person's gaze reflects our understanding of the differences in each other's fields of view and the spatial relationships of objects around us. It shows our awareness of the communicative significance of eyes and our recognition that the gaze of a person reflects their mental state. It reflects our understanding that the gaze of a person may not always be informative of attention (e.g. staring plainly at something) and can be manipulated to deceive. We perceive, process, and interpret a person's gaze and make judgements about its underlying meaning instantaneously and instinctively. Further, other cues, such as the facial expression of the person, modulate how his or her gaze is utilised by an onlooker (Bayliss, Frischen, Fenske, & Tipper, 2007). Additional cues such as facial expression may be completely missing when gaze is communicated artificially. It is debatable whether communicating the gaze of a person as an artificial visual cue may be followed as preferentially, perceived as intuitively, or decoded as effortlessly as compared to perceiving gaze of a person by looking at their face.

On the other hand, presenting the partner's gaze information artificially provides several pragmatic benefits, that of a pointer that automatically and intuitively conveys our spatial attention. Further, our belief that an

artificially presented cue is representative of the gaze of a person can modulate our low-level mechanisms of visual attention (Tufitt, Gobel, & Richardson, 2015). When viewers believe that an artificial cue, such as a dot in a video, is communicating the focus of attention of another person, they respond differently to when they believed that the cue was not associated with an implied social context.

Velichkovsky (1995) was the first to study the value of gaze awareness in task-oriented collaboration in an applied setting. With a shared view of the computer screen, two remote users collaborated to perform a puzzle-solving task. One of the users knew the solution to the task, but could not act on the puzzle. The other user could perform the task, but lacked the knowledge of how to solve the puzzle. Instead of presenting the video of the face of the partner to provide gaze awareness, the study used gaze tracking to estimate the point of regard of the user on screen and visualised this information on the display of the remote partner, in the form of a semi-transparent dot. The study showed that sharing gaze information can improve the efficiency of collaboration and change the nature of dialogues between the collaborators.

Following the promising study by Velichovsky in 1995, numerous others have investigated the value of applied gaze awareness in various collaborative and communicative-use contexts. Numerous workshops have been organised under the theme of Dual-Eye Tracking (e.g. DUET 2011, 12, 13), and the collaborators' real-time gaze-tracking technique has been proposed as a novel methodology to not just support collaboration, but to also study the dynamics and quality of collaboration (Jermann, Nüssli, & Li, 2010), predict expertise of the collaborators (Y. Liu, Hsueh, & Lai, 2009), predict the social context of the collaboration (W. Li, Nüssli, & Jermann, 2010), and detect misunderstandings during collaboration (Cherubini, Nüssli, & Dillenbourg, 2008). Other research has also focused on developing novel technological frameworks and software tools that enable fast and synchronised sharing and recording of gaze information during collaboration (Nyström, Niehorster, Cornelissen, & Garde, 2017).

There is also increasing commercial interest in shared-gaze interfaces. For example, Tobii Ghost⁶ allows real-time livestreaming of casual gaming and e-sport sessions, with gaze overlay to the audience. Sprint⁷ by Tobii is another service that allows users to share a desktop screen with gaze overlay on it with remote collaborators in real time. Sprint is a platform for designers and researchers to effortlessly harness the power of gaze tracking in remote-user testing.

⁶ <https://gaming.tobii.com/software/ghost/> (accessed 04 March 2019)

⁷ <https://www.tobiipro.com/sprint/> (accessed 04 March 2019)

As the domain of shared gaze interfaces is expanding, there is currently a lack of coherent understanding of the type of previous work that has already been undertaken in this area, the results these studies have provided in their specific contexts of evaluation, and the subtle differences between the different studies.

A comprehensive literature review that provides a holistic view of the topic is currently missing. Such a literature review would also help in understanding the originality and contribution of this thesis, in the larger context of research in this domain. In this and the following chapter, such a review is presented. In this chapter, I present an overview and classification of the research domain. In Chapter 5, I focus on previous research on shared gaze interfaces for real-time remote collaboration.

The literature review presented here analyses all the publications so far on gaze sharing in computer-mediated communication, including the publications produced as part of this thesis, and the more recent publications afterwards. Such an approach is taken to present a coherent and complete review of the work in the domain and to do justice to the cumulative nature of research undertaken in this domain. Wherever relevant, I will highlight the research performed and the contributions made as part of this thesis.

4.2 METHODOLOGY

The focus of this literature review is on task-oriented video-mediated communication, where the gaze information of the partner is communicated artificially as an abstract visual cue as opposed to users interpreting the gaze of the partner directly from the video of his or her face. For this review, I used two approaches to gather the relevant publications. First, I selected a few of the popular and pioneering works in the domain, such as that of Velichkovsky (1995), Brennan et al. (2008b), Stein and Brennan (2004), and Qvarfordt et al. (2005). Then, I used a snowball sampling technique to find relevant papers that either cited or were cited by these publications. Second, I used Google Scholar to search for relevant papers, using the focused search query “Communication” AND (“Gaze Sharing” OR “Gaze transfer” OR “Shared Gaze”) AND “Video” to find papers that might have been missed in the earlier approach. The query returned 475 results (on 20 June 2018). I then reviewed the titles and abstracts of the results to judge the relevance of the papers to the literature review. I removed papers that did not meet the criteria from the collection. I did not include any possible duplicates and publications that were not from peer-reviewed venues (e.g. master’s theses, white papers, etc.).

At the end of this stage, 92 papers remained, all focused on the topic of interest. However, a few of these papers focused on concept presentations or related technology frameworks (e.g. Nyström et al. 2017) without an experimental evaluation. Such papers were filtered out. Lastly, there were a few studies related to perceptual skill transfer from an expert to a novice that did not explicitly use the gaze data of the expert directly to train the expert but instead used either verbal instructions or simulated gaze representation to convey expert gaze pattern. Such studies were also excluded from the analysis.

Finally, there were 73 peer-reviewed publications. Figure 5 shows the distribution of the publications included in the literature review, according to the year of publication.

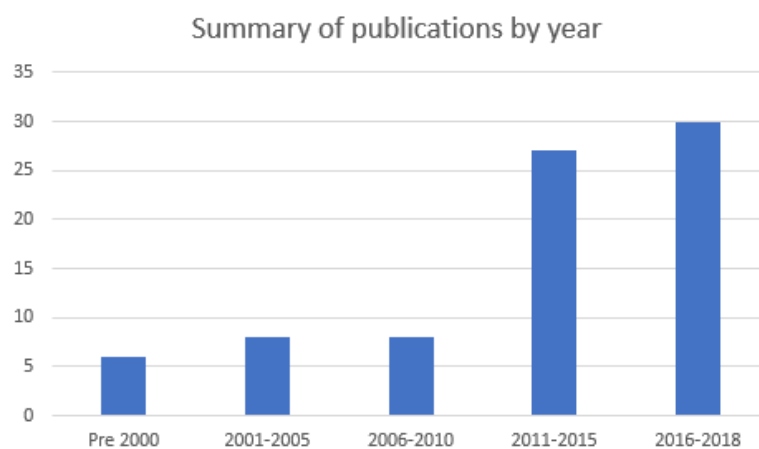


Figure 5. Distribution of publications included in the literature review. The literature review was conducted in June 2018.

4.3 CLASSIFICATION BASED ON JOHANSEN'S TIME-SPACE MATRIX

An obvious technique to classify the shared gaze communication systems is to use Johansen's time-space matrix (Johansen, 1988). The time-space matrix involves a 2 X 2 matrix based on the time (whether the system is designed for real-time or non-real-time use) and space (whether the system is designed for collaboration between collocated or remote users) characteristics. Table 1 shows the time-space matrix for shared gaze interfaces, with an example for each.

For some of the studies in the literature review, the time-space characteristics were not evident. For example, systems intended for remote collaboration were evaluated in an experimental setup in which the participants were collocated but separated visually using a physical barrier (e.g. D'Angelo & Begel 2017). Such a setup has pros and cons. It provides experimental simplicity, without the need for any Internet-based audio/video-streaming requirements. On the other hand, it reduces the

generalisability of the findings to “in the wild” contexts. Furthermore, systems intended for real-time collaboration were experimentally evaluated in studies conducted in multiple phases without real-time collaboration (e.g. Akkil & Isokoski 2016a). Such a study design provides better focus on the research question by eliminating the effect of other variables introduced as part of the complexities of real-time collaboration. The time-space matrix presented in this literature review is based on the intended use of the collaboration system, not on the design of the studies used in the evaluation.

Furthermore, for most studies involving non-real-time communication, the location characteristics were not clear or not important (e.g. systems designed for skill transfer from an expert to a novice). In these systems, it made little difference whether the video of the expert performing the task was recorded in the same or a different location. Thus, publications with non-real-time characteristics in which the location was not an important factor were grouped together in the “remote” category.

Time-space characteristics	# of publications	Example
Non-real-time, remote	29	Augmenting Massive open online course (MOOC) video with the gaze of the instructor (Sharma, Jermann, & Dillenbourg, 2015a)
Non-real-time, collocated	2	Sharing the gaze information of students involved in a reading a reading activity to the class teacher for later analysis (Špakov, Siirtola, Istance, & Riih�, 2017)
Real-time, remote	35	Gaze sharing during remote pair programming (D’Angelo & Begel, 2017)
Real-time, collocated	7	Sharing the gaze of a co-passenger with the driver during collaborative car navigation (Maurer, Tr�sterer, et al., 2014)

Table 1. Classification of previous work on shared gaze interfaces based on Johansen’s time-space characteristics, with an example for each from previous literature.

Table 1 shows the classification of previous work based on Johansen’s time-space characteristics.

(a) Gaze Sharing in Non-Real-Time Remote Communication

A majority of the studies in this time-space matrix investigated skill transfer (e.g. from an expert to a novice) in a variety of different domains, such as inspection, classification, reading, programming, and psychomotor tasks. They focused on answering the following question:

Does viewing a video of a model with the gaze overlay help novice users to perform the same task more efficiently?

Table 2 presents a summary of the 29 publications with their domain of investigation.

Publication	Domain of Investigation
(Litchfield, et al. 2008, 2010; Nalanagula, Greenstein, & Gramopadhye, 2006; Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005; Seppänen & Gegenfurtner, 2012; Sridharan, McNamara, & Grimm, 2012)	Inspection and interpretation
(Causer et al. 2014; Moore, Vine, Smith, Smith, & Wilson, 2014; Joan N. Vickers, Vandervies, Kohut, & Ryley, 2017; Vine, Moore, & Wilson, 2011; Vine & Wilson, 2010, 2011; Wilson et al., 2011; Wood & Wilson, 2011, 2012)	Precision skill training (e.g. surgery, sports, marksmanship)
(Gallagher-Mitchell, Simms, & Litchfield, 2017; Jarodzka et al., 2012; Jarodzka 2009; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Wu, Shimojo, Wang, & Camerer, 2012)	Classification and estimation
(Kushalnagar, Kushalnagar, & Manganelli, 2012; Sharma, D'Angelo, Gergle, & Dillenbourg, 2016; Sharma, Jermann, & Dillenbourg, 2015b)	MOOC video-based learning
(Mason, Pluchino, & Tornatora, 2015, 2016)	Reading
(van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009; van Marlen, van Wermeskerken, Jarodzka, & van Gog, 2016)	Procedural problem-solving
(Stein & Brennan, 2004)	Software program debugging
(Litchfield & Ball, 2011)	Perceptual problem-solving

Table 2. Research on shared gaze for remote, non-real-time applications

The eye movement strategy greatly varies between expert and novice users in complex tasks. Reingold and Sheridan (2012) noted two characteristic differences in eye movement patterns of experts and novices. First, experts exhibit superior encoding of visual information. When presented with a visual scene, experts are faster to gain a global understanding of the scene by following an optimal scan path, often consisting of fewer but longer fixations. Second, experts exhibit specific gaze patterns as a result of the implicit task-related knowledge that they might have obtained through their years of experience and acquired expertise. Experts fixate on task-relevant areas more often and ignore task-irrelevant areas, sometimes even without being consciously aware of it. For example, elite hitters in fast ball sports, such as cricket, baseball, tennis, and squash, track the fast-moving ball for longer periods than amateurs do (a T. Bahill & Laritz, 1984). Elite hitters also make predictive saccades to the anticipated position of the ball much earlier than amateurs. The pronounced differences in gaze patterns between experts and novices go far beyond the field of elite sports and could very well exist in everyday situations, such as reading, searching, driving, and problem-solving tasks (Land & McLeod, 2000).

Similarly, Stein and Brennan (2004) showed that viewing the gaze pattern of an expert debugging a software program can enable novice users to solve the same task faster. They noted that eye gaze, even if produced instrumentally without the explicit intention to communicate information, can be beneficial to others performing the same task. Litchfield and Ball (2011) obtained similar results in a Duncker's radiation problem, another perceptual problem-solving task.

Similar results were found in inspection tasks. Nalanagula et al. (2006) studied whether viewing the gaze pattern of an expert would help novice users in a printed circuit board inspection task. They found that novice users, when trained using dynamic gaze visualisation of the expert, showed an improved transfer performance compared to that with static gaze visualisation or with no visualisation at all. Their results suggest that visualisation of the gaze of the model influences the benefits of the training videos and that such novel training methods can foster learning. Sadasivan et al. (2005) found similar results in an aircraft inspection task. Novice inspectors performed better when they were shown the gaze pattern of an expert, although their task completion times increased. Seppänen and Gegenfurtner (2012) showed that seeing an expert's gaze enabled novice radiographers to focus on task-relevant areas and fostered learning.

Litchfield et al. (2008, 2010) studied the effect of the expertise of both the model and observers on task performance in a task requiring identification of pulmonary nodules in chest X-rays. They found that both novice and

expert radiographers benefitted from gaze visualisation of a model performing the same task, irrespective of the expertise of the model. Their results suggest that, along with expertise-related eye movement, there might be task-related components of eye movement that could also be informative to viewers performing the same task.

In inspection and problem-solving tasks, seeing another person's gaze can help a person to perform the same task faster compared to receiving no other cues. This is, of course, no surprise. Knowing where others looked while solving the same problem could give the user insights into the task-relevant regions. One could also argue that any additional information in such a context could be more useful than no information at all (e.g. knowing how others used their mouse while solving the same problem, or hearing their think-aloud comments, could also arguably lead to some benefits). These studies do not necessarily show that gaze is a superior information signal when compared to other possible alternatives.

Researchers have compared gaze sharing with alternative communication channels in skill transfer as well. Sridharan et al. (2012) showed that showing an expert's gaze point or explicit selection using the mouse can be helpful in mammography training, with gaze sharing showing a slightly better short-term transfer effect. Similarly, Gallagher-Mitchell et al. (2017) found that viewing training videos of an expert performing a number line estimation task, with either the gaze or mouse cursor visualised, led to better performance than a control condition involving self-training. Interestingly, gaze- and mouse-based learning led to similar performance, with the mouse being marginally more accurate. Sharma et al. (2016) compared gaze visualisation and pen pointer visualisation to a control condition of no visual aid to convey deixis in MOOC videos. They found that gaze visualisation led to improved learning compared with having no visual aid, and no difference between gaze and pen visualisation was found. However, they observed that students spent more time looking at the task-relevant areas in the gaze condition than in the pen-based visualisation condition. Overall, the value of gaze-augmented training videos is task dependent and might not always be more effective than overlaying mouse/pen visualisation to indicate active areas.

In all the previous examples, novice users viewed a video of a model performing a perceptual task, with the gaze of the model overlaid on the video. The novice users could see where the model fixed his or her gaze at different points during the task. However, the novice viewers might not have always been able to understand why the model looked at those areas.

Eye movement modelling examples (EMMEs) (Jarodzka et al., 2012; Jarodzka, Holmqvist, & Gruber, 2017; Jarodzka et al., 2009, 2013; van Gog et al., 2009) are gaze-augmented videos of experts, often produced

didactically along with explicit verbal descriptions, designed to teach novice users to perform complex perceptual tasks. When shown the video of an expert performing a similar perceptual task, with the gaze of the expert overlaid on the video, novice users can gain an understanding of the tacit strategies and perceptual processes followed by the expert. When accompanied with explicit verbal instructions, novice users can learn where the expert looked at specific points during the task and why he or she looked at those points.

Jarodzka et al. showed the value of EMMEs in tasks requiring clinical reasoning (2012) and educational classification (2009, 2013). They noted that in visually rich learning materials, using visualisation of an expert's gaze fosters learning (Jarodzka et al., 2013). Mason et al. (2015, 2016) showed that EMMEs can help children in reading tasks by enabling them to better integrate texts with related illustrations. Children receiving EMME training spent more time transitioning from text to visual representation and strategically spent a longer time re-inspecting the pictures while rereading the text. They also showed improved verbal and graphical recall.

In contrast to studies that showed a positive effect of EMME training, Gog et al. (van Gog et al., 2009) showed that EMME training might not be more effective than normal videos in helping users solve procedural tasks. Classification tasks and reading strategy tasks used in other EMME studies involved users inspecting or viewing the content on the screen, without explicitly acting on it. In contrast, procedural problem-solving involves the model acting on the on-screen content either by using a mouse or by typing. Such overt actions also indirectly communicate the attention of the model. In such cases, the redundancy offered by the gaze does not help in learning and can, in fact, be detrimental. Similar results were obtained by Marlen et al. (2016).

In summary, showing the gaze information of an expert involved in a similar task can help communicate the perceptual processes of the expert. Such modelling examples can be useful in a wide variety of classification, inspection, estimation, reading, and problem-solving tasks. They provide two benefits. First, they can help novice users performing the same task to improve their efficiency. Second, they can enable novice users to learn the problem-solving strategy of the expert and transfer this knowledge in novel situations. It is unclear if gaze visualised as an abstract visual cue overlaid on the display is a more useful signal for problem-solving than other possible signals, such as communicating mouse position or explicit verbal instruction. Also, when the attention of the model is already available in the form of other interactions such as mouse movement, redundancy provided by gaze sharing does not aid in learning and could be detrimental.

Shared Gaze for Psychomotor Training

So far, all of the studies discussed involved tasks performed on a computer display, where other modalities such as mouse and pen pointers are also feasible. An important question to ask is, can communicating the gaze information of a model be helpful in physical tasks? A significant differentiation here is that in physical tasks, other modalities are less likely to provide the extent of information made possible by gaze sharing. For example, in basketball training, the gaze of an expert could uniquely provide key insights into his or her visual processing strategy. The hands of the expert might already be occupied in the task, and providing verbal instructions might not be feasible in such a fast-paced scenario. Furthermore, some of the strategies used by players in these complex scenarios might be subconscious, such that they themselves are not fully aware of them. Gaze tracking could thus be a valuable tool in such a training routine.

Previous research in the area of precision skill training and hand-eye coordination has highlighted the role and importance of a “quiet eye” period in tasks that require precision psychomotor skills such as aiming and interceptive tasks (Joan N. Vickers, 1996). The quiet eye period is defined as the period during which a performer fixates on or tracks the critical object, before the initiation of a motor action (e.g. the period during which a volleyball player tracks the incoming ball before receiving a serve, the period during which a golf player fixates on the ball before putting, or the period during which a basketball player fixates on the hoop or backboard before throwing). Even though the underlying cognitive and perceptual processes are not very well understood, it is believed that during the quiet eye period, task-related cues are processed and motor plans are coordinated to successfully perform the task. Functionally, the quiet eye period allows for reorganising the neural networks responsible for movement and pre-programming of movement parameters that are required for precision psychomotor tasks. A longer quiet eye duration is a characteristic of skilled performers (Gonzalez et al., 2017).

Vickers and Adolphe (1997) studied members of the Canadian men’s national volleyball team and compared the gaze characteristics of the individual players with their yearly performance statistics. They found that players with better serve reception and pass statistics also exhibited improved tracking of the ball prior to receiving it, with minimal interference from other motor behaviours. These players had a clear and distinct quiet eye period of 432 ms, during which they quietly gathered the visual information required for their upcoming motor action. In contrast, the others did not have a clear quiet eye period. Researchers noticed similar results showing the relationship between extent of the quiet eye period and expertise in a variety of aiming and interceptive tasks, such as billiard shots (Williams, Singer, & Frehlich, 2002), golf putting (J.N.

Vickers, 1992), ice hockey goal tending (Panchuk, Vickers, & Hopkins, 2017), and basketball free throwing (Joan N. Vickers, 1996).

This has led to growing interest in the field of psychomotor training to develop smart training interventions that teach the appropriate quiet eye behaviour to trainees. Vine et al. (2014) noted that the quiet eye period is not a by-product of expertise but rather a mediator of skilful performance. Quiet eye training (QET), or training novice users to follow the quiet eye gaze patterns of an expert, is known to improve learning of psychomotor performance. Some of the QET studied used explicit verbal instructions to the participants on how to control their gaze behaviour and sometimes employed feedback sessions in which the participants viewed their own gaze data overlaid on a video (e.g. Vine & Wilson 2010). Other studies used training videos of expert users with gaze augmentation as a training aid, along with explicit verbal instructions emphasizing the critical gaze patterns of the expert that need to be followed (e.g. Vickers et al. 2017). Please read the work of Vickers (2016) for a review of the origin, typical training methodology, and current research progress in the field of QET.

Adolphe et al. (1997) employed a 6-week QET intervention for “near expert” volleyball players. They employed a comparative video feedback session during which the players could view their own gaze behaviour compared with that of two expert players in terms of four key gaze characteristics. They found significant pre-to-post improvement in the quiet eye gaze characteristics. Their results confirmed that quiet eye skills are trainable. Similarly, Vine et al. (2011) studied the benefits of QET in golf putting. The trainees watched a video comparing their gaze pattern while putting to that of an elite model. This was followed by a discussion with the trainees to cognitively probe their understanding of the gaze pattern of the expert model and the observed differences between their own gaze pattern and that of the model. They found that their lab-based QET intervention improved trainees’ putting performance, which transferred to real golf courses. In a follow-up study, they experimentally manipulated the anxiety level of trainees and found that QET offers two key advantages: resilience to anxiety and expedited rate of skill acquisition compared to the control group that did not receive gaze behaviour-specific training (Vine et al., 2011).

Other researchers found positive benefits of employing QET by showing a video and gaze of an expert in a wide variety of tasks, such as basketball free throwing (Joan N. Vickers et al., 2017), shotgun shooting (Causer, Holmes, & Williams, 2011), maritime marksmanship (Moore et al., 2014), soccer penalty taking (Wood & Wilson, 2012), laparoscopic technical skill acquisition (Wilson et al., 2011), and surgical knot tying (Causer et al., 2014).

Also, it should be noted that QET is an area of research that has been gaining a lot of interest recently and is growing rapidly. It is very likely that the strategy used to gather publications for the literature review resulted in missing a large proportion of this work. A Google Scholar search using the focused query “quiet eye training” returned 390 results, with 190 of those published since 2015. The research mentioned in this section is by no means meant to be an exhaustive review of the work on QET but rather indicative of the variety of previous work that has been conducted. Please read the work of Vickers (2007) and Vine et al. (2014) for a more thorough review of QET.

(b) Gaze Sharing in Non-Real-Time Collocated Communications

There were two publications involving non-real-time collocated communication using gaze sharing. Both studies involved gaze sharing between teacher and students. Cheng et al. (2015) developed SocialReading, a system that shares a teacher’s gaze information while he or she is reading an academic paper to students. Instead of using raw gaze point, SocialReading uses higher levels of abstraction in the visualisation by converting each paragraph into an area of interest (AOI): grey shading to visualise reading speed, border thickness to indicate frequency of rereading, and lines to indicate transition from one paragraph to another. They found that such gaze-based annotations improved the reading comprehension of the students and led to increased similarity in reading pattern between teacher and students.

Spakov et al. (2017) presented a system that supports different visualisations of the reading progress of young children during classroom reading to aid the teachers. The system supports both real-time and non-real-time visualisations, such as the raw gaze data of the students with or without a scan path, the reading progress of all students in tabular form, and a summary of students’ reading, such as average reading speed and fixation length. The system was evaluated by surveying the teachers who tried the system, and the researchers found that different visualisations serve different purposes. The teachers particularly appreciated the possibility of analysing the reading behaviour of individual students and collectively of the class, of identifying problematic words after a lesson is over, and of communicating the progress to parents.

Surprisingly, the research on gaze sharing in non-real-time collocated contexts is limited to teacher-student interactions. With the increasing popularity of wearable gaze trackers and displays, gaze sharing for non-real-time collocated communication could be an important avenue for future research and applications. For example, imagine walking through a museum and seeing the gaze representation of previous visitors presented using ambient lights or walking into a store and seeing the abstract visualisation of what other shoppers paid attention to on AR smartglasses.

(c) Gaze Sharing in Real-Time Collocated Communication

All of the publications on real-time collocated communication, except that of Spakov et al. (2016), explored scenarios in which two collocated individuals were involved in the collaboration. In contrast, Spakov et al. (2016) studied gaze sharing in a context in which the gaze of a speaker was shown to the presentation audience. Table 3 provides a summary of the seven publications with their domain of investigation.

Publication	Positioning of Collaborators	Domain of Investigation	Direction of Gaze Sharing
(Trösterer, Gärtner, et al., 2015; Trösterer, Wuchse, Döttlinger, Meschtscherjakov, & Tscheligi, 2015)	Side by side	Driver-passenger collaboration	Passenger to driver
(Zhang et al., 2017)	Side by side	Collaborative visual search on public display	Bi-directional
(Maurer, Aslan, Wuchse, Neureiter, & Tscheligi, 2015)	Side by side	Player-spectator collaboration	Spectator to gamer
(Guo & Feng, 2013)	Side by side	Parent-child shared storybook reading	Parent to child, child to parent
(Pfeuffer, Alexander, & Gellersen, 2016)	Face to face	Gaze-aware collaborative tabletop gaming	Bi-directional
(Špakov et al., 2016)	Face to face	Presentation aids for lecture	Presenter to audience

Table 3. Research on shared gaze for collocated, real-time applications

Depending on how collaborators are positioned, people working together in a collocated setting can estimate the direction of their partner's gaze by observing his or her facial orientation and eye position. When the individuals are positioned facing each other, the accuracy of perception of gaze direction is high (Cline, 1967). However, it degrades when the collaborators are side by side or positioned such that they do not see each other's faces (e.g. one user partially behind the other). Most of the previous studies explored scenarios in which two or more users were sitting or standing in front of a display (e.g. public display, gaming display, or driving simulator). Other studies explored scenarios in which

users were face to face (e.g. teacher–student in classrooms, using tabletop computers while sitting facing each other).

Trösterer et al. (2015a, 2015b) explored how sharing the gaze of a co-passenger with the driver can be useful in a collaborative navigation scenario. Using a driving simulator, they evaluated the value of sharing the gaze of a co-passenger, either continuously or after explicit activation by the co-passenger. They compared this with a baseline where the driver and co-passenger communicated verbally, without any shared gaze visualisation (Trösterer, Gärtner, et al., 2015), in a complex lane change task. Even though the collaboration between the driver and co-passenger using gaze sharing did not improve driving performance, the researchers found that it reduced the cognitive demand and perceived workload of the driver by enabling faster and more efficient communication. In another study, Trösterer et al. (2015b) compared direct visualisation of the gaze of a co-passenger on the windscreen to a more subtle visualisation using LED strips to present the horizontal position of the co-passenger’s gaze. They found that LED strips have the advantage of reduced driver distraction, at the cost of reduced accuracy and trust of the co-passenger.

Maurer et al. (2015) and Pfeuffer et al. (2016) used gaze as an input technique to interact with games in a multi-user setting: to integrate the game spectator into the game play (Maurer et al., 2015) and as an input mechanism in multiplayer tabletop games (Pfeuffer et al., 2016). Such gaze-aware multi-user applications communicate attention between players and allow for novel gameplay mechanics by requiring the partners to maintain shared attention or shift their attention in specific ways to collaboratively play the game and promote novel ways of engagement.

Zhang et al. (2017) studied the effect of bi-directional gaze sharing between collaborators and the effect of different gaze visualisations in a collaborative visual search task. They compared four different gaze visualisations (cursor, trajectory, spotlight, and highlight) with a baseline of a no shared gaze condition. They found that gaze sharing improved visual search performance and that the subtlety of gaze visualisation influences the quality of collaboration. Participants generally prefer subtle, yet visible, visualisations of gaze.

Guo and Feng (2013) studied the effect of gaze sharing between parent and child during shared storybook reading. They found that gaze sharing of parent to child, or vice versa, improved the instances of joint visual attention between parent and child. Such interventions also provided significant learning benefits to the children.

Lastly, Spakov et al. (2016) compared the value of gaze sharing during presentations. The gaze point of the presenter was overlaid on the PowerPoint presentation and shown to the audience as a tool for pointing.

They compared gaze with a conventional handheld laser pointer and mouse pointer. Overall, gaze and mouse cursor were noticed faster than the handheld laser pointer.

In summary, gaze sharing can be useful for real-time communication, even when individuals are collocated and can potentially naturally perceive the direction of their partner's from his or her face. Augmenting gaze information on the shared visual content enables more intuitive and accurate awareness of attention, even in a cognitively challenging situation such as driving.

(d) Gaze Sharing in Real-Time Remote Communication

The focus of this thesis is on real-time remote communication; thus, the previous work in this category is the most relevant to this thesis.

Real-time gaze sharing in video-based remote communications imposes two main challenges on the usability of a shared gaze when compared to non-real-time use cases. First, typical video communication over the Internet in the current state of technology can introduce a delay of several hundred milliseconds after a visual or gaze event has occurred until it is perceived at the remote end (Berndtsson, Folkesson, & Kulyk, 2012). The delay could occur due to a wide variety of issues, such as video compression, transmission of the data over the Internet, delay incurred due to image acquisition and processing by the gaze tracker, or as a result of the processing and visualisation of the information at the receiving end.

Second, the robustness and accuracy of gaze tracking can be an important factor that influences the use of a gaze pointer. In non-real-time use cases, we can perform post-calibration of gaze data to correct the possible inaccuracies in tracking and ensure that the technology works reliably. However, this is not always possible when the gaze of the user is transferred in real time to the remote participant.

Previous studies have used several methods to avoid these two challenges. First, the possible delay in video and gaze sharing is reduced by experimentally evaluating the value of gaze sharing in controlled lab setups that limit latency in gaze transfer (e.g. by having the collaborators in the same physical location or using a dedicated high-speed local area network for data sharing). Second, the issue with gaze-tracking accuracy is often tackled by calibrating the user multiple times or, in some cases, excluding the "bad" gaze data from the analysis.

Task	Publications
Joint construction	(E. G. Bard, Hill, Foster, & Arai, 2014; Carletta et al., 2010; D'Angelo & Gergle, 2016; Harrer, Schlosser, Schlieker-Steens, & Kienle, 2015; C. Liu, Kay, & Chai, 2011; Müller, Helmert,

	Pannasch, & Velichkovsky, 2013; Schlösser, Schlieker-steens, & Kienle, 2015; B. M. Velichkovsky, 1995) (Akkil & Isokoski, 2019; Akkil, James, Isokoski, & Kangas, 2016; Akkil, Thankachan, & Isokoski, 2018; Billinghamurst et al., 2017; S. R. Fussell et al., 2003; Gupta, Lee, & Billinghamurst, 2016; Higuch, Yonetani, & Sato, 2016)
Visual search and consensus	(Brennan et al. 2008; Neider et al. 2010; Wahn et al. 2016; McDonnell et al. 2017; Messmer et al. 2017; Yamani et al. 2017; D'Angelo & Gergle 2018)
Computer gaming	(Lankes, Maurer, & Stiglbauer, 2016; Lankes, Rammer, & Maurer, 2017; Maurer, Lankes, Stiglbauer, & Tscheligi, 2014; Newn, 2018; Newn, Velloso, Allison, Abdelrahman, & Vetere, 2017)
Video and text communication	(Roberts et al., 2009; Schlösser, Schröder, Cedli, & Kienle, 2018; Shikida, 2016)
Spatial referencing	(Akkil & Isokoski, 2016a; Duchowski et al., 2004)
Computer programming	(Bednarik & Shipilov, 2011; D'Angelo & Begel, 2017)
Trip planning	(Qvarfordt et al., 2005)
Collaborative learning	(Schneider & Pea, 2013)
Collaborative navigation	(Akkil & Isokoski, 2016b)

Table 4. Overview of previous studies based on the context of evaluation

Table 4 presents a summary of the previous studies based on the task used. Fifteen of the previous publications explored the value of gaze sharing to facilitate remote guidance (e.g. an expert user guiding a novice worker) in tasks involving arrangement, assembly, and repair of objects, collectively categorised as joint construction. In addition, there were seven publications in the area of visual search (i.e. two or more collaborators looking for a specific object in the shared visual field) and five publications in the domain of computer gaming.

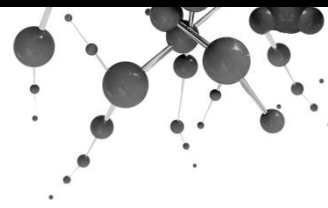
An important factor to note here is that the design of the shared gaze interface and criteria for evaluating the success of gaze sharing for different tasks are not the same. For example, in the context of joint construction, an important function of gaze sharing is to enable grounding and improve efficiency of communication. The purpose of gaze sharing in

the domain of computer game streaming might be to communicate the cognitive processes of players to spectators and increase engagement of the spectators. Thus, a gaze cue that provides a fine level of information about the visual strategy of the player might be appreciated in such scenarios. On the other hand, in cognitively challenging tasks such as collaborative learning, gaze cues that are even only slightly distracting could be detrimental to the activity. The potential benefits and limitations of gaze sharing are dependent on the collaborative task.

In Chapter 5, I present a more in-depth analysis of the previous work on gaze sharing in real-time remote communication.

Summary of the chapter

- *Gaze sharing for (real-time and non-real-time) collocated applications is a potential avenue for future research.*
- *Gaze sharing for psychomotor training is an area that is gaining increasing research interest.*
- *Gaze sharing for real-time applications presents two additional challenges related to delay in gaze transfer and quality of gaze tracking.*



5 Shared Gaze in Real-Time Remote Collaboration

The focus of this thesis is on gaze sharing in real-time remote video-based collaboration. In this chapter, I present a more detailed analysis and taxonomy of the previous studies in this area, followed by analysing the benefits and limitations of gaze sharing highlighted in the literature.

5.1 CLASSIFICATION OF PREVIOUS STUDIES ON SHARED GAZE

There are multiple factors that can be used to classify previous literature on gaze sharing in real-time remote collaboration. From the perspective of the thesis, four factors are specifically interesting: i) characteristics of the task, ii) symmetry of collaborator roles, iii) type of gaze visualisation used, and iv) level of awareness of gaze sharing. The categories I used for the classification are shown in Figure 6 and described in more detail below.

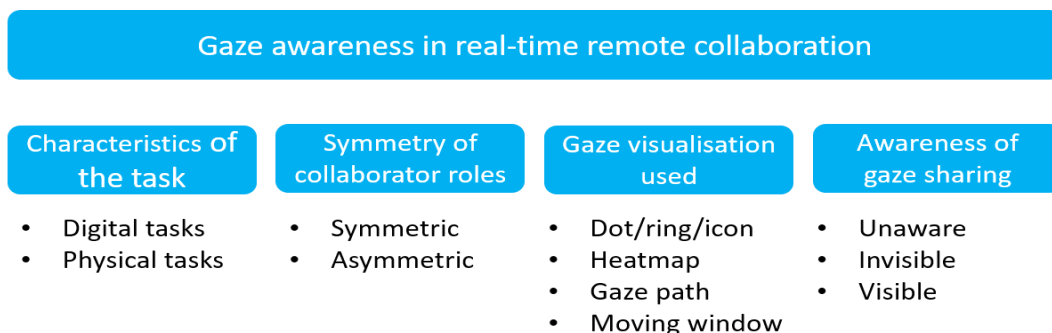


Figure 6. Factors used in the classification of the literature on shared gaze in real-time remote collaboration

Characteristics of the Task

There are multiple ways of classifying a task based on its characteristics (e.g. based on its cognitive demands, based on how well defined the task is). From the perspective of literature on gaze sharing for remote collaboration, an important distinction must be made between digital tasks and physical tasks.

Digital tasks are tasks performed exclusively on a 2D computer display (e.g. pair programming). Collaboration to accomplish a digital task normally involves sharing the screen between individuals so that both collaborators have a consistent and full view of the desktop screen. Tasks that require explicit user interaction are accomplished by using the mouse or touching the screen to act on the virtual objects. From an interaction mechanics point of view, the vast majority of digital tasks that are performed on a computer display are pointing-intensive interactions (e.g. menu navigation, clicking hyperlinks). Also, typically, the screen-sharing software available enables sharing control of the mouse cursor. This means that the task of performing the required interactions can be delegated to the remote collaboration partner. Figure 7 shows a typical shared display collaboration setup involving two collaborators.

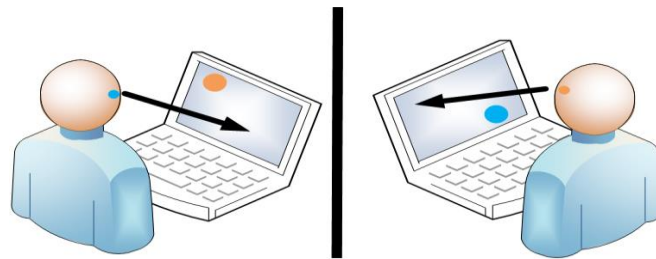


Figure 7. A typical shared display collaboration setup involving two users. The users have their display shared and can see the gaze point of the collaboration partner overlaid on their screen.

Physical tasks are tasks that require actions in the 3D physical world, such as manipulating and analysing objects (e.g. operating a coffee machine). Collaborative physical tasks involve one or more individuals using a camera to show part of the physical world to the remote collaborator(s). Thus, the remote collaborator(s) might not always have a consistent, or full, view of the task space. The view provided to the remote collaborator(s) is influenced by the relative positioning of the camera, the limited field of view of the camera, and the 3D nature of the task objects. In addition, the camera arrangement used for the collaboration can be stationary or mobile (e.g. using cameras on mobile phones or smartglasses). The camera arrangement and the mobility of tasks can thus introduce additional complexities in the collaboration. Furthermore, in collaborative physical tasks that require physical manipulation, only the individual who is in physical proximity to the task objects can perform the

physical manipulations. This limitation, due to the physical nature of the task, introduces a clear asymmetry of collaborator roles.

Another key distinction between physical and digital tasks is in the task complexity and the interaction mechanics. While our interactions with digital artefacts on the computer display are largely “point and click”, interactions with the physical world involve a series of more complex 3D manipulations (e.g. turning and flipping objects) to be performed directly, using the hands or with specific tools operated using the hands (e.g. a screwdriver). It is easy to gauge the difference in the interactions using an example. Imagine a common interaction in the physical world, that of assembling a new item of furniture. This physical task requires the operator to first locate the right blocks, followed by precisely orienting the blocks and aligning them relative to one other. While holding the joined structure together, the operator needs to locate the appropriate screw and screwdriver. Next, without removing his or her hand used to hold the structure together, the operator tightens the screw using the screwdriver, making a clockwise movement of the hand, while exerting enough force inwards. Physical tasks are more complex than the typical “point and click” interactions involved with virtual objects on a 2D computer screen.

Most of the previous work on shared gaze interfaces studied collaborative digital tasks. In contrast, the research focus on physical tasks is relatively new. Except for the earliest study by Fussell et al. (2003), all the work on gaze awareness in collaborative physical tasks was published after 2014. See Table 5 for an overview.

Characteristics of the task	Publications
Digital task	(Velichkovsky 1995; Qvarfordt et al. 2005; Cherubini et al. 2008; Neider et al. 2010; Bednarik & Shipilov 2011; Liu et al. 2011; Müller et al. 2011, 2013, 2014; Schneider & Pea 2013, 2014; Maurer et al. 2014a; John et al. 2014; Schlösser et al. 2015, 2018; Harrer et al. 2015; Wahn et al. 2016; D’Angelo & Gergle 2016, 2018; Lankes et al. 2016, 2017; Li et al. 2016; Newn et al. 2017, 2018; Niehorster et al. 2017; D’Angelo & Begel 2017; Yamani et al. 2017; Messmer et al. 2017)
Physical task	(Fussell et al. 2003; Akkil & Isokoski 2016b, a, 2019; Akkil et al. 2016, 2018; Gupta et al. 2016; Higuch et al. 2016)

Table 5. Classification of previous work on shared gaze based on the characteristics of the task.

The purpose of this thesis is to extend knowledge on the costs and benefits of gaze sharing in collaborative physical tasks. Four out of the eight publications on collaborative physical tasks were produced as part of this thesis.

Symmetry of Roles of the Collaborator and Direction of Shared Gaze

Two friends chatting with each other using an instant messenger is a good example of an interaction with symmetry of roles. Both the users have the same sub-task, i.e., to read, process, understand the conversation so far, and respond in order to facilitate the exchange of views, opinions, or thoughts. Collaboration between individuals with a symmetry of roles means they have a comparable sub-task, extent of participation, and mental and physical effort.

On the other hand, many everyday situations involve collaboration between individuals who do not have symmetry of roles (e.g., a remote expert teaching a novice to perform a task). Multiple people working together to accomplish a task may have different visual environments, activities within the task, knowledge of the task at hand, or abilities for performing the actions to accomplish the task. Asymmetries in collaborator roles may mean that even though two individuals are collaborating to accomplish a common task, they are involved in different activities, leading to different mental and physical efforts. Schneider and Pea (2013) noted that even in collaboration with theoretically symmetrical roles, asymmetry may emerge as the collaboration progresses because the collaborators may show different levels of interest and initiative in the task.

In terms of directionality of shared gaze, there are different ways of implementing a shared-gaze collaborative system. Gaze can be shared in one direction, from a specific collaborator to others (e.g. from an expert to the novice), or the gaze of every collaborator can be broadcast to others (e.g. three users performing a collaborative search with each other's shared gaze).

The symmetry of collaborators' roles often influences the directionality of shared gaze. Asymmetry of roles introduces scenarios where sharing the gaze of one of the collaborators may be more beneficial, more relevant to the collaboration (e.g. a teacher gaze sharing with students or a game player with a game viewer), or technically easier (e.g. a desktop computer user collaborating with a mobile phone user).

Previous research on shared gaze involving asymmetrical roles has focused on the asymmetries introduced due to the knowledge possessed by the collaborators, e.g., expert and novice (Akkil et al., 2016; B. M. Velichkovsky, 1995), asymmetries due to different visual environments, e.g. collaborative visual search using a gaze contingent moving window (McDonnell et al., 2017; Müller et al., 2014), asymmetries as a result of different abilities for performing actions on the task objects, e.g. game player and viewer (Lankes et al., 2017), asymmetries induced due to the medium of collaboration, e.g. a desktop computer user collaborating with a mobile phone user (Akkil et al., 2018), or asymmetries due to their

different combinations, e.g. a remote expert guiding a field worker through mobile video communication to accomplish a physical task introduces asymmetry of knowledge, visual environment, and abilities (Akkil et al., 2018).

Table 6 shows an overview of previous publications based on the symmetry of roles and direction of shared gaze. A clear pattern emerged from the analysis of previous work; i.e. all the previous studies in real-time remote collaboration involving symmetrical collaborator roles used multidirectional gaze sharing, while the majority of the studies involving asymmetrical collaborator roles studied the value of shared gaze in a unidirectional context.

Collaborator roles	Direction of shared gaze	Publications
Asymmetrical	Unidirectional	(Bednarik & Shipilov, 2011; Foulsham & Lock, 2015; C. Liu et al., 2011; McDonnell et al., 2017; Müller et al., 2014, 2013; Newn et al., 2017; Qvarfordt et al., 2005; Shikida, 2016; B. M. Velichkovsky, 1995) (Akkil & Isokoski, 2016b, 2016a, 2019; Akkil et al., 2016, 2018; S. R. Fussell et al., 2003; Gupta et al., 2016; Higuch et al., 2016; Newn et al., 2017)
	Bidirectional/ multi-directional	(E. G. Bard et al., 2014; E. Bard, Hill, Arai, & Foster, 2009; Duchowski et al., 2004; Lankes et al., 2017)
Symmetrical	Unidirectional	None
	Bidirectional/ multi-directional	(Siirtola et al.; Vertegaal 1999; Brennan et al. 2008; Neider et al. 2010; Schneider & Pea 2013, 2014; Bard et al. 2014; John et al. 2014; Maurer et al. 2014a; Schlösser et al. 2015, 2018; Harrer et al. 2015; Wahn et al. 2016; D'Angelo & Gergle 2016; Lankes et al. 2016; Niehorster et al. 2017; Yamani et al. 2017; D'Angelo & Begel 2017; Messmer et al. 2017; D'Angelo & Gergle 2018; Newn et al. 2018)

Table 6. Classification of previous research on shared gaze based on collaborator roles

Video-based remote collaboration to accomplish physical tasks introduces a clear asymmetry of collaborator roles. All the work reported in the thesis involved unidirectional sharing of gaze information. In Studies III and IV, gaze of the person performing the physical task was shared to the remote collaborator. In contrast, in Studies V and VI, gaze of the remote user was shared to the collaborator performing the physical task.

Visualisation of Gaze Information

There are different ways of visualising the gaze information of a collaborator in relation to the shared visual space. Table 7 summarises gaze visualisations used in previous literature and Figure 8 presents an example for each. The most common visualisation technique is to present the current gaze position as a cursor, i.e., an abstract visual element such as a semi-transparent dot (e.g., (Akkil et al., 2018; Qvarfordt et al., 2005), ring (e.g., Brennan et al. 2008; Neider et al. 2010), crosshair (e.g., (Yamani et al., 2017), or icon (e.g., (D'Angelo & Gergle, 2016; Müller et al., 2014). A cursor visualisation has multiple advantages. First, it is relatively simple to implement because it directly visualises the gaze point returned by the tracker and does not require complex processing of the historical gaze data or separating different eye movements, such as fixation and saccades. Second, it is very flexible to use because it can be used with any simple gaze-smoothing technique to make the gaze cursor more or less responsive according to the task's requirements. Third, the cursor visualisation works for any on screen content or task without the need for task- or content-specific fine tuning.

Visualisation used	Publications
Dot/ring/icon/crosshair	(E. Bard et al., 2009; S. E. Brennan et al., 2008; D'Angelo & Gergle, 2016; Lankes et al., 2016, 2017; C. Liu et al., 2011; Müller et al., 2013; Neider et al., 2010; Qvarfordt et al., 2005; Schneider & Pea, 2014; B. M. Velichkovsky, 1995) (Bednarik & Shipilov, 2012; Duchowski et al., 2004; Foulsham & Lock, 2015; C. Liu et al., 2011; Messmer et al., 2017; Newn et al., 2017; Schlösser et al., 2015; Siirtola et al., 2019; B. M. Velichkovsky, 1995; Vertegaal, 1999; Yamani et al., 2017) (Akkil & Isokoski, 2016b, 2016a, 2019; Akkil et al., 2016, 2018; S. R. Fussell et al., 2003; Gupta et al., 2016; Higuch et al., 2016)
Moving window or variants	(McDonnell et al., 2017; Müller et al., 2014)
Heat map or variants	(John et al. 2014; Newn et al. 2017; D'Angelo & Gergle 2018)
Scan path	(Newn et al. 2017; D'Angelo & Gergle 2018)
Area of interest	(Harrer et al., 2015; Newn et al., 2017; Schlösser et al., 2015)

Shared attention area	(Lankes et al. 2017; D'Angelo & Gergle 2018)
Task/screen content-specific visualisation (subtle highlighting of line)	(D'Angelo & Begel, 2017; Schlösser et al., 2018)
Audio and vibrotactile feedback	(Wahn et al., 2016)

Table 7. Different shared-gaze visualisations used in literature

Previous literature also utilised other visualisations such as heat maps, scanpaths, and shared attention area, albeit rarely. More recently, non-visual modalities, such as vibrotactile feedback, have been proposed as feasible feedback modalities for gaze in remote collaboration (Wahn et al., 2016). See Figure 9 for example visualisations.

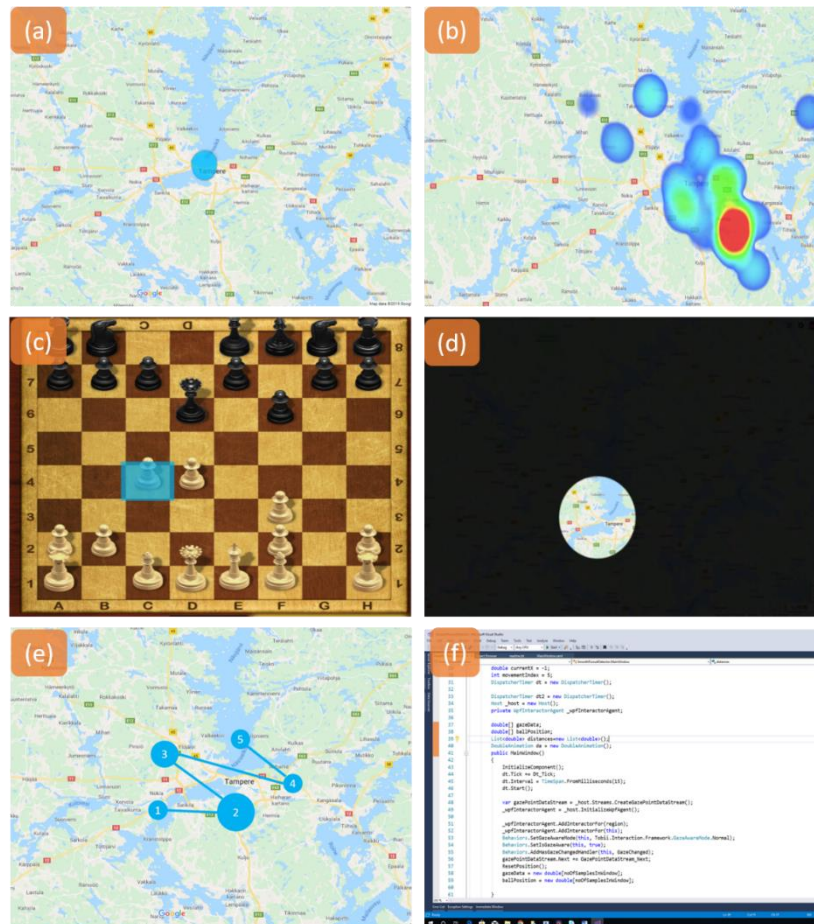


Figure 8. Example visualisations used in previous studies: (a) dot, (b) heat map, (c) area of interest if visualisation is based on gaze of one of the collaborators and shared attention area if visualisation is triggered upon joint attention, (d) moving window paradigm, (e) scanpath visualisation, and (f) content-specific visualisation, in which the coloured line highlighting to the left of the screen indicates the region of attention of the collaborator

Previous studies have also focused on comparing different gaze visualisations for remote (D'Angelo & Gergle, 2018; Harrer et al., 2015; Newn et al., 2017; Schlösser et al., 2015) and collocated (Špakov et al., 2016; Trösterer, Wuchse, et al., 2015; Zhang et al., 2017) collaborations. The different gaze visualisations such as cursor, heat map, AOI, shared attention area, and scanpath communicate different information to the viewer, have different levels of conspicuousness and distraction, and afford different interpretations (D'Angelo & Gergle, 2018). For example, a heat map visualisation enables easy interpretation of historical gaze data (i.e. did the collaborator look at a specific AOI in the last few seconds?). In contrast, a dot representation of gaze shows only the current gaze point. Though at the cost of higher cognitive effort, a viewer may still be able to interpret whether the collaborator recently viewed a certain area. On the other hand, the current gaze point may be less visible in a cumulative heat map visualisation.

The challenge in visualizing the gaze point is to understand the gaze information that is most relevant to a given task and communicate it such that it enables easy interpretation with minimal cognitive effort. Previous research has noted the challenge of balancing the visibility, visual information, and distraction of gaze markers in remote collaboration (Newn et al., 2017). Zhang et al. (2017) studied four gaze visualisations (cursor, trajectory, spotlight, and highlight) in a collocated collaborative visual search. They found that the subtle visualisation of gaze often leads to reduced efficiency in completing a task. On the other hand, prominent visualisation may be distracting.

It is evident from the previous work that there is not one gaze visualisation that fits all scenarios. The task and context should determine the gaze visualisation. For example, Newn et al. (2017) compared nine gaze visualisations in a competitive game setting. They found that heat map visualisation was the most preferred and efficient visualisation for enabling intention prediction. Heat map visualisation maintains the gaze information for a small amount of time, allowing viewers to comfortably gather the recent historical gaze points and infer the intention. On the other hand, D'Angelo and Gergle (2018) compared three gaze visualisations (heat map, scanpath, and shared attention area) for a collaborative search and consensus task. They found that heat map visualisation was the least useful and least subjectively preferred. The heat map was also considered the most distracting. With heat map visualisation, the current gaze point is not very prominent because the visualisation takes into account previous gaze points within a specific time window. Current gaze location is often important in tasks where gaze is used for explicit deictic referencing. Another drawback of heat map visualisation is that it can occlude the task space.

Interestingly, the cursor visualisation of gaze performed moderately well compared to other visualisations in intention prediction (Newn et al., 2017), collocated (Zhang et al., 2017), and (D'Angelo & Gergle, 2018) remote visual search tasks. This indicates the flexibility of use and interpretation the relatively simple cursor visualisation provides. On the other hand, cursor visualisation can be potentially distracting due to the “jumpy” and “jittery” movements of the cursor. Thus, consideration should be given to smoothening the gaze data before presentation (Akkil et al., 2016; D'Angelo & Gergle, 2016; Qvarfordt & Zhai, 2005) and identifying situations where the gaze visualisation could be useful or distracting to automatically enable or disable the visualisation.

Four key conclusions based on the analysis of gaze visualisations in mediated collaboration are as follows:

- Cursor-based visualisation is the most commonly used gaze visualisation in mediated communication.
- The visualisation used can influence the collaboration performance and benefit of gaze awareness.
- There is not one gaze visualisation that is best suited for all tasks and contexts.
- Cursor-based visualisation is a simple and flexible visualisation that performs moderately well for different tasks and contexts.

The work in this thesis used cursor-based visualisation of shared gaze.

Level of Awareness of Gaze Sharing

In remote collaboration, a shared-gaze cursor provides two different utilities. It can function as an explicit communication mechanism between the collaborators, and it can function as an implicit information channel when the eye movements are also task relevant. An example of the explicit use of gaze is to use the gaze cursor as a spatial pointer in the communication (e.g. “*Place the object here*” while staring at a spot). Brennan et al. (2008b) demonstrated that gaze pattern that is naturally produced as part of performing a task, as opposed to explicitly produced for communicating, can be beneficial in remote collaboration. In their collaborative visual search task, collaborators could covertly attend to the gaze of their partner and allocate their own attention based on an “I look where you are not looking” strategy.

Brennan et al. (2012) presented a differentiation between explicit/communicative and implicit/informative signalling. They note that for a signal to be explicit, it needs to have three characteristics: First, the signal must be informative, second, the signal receiver must be able to perceive and process it, and third, the signal must be produced with the

intention to communicate. On the other hand, an implicit/informative signal may naturally contain information the receiver can perceive and process. However, it is not produced with the intention of communicating.

Studying implicit and explicit use of gaze in remote collaboration can be difficult, as it requires understanding the user's intention. An easier way to study the difference between the implicit and explicit use of gaze is to manipulate the collaborators' awareness regarding shared gaze. When the producer of the gaze is not aware of gaze sharing, all of the eye movements are naturally occurring and are produced as part of the task, without the producer's explicit intention to communicate using his eyes.

On the other hand, though the producer of the gaze is aware that his gaze is being shared with the collaboration partner, this does not mean all eye movements are explicitly produced with the intention to communicate. The collaborator may manipulate his natural gaze behaviour to be less confusing (e.g. avoid looking at certain places) or use it explicitly to communicate. However, they may still exhibit eye movements necessary for perception (e.g. search for an item). The awareness that his gaze is being shared and the absence of other remote gesturing mechanisms would simply increase the probability of the collaborator using gaze as an explicit channel of communication.

In addition to the awareness of shared gaze, another aspect that can potentially influence the extent and accuracy of use of the gaze cursor is whether the producer of the gaze can see his own gaze point, which is being transferred to the collaborator. The direct feedback of one's own gaze can theoretically help in multiple ways. First, it enables the producer of the gaze to be aware of his eye movements, allowing him to proactively correct them when they could be potentially misleading. Second, it allows the producer of the gaze to be aware of gaze-tracking accuracy and proactively work to overcome any offset that may exist (e.g., recalibrate the tracker, look slightly away so that the gaze cursor is on target, or adjust the tracker verbally). Third, it may allow the producer of the gaze to use the channel for collaboration more confidently because he can see the exact gaze point that his partner can also see.

In certain collaborative contexts, seeing one's own gaze point may be unavoidable. For example, Zhang et al. (2017) studied collaborative visual search among co-located users on a large display. In such cases, because both users share the same display, gaze sharing between participants would mean that the producer of the gaze also views his own gaze. Similarly, Akkil et al. (2016a) and Higuch et al. (Higuch et al., 2016) studied a remote collaboration setup in which a collaborator's gaze was physically projected to the partner's task space. The physical projection of gaze is thus also visible in the shared visual space for the gaze's producer. Furthermore, in certain collaborative contexts, seeing one's own gaze

point may be a solution to avoid the potential privacy issues associated with gaze sharing (Vertegaal, 1999).

Previous research on shared gaze can be categorised according to the gaze producer's level of awareness regarding the gaze sharing. Figure 9 shows an illustration of the three classifications in the context of a one-directional shared display remote collaboration, and Table 8 shows the previous studies based on the level of awareness of the producer of the gaze regarding gaze sharing. Some of the studies did not mention in the paper if the producer of the gaze could see his own gaze point (e.g., (Bednarik & Shipilov, 2012; Niehorster et al., 2017)). These studies were excluded from analysis. Some studies presented here were done in two phases (e.g., (Akkil & Isokoski, 2016b, 2016a; Foulsham & Lock, 2015; Newn, 2018)) and did not involve real-time collaboration. In such cases, the awareness of shared gaze was determined, not based on whether the producer of the gaze was aware of gaze tracking but whether he was aware of the task for subsequent participants who would view the video. Without knowing the task for subsequent participants, it is unlikely that they would have altered their natural eye movements to explicitly communicate to the viewer.

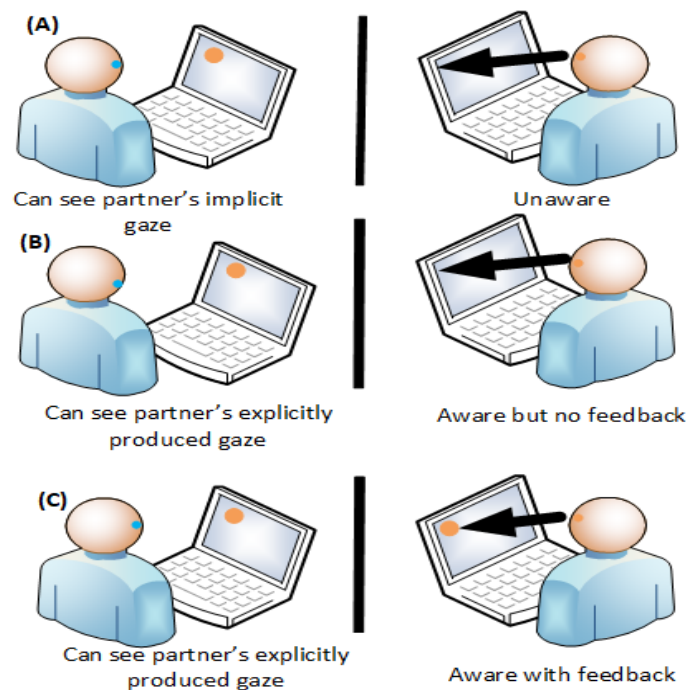


Figure 9. Three gaze-sharing configurations based on the level of awareness of the gaze producer. (A) The collaborator on the right is not aware of gaze sharing. (B) The collaborator on the right is aware of their shared gaze but cannot see the gaze point. (C) The collaborator on the right is aware of their shared gaze and can also see the gaze point that is transferred. In all cases, the collaborator on the left can see the shared gaze

Level of awareness of shared gaze	Publications
Not aware	(Akkil et al., 2018; C. Liu et al., 2011; Newn et al., 2018, 2017; Qvarfordt et al., 2005)
Aware but no direct feedback	(E. G. Bard et al., 2014; S. E. Brennan et al., 2008; D'Angelo & Begel, 2017; D'Angelo & Gergle, 2016; Lankes et al., 2016, 2017; Messmer et al., 2017; Müller et al., 2013; Neider et al., 2010; B. M. Velichkovsky, 1995; B. Velichkovsky, Pomplun, & Rieser, 1996; Yamani et al., 2017)
Aware and received direct feedback	(Akkil & Isokoski, 2019; Akkil et al., 2016; Carletta et al., 2010; Duchowski et al., 2004; Harrer et al., 2015; Higuch et al., 2016; McDonnell et al., 2017; Müller et al., 2014; Schlösser et al., 2015)

Table 8. Classification of shared-gaze research based on level of awareness of shared gaze

Results from previous publications indicated that shared gaze in all three configurations provides benefits to collaboration. For example, Qvarfordt et al. (2005) found that implicitly produced gaze can be useful in spatial referencing, aid topic switching, and reducing ambiguity in communication. Others found that naturally produced eye movements can help enable referential grounding (C. Liu et al., 2011) and prediction of intention (Akkil & Isokoski, 2016b; Newn, 2018). Similarly, numerous publications have found use of gaze as an explicit pointing mechanism (D'Angelo & Gergle, 2016; Neider et al., 2010).

The cost and benefits of seeing one's own gaze point is unclear from the previous publications. Seeing one's own gaze point can be potentially distracting, especially when gaze tracking is not accurate or when there is a delay in updating the visuals. D'Angelo and Gergle (2016) noted that showing one's own gaze point "can produce a feedback loop that causes people to follow their own cursor". On the other hand, Maurer et al. (2014a) studied shared gaze in cooperative online gaming. Their participants commented that they would like to see their own gaze point visualised along with the partners.

One of the publications (Study VI) in this thesis is specifically designed to answer this question: Does the level of the gaze producer's awareness regarding the gaze sharing influence the utility of shared gaze? We compared the three gaze-sharing configurations in a mobile, video-based, collaborative physical task.

5.2 TASK COUPLING

In addition to the four factors used to classify the previous work on shared gaze, task coupling is another aspect that influences the utility of shared gaze for collaboration.

Task coupling is a measure of dependencies between the collaborators and refers to the extent of communication required to accomplish the task (G. M. Olson & Olson, 2000). A parallel visual search where both collaborators can independently search for the object with minimal coordination with each other is an example of a loosely coupled task. On the other hand, a similar visual search task with an additional consensus phase where both collaborators need to locate the object and reach a consensus introduces additional dependencies and a coordinated effort. The collaborator who first finds the object needs to coordinate its location with the partner to jointly arrive at a decision. The visual search and consensus task is more tightly coupled than the parallel visual search.

A loosely coupled task requires less frequent or less complex interaction between the collaborators. Likewise, a tightly coupled task may require more frequent or more complex interactions (Neale, Carroll, & Rosson, 2004; G. M. Olson & Olson, 2000). Task coupling is a measure that is inherently associated with the nature of the task and may be influenced by other contextual factors such as the collaborators' familiarity with the task, task-specific common knowledge shared by the collaborators, and even familiarity between the collaborators (G. M. Olson & Olson, 2000).

Previous studies have experimentally manipulated the task coupling to understand how the value of task coupling in loosely coupled and tightly coupled tasks influences shared gaze (D'Angelo & Gergle, 2018; Müller et al., 2013). Muller et al. (2013a) used the term "task autonomy" instead of task coupling. In their work, "low autonomy" and "high autonomy" were indicative of tight coupling and loose coupling, respectively.

The previous literature discusses the extent to which task coupling can influence the value of shared gaze, with more tightly coupled tasks benefitting more from shared gaze than loosely coupled tasks (D'Angelo & Gergle, 2018). This is in line with the more general understanding that a tightly coupled remote task which requires more complex and frequent interactions between the collaborators may require more sophisticated awareness mechanisms that provide the collaborators with the relevant contextual cues (Neale et al., 2004). Shared gaze may not provide any substantial benefit in loosely coupled tasks. Furthermore, Muller et al. (2013) showed that gaze sharing in loosely coupled tasks may be counterproductive. When collaborators can work on the tasks independently, showing the partner's gaze can be distracting and detrimental to the efficiency of collaboration.

While gaze may be generally more useful in tightly coupled collaborative tasks than loosely coupled tasks, one should also be careful about generalizing this result to all types of tasks. For example, Muller et al. (2014) studied the value of shared gaze in a tightly coupled “moving window paradigm”-style hierarchical decision-making task and found that gaze sharing was not as useful as sharing a mouse. This highlights the fact that not all tightly coupled tasks may show clear benefits of shared gaze. Task characteristics beyond simple task coupling, such as how grounding takes places in the specific collaborative context need to be considered before applying shared gaze for remote collaboration.

5.3 BENEFITS OF SHARED GAZE IN REMOTE COLLABORATION

In 1995, Velichkovsky conducted the seminal study comparing gaze and mouse pointers in a collaborative construction task involving on-screen puzzle solving. He found that sharing the gaze and mouse of a remote expert can help improve the task performance and found no overall performance difference between gaze and mouse transfer for collaboration. Both gaze and mouse sharing were better than a no pointer condition. Interestingly, Velichkovsky (1995) found a noticeable difference in the rate of learning the two pointers, with gaze being faster in the initial trial and the mouse being faster in later trials. He noted that communication using gaze may be especially efficient in situations involving high complexity and low redundancy.

Many subsequent studies have evaluated shared gaze in real-time collaboration in a variety of contexts. Table 9 shows a cumulative summary of the previous findings. The purpose of this analysis is to present a high level summary of the value of shared gaze and does not take into account any difference in context and collaborator roles.

Benefit of shared gaze	Publications
Spatial/deictic referencing	(Akkil & Isokoski, 2016a; Akkil et al., 2016; E. Bard et al., 2009; Cherubini et al., 2008; D’Angelo & Gergle, 2016, 2018; Gupta et al., 2016; Higuch et al., 2016; Maurer, Trösterer, et al., 2014; Müller et al., 2013; Neider et al., 2010; Qvarfordt et al., 2005; Schneider & Pea, 2013; Špakov et al., 2016; Trösterer, Gärtner, et al., 2015; van Rheden, Maurer, Smit, Murer, & Tscheligi, 2017; B. M. Velichkovsky, 1995)
Establishing joint attention	(Akkil & Isokoski, 2016b; D’Angelo & Begel, 2017; Guo & Feng, 2013; Harrer et al., 2015; Qvarfordt et al., 2005; Schneider & Pea, 2013; B. M. Velichkovsky, 1995; Zhang et al., 2017)
Enabling grounding	(S. E. Brennan et al., 2008; John et al., 2014; C. Liu et al., 2011; Siirtola et al., 2019; B. M. Velichkovsky, 1995;

	Zhang et al., 2017)
Increasing task engagement/enjoyment	(Akkil & Isokoski, 2019; Akkil et al., 2016, 2018; Gupta et al., 2016; Lankes et al., 2016, 2017; Maurer et al., 2015; Pfeuffer et al., 2016; Zhang et al., 2017)
Increasing redundancy and reduced ambiguity in communication	(Akkil & Isokoski, 2019; Akkil et al., 2016, 2018; Gupta et al., 2016; Higuch et al., 2016; Maurer, Trösterer, et al., 2014; Qvarfordt et al., 2005)
Communicating interest, preference, and intention	(Akkil & Isokoski, 2016b; Foulsham & Lock, 2015; Higuch et al., 2016; Newn et al., 2018, 2017; Qvarfordt et al., 2005)
Increasing understanding of collaborators' task status	(Akkil & Isokoski, 2016a, 2019; Akkil et al., 2016; Qvarfordt et al., 2005; Schlösser et al., 2018; Zhang et al., 2017)
Improving feeling of presence	(Akkil & Isokoski, 2019; Akkil et al., 2016; Gupta et al., 2016; Lankes et al., 2016, 2017)
Improving perceived quality of collaboration	(Akkil & Isokoski, 2019; Akkil et al., 2016, 2018; Gupta et al., 2016; Higuch et al., 2016)
Improving confidence in communication	(Akkil & Isokoski, 2016b, 2016a; Akkil et al., 2016; Qvarfordt et al., 2005)
Enabling learning	(Guo & Feng, 2013; Harrer et al., 2015)
Coordinating efficiently in time-critical tasks	(S. E. Brennan et al., 2008; Neider et al., 2010)
Allowing consistency of use	(Akkil et al., 2018)

Table 9. Benefits of shared gaze

Furthermore, as a consequence of these benefits, many studies have also reported improved efficiency of task performance. For example, Brennan et al. (S. E. Brennan et al., 2008) reported improved task completion time in collaborative visual search tasks. On the other hand, D'Angelo and Gergle (2016) reported that while gaze sharing was helpful in accurately referring to linguistically complex objects, it did not improve task completion time. This suggests that the benefits of shared gaze do not always translate to performance improvement.

Analysis of the reported benefits of shared gaze highlights four important points:

- A relatively simple intervention such as showing where a collaborator is looking can provide numerous benefits to the collaboration.

- Shared gaze benefits both objective (e.g. increased deictic referencing) and subjective (e.g. improved perceived quality of collaboration) aspects of the collaboration.
- While gaze provides many different benefits in collaboration, one of the most commonly reported benefits of shared gaze is its use as a pointer for explicit deictic referencing.
- The collaborative task and the visualisation of the gaze cursor influence the benefits of shared gaze. A fine-level analysis of the task characteristics and designing the visualisation to intuitively present the eye movement pattern relevant for the task ensure the benefits of shared gaze.

5.4 LIMITATIONS OF SHARED GAZE IN REMOTE COLLABORATION

Gaze sharing between collaborators can provide numerous benefits, as reported in the earlier section. This brings us to the other pertinent question: What are the limitations of shared gaze? Table 10 presents a summary of previous studies based on the limitation of shared gaze that they highlighted. Some of the limitations may be overcome to a certain extent by a better design of the gaze-sharing system (e.g., better visualisation can reduce the distraction caused by shared gaze), and all the limitations may not be intrinsic to the concept of gaze sharing.

Limitations	Publications
Visualisation of the gaze cursor (fast and jittery) can be distracting	(Akkil et al., 2018; Bednarik & Shipilov, 2012; D'Angelo & Gergle, 2016; Lankes et al., 2017; Newn et al., 2017; Trösterer, Gärtner, et al., 2015; Trösterer, Wuchse, et al., 2015; Zhang et al., 2017)
Accuracy of tracking can influence value of shared gaze	(Akkil & Isokoski, 2019; Akkil et al., 2018; D'Angelo & Gergle, 2016; van Rheden et al., 2017; Zhang et al., 2017)
Not as flexible as an explicit gesturing mechanism	(Akkil & Isokoski, 2019; Akkil et al., 2018; Higuch et al., 2016; Müller et al., 2014, 2013)
Can be ambiguous (Midas touch) and thus complicate grounding	(Akkil & Isokoski, 2019; Akkil et al., 2018; Müller et al., 2011, 2013)
Potential privacy issues	(Akkil et al., 2016; van Rheden et al., 2017; Zhang et al., 2017)
Without shared visual context, shared gaze may not be useful	(C. Liu et al., 2011; Müller et al., 2014)
Can cause eye strain	(Akkil et al., 2018)

Table 10. Limitations of shared gaze

Need for Shared Visual Context

Liu et al. (2011) studied the usefulness of gaze as a means to support collaboration in an object arrangement task. One of the participants (an expert) knew the arrangement to make, while only the remote partner could act on the objects. The expert's gaze was transferred to the remote partner. They found that gaze transfer improved collaboration when both partners had the exact view of the task space. In comparison, gaze was less helpful when the collaborators had a mismatched view of the shared space. Similarly, Muller et al. (2014) noted that without a shared visual context, gaze may not be useful in collaboration. The meaning encoded in eye movement can only be interpreted when one understands the visual context in which the eye movement is made. For example, knowing that the collaboration partner looked at (x,y) position on the screen and knowing that the person looked at a specific object in that visual context affords different interpretation.

Accuracy of Gaze Tracking

The second limitation of shared gaze is associated with its accuracy. Numerous previous publications have qualitatively highlighted the fact that inaccuracies in tracking can complicate the use of the shared gaze. For example, D'Angelo and Gergle (2016) noted that when the gaze cursor is not accurate, collaborators rely on extensive verbal instructions to achieve conversational grounding. Rheden et al. (2017) noted that when gaze tracking is inaccurate, the value of shared gaze decreases.

As part of this thesis, we performed an objective analysis of how gaze-tracking accuracy influences task completion time and verbal effort (Akkil & Isokoski, 2019; Akkil et al., 2018).

Distraction of Gaze Visualisation

The visualisation of shared gaze may distract the collaborator from the task. Viewers of gaze data often find gaze data to be "jittery" and "jumpy". The jitteriness of the gaze cursor is due to the low precision of gaze trackers. On the other hand, the frequent and swift saccadic motion of the eyes causes the jumpiness of the gaze cursor. The fast and frequently moving gaze cursor can often take the attention of the collaborator. Filtering the jitteriness and smoothening the jumpiness of the gaze data could reduce the distractibility associated with the gaze cursor. The distractibility associated with the gaze cursor is specifically impactful when the gaze cursor is not directly relevant to the current sub-task, e.g., in loosely coupled tasks (Müller et al., 2013).

Privacy Issues Associated with Shared Gaze

Another key and often under-discussed aspect of shared gaze is the potential privacy issues associated with it. Several factors, such as the characteristics of the scene (e.g., salience), information requirements of the current task, and characteristics of the person (e.g., personality can

modulate eye movements), modulate a person's gaze. Theeuwes et al. (1998) noted that while goal-directed eye movements are voluntary, stimulus-directed eye movements may be produced reflexively. For example, if a new object suddenly appears in the scene, people tend to look at it involuntarily. Some eye movements may not be within a person's control. Another troubling aspect of gaze allocation in the context of shared gaze is that people are not always aware of their own low-level voluntary eye movements (Kok, Aizenman, Vö, & Wolfe, 2017). When asked to report the areas where the person fixated in a previous scene, viewers often only report a small subset of the actual fixations. The disparity between the number of reported and actual fixations increases as the complexity of the scene increases (Marti, Bayet, & Dehaene, 2015).

Eye movements are considered to be a window to a person's mind because they can give deep insights into much of one's potentially private information, such as personal characteristics, emotional state, current interest, future intention, and other cognitive processes. Considering that people have neither full control nor awareness of their eye movements, it one could inadvertently share such information with a communication partner. For example, Zhang et al. (2017) reported that in a collaborative visual search, collaborators often agree upon a "divide and conquer" search strategy. When gaze is being shared, any deviation from this agreed strategy becomes evident and may signal a lack of trust between collaborators.

The privacy issues associated with shared gaze may also have several practical implications. The two most important factors are relating to user consent and user awareness. Shared-gaze systems would require explicit user permission instead of being an always ON feature. Users need to have the flexibility of being aware of the status of shared gaze and the potential to toggle the feature ON and OFF during a collaborative activity. Furthermore, there should also be awareness mechanisms that enable users to have fine-level understanding of their own eye movements (e.g., by showing their eye movements; (Vertegaal, 1999). Such awareness tools would allow users to be more aware of the gaze sharing and the possible interpretations of the meaning it communicates. It also seems likely that users may be more willing to share their gaze only in certain contexts and with certain trusted collaborators (Zhang et al., 2017). Collaborative systems thus need to provide alternative remote gesturing and awareness mechanisms so that even the user's opting out of gaze sharing has limited impact on the collaboration.

Ambiguity Associated with the Gaze Cursor

Ambiguity associated with a gaze cursor that is continually moving is evident when shared gaze is compared to a more explicit shared mouse pointer. Eighteen years after the pioneering study by Velichkovsky (1995),

Muller et al. (2013a) replicated it, and similar to the results by Velichkovsky (1995), they found that both gaze and mouse transfer lead to similar task performance. However, gaze sharing introduces ambiguity in communication and complicates grounding in spatial referencing. In case of an explicit gesturing mechanism, such as the mouse, every move communicates an intention that is relevant to the task. Collaborators can thus trust this movement for its communicative value, and use the cursor confidently, e.g. “Don’t think; just follow my mouse” (Müller et al., 2014). In contrast, not all eye movements are made with the intention to communicate. Collaborators may be less confident to use the gaze without confirming the intention behind the eye movement. This is very similar to the Midas-touch problem associated with gaze interaction.

Gaze sharing can be valuable in the absence of any other remote gesturing mechanisms. However, when compared to an explicit pointer such as the mouse, gaze induces uncertainty and increased reliance on verbal instructions to complete a task.

Limited Flexibility Compared to Other Explicit Pointing Mechanisms

The physiological constraints of how our eyes move, how visual perception functions, and the dual role that eyes play limit the flexibility gaze offers as a remote gesturing mechanism. An explicit pointer such as the mouse performs only one role, that of a communicating device. Thus, it enables a certain degree of flexibility. It can be moved quickly or slowly, depending on the task demands. It can even be moved to closely replicate visual attention to a certain degree (Müller et al., 2014).

The work done as part of this thesis extends these findings to the case of collaborative physical tasks. A mouse (or one could argue hand gestures and other explicit gesturing mechanisms) can be used to communicate complex procedural instructions by drawing shapes and representing actions (Akkil & Isokoski, 2019; Akkil et al., 2018). Such instructions make up a large part of people’s collaboration to accomplish physical tasks (S. R. Fussell et al., 2003).

One could also argue that the purpose of shared gaze is not to use it to explicitly communicate but for the implicit benefits it provides. Our results suggest that in the context of collaborative physical tasks involving the complex manipulation of objects, shared gaze may not be enough to efficiently complete the task.

Can Be a Cause of Eye Strain

The most widely identified benefit of shared gaze in previous studies is its value as an explicit gesturing mechanism for deixis. The explicit use of gaze to communicate spatial references often makes it unnatural (e.g., staring at a place for a long period) (Chitty, 2013; Kangas et al., 2014).

The work in this thesis also supports this claim. When gaze is the only shared gesturing mechanism in the context of a collaborative physical task, users tend to use gaze, to communicate complex procedural instructions (e.g. “*Turn it this way*” while trying to make a circular eye movement). Such unnatural eye movement can lead to eye strain. Many of our participants also reported eye strain from viewing their own gaze point, especially when gaze tracking was not accurate.

In summary, gaze sharing information between collaborators has its share of limitations:

- The context of use (i.e., characteristics of the task, availability of other gesturing modalities, accuracy of tracking, availability of shared visual context, etc.) influence the usefulness of shared gaze.
- The visualisation and responsiveness of the gaze representation should be designed to facilitate the task
- Accuracy of tracking is an important technical aspect that limits the usefulness of shared gaze.
- Applications designed to facilitate remote collaboration should include multiple remote gesturing mechanisms to ensure the best experience for all users in all scenarios.

5.5 HOW HAVE PREVIOUS STUDIES ADDRESSED GAZE-DATA QUALITY?

Previous studies used different gaze-tracking hardware setups (e.g., monocular vs. binocular and remote vs. head-mounted), different calibration schemes (e.g., 5-, 9-, or 15-point calibration), possibly different experimental contexts (e.g., different ambient lighting and screen brightness), and different demographics of participants. It is inevitable that they had different levels of gaze data quality.

Interestingly, no previous studies on shared gaze in remote collaboration before this thesis have reported the gaze-data quality achieved. This makes it difficult to interpret the results and compare them to other publications. The most common way of addressing the accuracy issue is to not mention it at all in the publication. Few publications report the manufacturer-provided accuracy as representative values. However, as stated earlier, this can be highly misleading.

Furthermore, studies used very subjective and subtly different approaches to address the issue of gaze-tracking quality, e.g. by calibrating the users multiple times or excluding the data from participants with “bad” gaze-data quality. Very few publications report, even qualitatively, the accuracy problems they faced and how they addressed this challenge in the research.

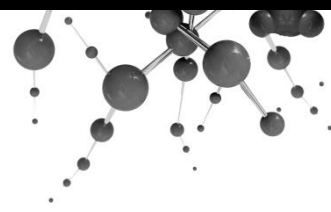
Schneider and Pea (2013) mentioned that they “cursorily watched the videos showing the participants’ gaze patterns, to ensure that no large deviation was present”. Li et al. (2016) “recalibrated as needed, in order to maintain accuracy”. D’Angelo et al. (2016) noted that they recalibrated roughly 40% of the participants. D’Angelo and Begel (2017) “relied on the participants to tell when they believed the accuracy had degraded enough to require recalibrating”. Messmer et al. (2017) reported a more objective approach, highlighting the fact that “the calibration routine was repeated until it produced no more than 1° of visual angle gaze error vertically and horizontally. Calibration was revalidated between blocks of trials”. Harrer et al. (2015) excluded data from 10% of the participants, due to “unsatisfactory gaze data”.

Gaze tracking, even in its current state-of-the-art form, is a difficult technology that does not work “satisfactorily” for all users all of the time. The results reported in many of the previous studies on shared gaze for remote collaboration may be presenting an “ideal world” view of the benefits of shared gaze and may not be representative of real-world scenarios in terms of gaze-data quality. It is difficult to conclusively say, because most of the studies do not report the gaze-data quality achieved.

Notably, this trend of overlooking gaze-data quality possibly applies to most fields of science using real-time gaze tracking as a research tool. Research that utilises gaze data for offline analysis often has the luxury of using post calibration to ensure reliable gaze data. However, studies that use gaze data in real-time would benefit from more objective reporting of gaze-data quality and grounds for recalibration/exclusion of data from analysis. One of the contributions of this thesis is to develop an easy-to-use tool to measure gaze-data quality to facilitate and encourage reporting such metrics in research publications.

Summary of the chapter

- *Characteristics of the task, symmetry of the roles of collaborators, visualisation of gaze, and level of awareness of gaze sharing are four factors of classifying literature on shared gaze for real-time collaboration.*
- *Benefits and limitations associated with shared gaze are dependent on contextual factors.*
- *Previous literature on shared gaze for real-time collaboration lacks methodological consistency in terms of gaze-data quality, making comparisons between studies more complicated.*



6 Methodology

The research reported in this thesis employed both qualitative and quantitative research methods, to gather a holistic understanding of the main focuses of the individual studies. Study II was exploratory in nature and used focus groups as the data collection method. Studies I, III, IV, V and VI used experimental research methods. Prototype systems were developed by iterative design. Each system enabled a specific use of gaze information in remote collaboration. The prototype systems were first informally evaluated in a series of pilot tests, and the learnings from the pilot tests guided further fine-tuning of the prototypes. The usability and utility of the prototypes was then evaluated in controlled user experiments.

In this chapter, I present an overview of the constructive and iterative approach followed in this research. I briefly describe the early pilot evaluations that were undertaken, the learnings from them, and how they guided the design of the six studies presented as part of this thesis. I conclude this chapter by presenting the ethical challenges associated with conducting research in the domain of shared gaze.

6.1 CONSTRUCTIVE AND ITERATIVE APPROACH

The studies reported in this thesis followed a constructive and iterative approach. In the research, I made use of off-the-shelf hardware devices, such as gaze trackers, camera modules, projectors, and mobile devices. I also involved the development of multiple software and hardware applications for facilitating the user studies and for distribution to other researchers and practitioners to utilise in their research.

The software systems were developed using either Microsoft .NET framework, or web technologies such as HTML and JavaScript. The features of the prototypes developed for conducting the user study were all guided by the focus of the study. The features of the software system that was developed for distribution (Study I) were guided by the literature and discussion with other gaze-tracking researchers at Tampere Unit for Human-Computer Interaction (TAUCHI), Tampere University, Finland.

The software systems were evaluated informally multiple times, to ensure suitability of use. This was followed by a series of pilot evaluations to fine-tune the different parameters involved.

6.2 EARLY PILOT EVALUATIONS

Before the start of this thesis, there was only one previous study, by Fussell et al. (2003), in the area of shared gaze for collaborative physical tasks. They used a head-mounted gaze-tracking system and overlaid the user's gaze on the egocentric video. This video was presented to a remote collaborator. Despite the theoretical advantage of such a system for collaboration, they did not find any measurable benefit of gaze-augmented egocentric video for collaboration. They concluded that such *"head-mounted camera systems [with gaze tracking] may not yet be robust enough for actual field applications."*

Early phase of the thesis was utilized in exploring the value of gaze-augmented egocentric videos in complex, real-world collaborative scenarios. The numerous short-term pilot evaluations that were conducted eventually led to finalising the environment and design of the six publications. Figure 11 shows a snapshot of the different informal pilot evaluations that were conducted.

Prototype systems that enabled sharing gaze between collaborators were developed as a plug-in for the Google Hangouts⁸ video-calling system. I used ETUDriver (Bates & Spakov, 2006), a middleware that allows an end-user application to seamlessly connect with multiple gaze trackers. This enabled us to use the same software, with multiple remote (e.g. Tobii T60, Tobii X2) and head-mounted (e.g. Pupil,⁹ Ergoneers Dikablis¹⁰) gaze-tracker configurations. This platform allowed us to explore the value of the technology outside the confines of the laboratory, as well as between collaborators who are truly geographically separated and involved in complex, real-world tasks (e.g. collaborative shopping in a large grocery store, showing city landmarks to a remote partner, or collaboratively choosing a specific book from the library).

⁸ <https://hangouts.google.com/> (Accessed 16 Feb 2019)

⁹ <https://pupil-labs.com/> (Accessed 16 Feb 2019)

¹⁰ <https://www.ergoneers.com/en/hardware/dikablis-glasses/> (Accessed 16 Feb 2019)

We faced multiple technical issues that limited the feasibility or usefulness of gaze-augmented video calling in such “in the wild” environments. Two of the biggest technical challenges we faced concerned video-calling delay and limited gaze-tracking quality and system stability in a mobile environment. In addition to the technical issues, we also observed that the task characteristics play a very important role in how people use shared gaze and perceive its value for collaborative physical tasks. Below, I discuss the key learnings from the early pilot evaluations.



Figure 10. Different real-world collaborative use cases of gaze-augmented egocentric video explored as part of the thesis: (a) remotely guiding a museum visitor, (b) collaboratively selecting a book from the library, (c) grocery shopping, with the guidance of a remote person, (d) troubleshooting a coffee machine, with the help of a remote guide, (e) real-world puzzle solving, (f) collaboratively exploring notice-board advertisements, (g) collaborative assembly task with physical projection of gaze, (h) collaborative exploration of different city landmarks, and (j) a LEGO-building task

Effect of delay in video calling

A typical audio call made through the cellular network is very fast, as carriers have dedicated bandwidth for the service. However, the quality of video calling over the Internet can be influenced by multiple network factors. Since the purpose of video-calling is to facilitate real-time interaction, users are normally sensitive to any noticeable delay. However, the sensitivity to video delay may also depend on the task and the context of the video call. Certain conversational tasks are less sensitive to video delay if they involve less-frequent turn taking (e.g. remote presentation). In typical conversational video calls, a delay of 500–700 ms is common and often considered acceptable (Berndtsson et al., 2012). Yu et al. (2014) studied typical video-calling delays between two devices located in the

same city using major video conferencing services such as Google Hangout, Skype, and FaceTime. They found that video delay can typically vary between 300 ms to less than 1 second, with frequent spikes of much longer delays (up to 10 seconds). They noted that different dynamic network conditions, such as the bandwidth variation and packet loss, can influence the video-calling experience. Further, other factors such as whether the call is made through a cellular network or Wi-Fi, as well as the reception quality of the network, can also influence the overall delay in video transmission.

We noticed that gaze-augmented video is sensitive to network delay. In one of the pilot tests, we evaluated the system in a collaborative shopping scenario involving a mobile user wearing a head-mounted gaze tracker in a grocery store collaborating with a remote stationary user (see Figure 11(c)). We observed multiple instances where the stationary user would refer to the gaze of the mobile user as a way for deictic referencing (e.g. *“take the one that you are looking at”*). Even a slight video delay reduced the usability of the gaze cursor. The mobile user would constantly shift the focus of attention from one object in the store to another object. Thus, the deictic references that involved the gaze pointer often led to misunderstanding (e.g. *you mean this?* [while taking the wrong piece]). Such situations required extensive verbal effort to correct (e.g. *“not this one, the red-coloured one you looked at earlier”*). The effect of delay in video is further accentuated, as the mobile users may not often be aware of their own shifts in attention at a fixation level (e.g. *“OK, what did I look at earlier? This one, maybe?”*). A noticeable delay in video communication can reduce the benefit of gaze augmentation.

Surprisingly, the delay in video communication and its effects on the usability of shared gaze are aspects that are not adequately discussed in the previous literature. This could partly be due to the fact that the previous studies were conducted in a lab environment making use of dedicated LAN connections. Despite our best efforts, we could not manage to consistently get a lag-free video-calling platform outside a controlled environment using a 4G cellular network. It should be noted that the overall video lag we experienced is the result of cumulative delay incurred due to processing of the gaze tracker, processing of the video at the sender and the receiver’s end, network transmission delay, and the delay in presenting the content on screen.

Although we could not make the shared gaze prototype work fast enough on the current cellular networks, the advent of 5G networks and other advancements in mobile network connectivity promise lower latency and higher bandwidth. It is likely that commercial applications using gaze-augmented video for collaborative physical tasks may become a reality in the near future.

Gaze-tracking robustness and accuracy in a mobile situation

The second option we explored was to conduct the study in a semi-controlled environment at the university cafeteria. Such an environment allowed us to better access and set up WLAN networks so as to overcome the technical problem associated with video delay. The study involved a collaborative shopping scenario where a mobile participant had to show all the items on sale to the remote user and buy certain items of the remote stationary participant's choosing from the cafeteria. Figure 12 shows the university cafeteria settings where the pilot evaluations were conducted.



Figure 11. The university cafeteria environments where the pilot evaluations were conducted

The task required the participant to navigate the different aisles of the cafeteria. We used the Ergoneer Dikablis binocular gaze-tracking system for the mobile participant. The constant movement hindered the accuracy and robustness of gaze tracking. The effect of limited accuracy of tracking was accentuated, as the objects of interest (e.g. drinks in the refrigerator, sandwiches on the shelf, etc.) were small in size, as seen through the head-mounted camera. This led to the understanding that accurate gaze tracking is critical to the usability of shared-gaze interfaces, which is another aspect that is not well represented in the previous work.

Even in this semi-controlled environment, we faced unforeseen technical problems associated with the quality of tracking and several hardware issues. The data collection could not be completed.

Almost 15 years have passed since the first study by Fussell et al. (2003). Gaze-tracking technology has seen rapid advances since then in terms of the accuracy of tracking and ergonomics of use. Despite that, our observations are similar. Wearable gaze-tracking technology may not yet be robust enough for consumer applications involving collaborative physical tasks. Collective research effort is required to ensure that the technology works reliably for all users for collaborative tasks that require

high accuracy in tracking, involve frequent mobility, and are difficult to track environments.

Linguistic complexity of the task

In a variety of other tasks that we evaluated, the value of shared gaze was not clearly visible in terms of task efficiency. For example, we evaluated the value of shared gaze in a relatively simple LEGO-based joint construction collaboration task. One of the collaborators had direct access to the puzzle block but did not know what shape to construct. The remote collaborator knew the shape to construct but did not have direct access to the LEGO pieces. The collaborator who had access to the LEGO blocks used the Ergoneer Dikablis head-mounted gaze tracker. The gaze-augmented egocentric video from the gaze tracker was transferred to the remote partner. The task was to ensure efficient collaboration so as to build the structure.

In this specific task, the LEGO blocks used had different simple shapes (e.g. cube, cylinder, etc.) and distinct colour, and blocks afforded a few specific methods of arrangement (see Figure 11(j)). This made the task linguistically simple, making it easy to refer to task objects and locations, using their salient properties such as colour or shape. The affordance of the LEGO blocks also made the orientation and placement of the blocks intuitive. Thus, shared gaze did not directly improve referencing objects and locations and did not appear to lead to any noticeable performance improvement. This is in line with the findings of Macdonald and Tatler (2017), in the context of collocated collaboration, and D'Angelo and Gergle (2016), in the context of shared-display collaboration, in that the value of knowing where your partner is looking is more useful when the task is linguistically complex.

The benefit of the shared gaze, and one could argue that this applies to all remote-gesturing mechanisms, is amplified in an environment that is linguistically complex (i.e. in scenarios where verbal instructions can be difficult) due to the complexity of the tasks or collaborators' language proficiency. This does not mean that, in linguistically simple collaborative scenarios, there may not be any benefits of shared gaze at all. It is possible and likely that there are still benefits of shared gaze in terms of improved quality of collaboration, increased redundancy in communication, and other subjective aspects. Also, it may be the case that there are small performance improvements with shared gaze in such linguistically simple collaborative scenarios. However, experimentally validating this claim would have required an experiment with an unreasonably large sample size.

The design of the studies undertaken in this thesis should be seen in the light of the learnings from these early pilot evaluations. Based on the lessons we learnt, as well as the review of previous literature, we

hypothesised the potential benefits of shared gaze in collaborative physical tasks (e.g. prediction of intention (Akkil & Isokoski, 2016b), spatial referencing (Akkil & Isokoski, 2016a)). We then deconstructed a collaborative physical task to find sub-tasks where clear statistically significant benefits of shared gaze would be available. We then carefully designed experimental studies in a controlled lab environment to evaluate our hypothesis. The publications (Studies III and IV) used such an approach.

In contrast, Studies V and VI involved real-time collaboration, and we conducted them in a controlled lab environment. The laboratory environment helped the research in multiple ways. It allowed us to tune the technology to work optimally within the limits and confines of the lab using dedicated high-speed WLAN and custom video-calling solutions. It also allowed us to overcome the potential social and privacy issues associated with performing such studies “in the wild.” Another advantage of the controlled laboratory environment was that it enabled us to define complex, artificial but representative, and experimental tasks that highlight the specific costs and benefits of shared gaze. We could also create multiple versions of the same task, with comparable complexity, enabling us to leverage the strength of a within-subject experimental design.

6.3 RESEARCH ETHICS

Even though the field of HCI is relatively nascent, it already has well-established codes of conduct associated with research ethics (e.g. the ACM code of ethics).¹¹ However, as a field that is rapidly advancing, ethics is an important aspect that requires constant reflection and discussion.

Three studies (Studies I, III, and VI) reported in this thesis involved benign situations, where the participants were not fully informed in the beginning about the purpose of the study or how the data collected would be utilised. Study I involved a “hidden” gaze-tracker calibration mechanism. We calibrated the users in the background while they were answering a survey questionnaire on the computer using mouse. In Study III, I recorded the gaze of two actors while they were driving a car simulator. While the actors knew their eyes were being tracked, they were not told beforehand that, in the second phase of the study, other participants would watch the video recorded through their head-mounted camera, with the gaze point overlaid in an intention-prediction task. In Study VI, I studied the value of sharing implicitly produced gaze in a remote-collaboration task. In a specific condition of the experiment, we told participants their gaze was not being shared, when in reality it was. In all

¹¹ <https://www.acm.org/code-of-ethics>

the studies, we told the participants about the deception immediately after the data collection and explained why such an approach was required. In all cases, we offered the participants an option to withdraw their data (immediately or later by email) without any penalty or pressure.

We conducted all three studies in carefully controlled laboratory environments with experimental tasks that were carefully chosen, such that deception did not pose any risk or provide any chance for the participant to inadvertently reveal any personal information. In all the studies, the benign deception was short-lived and did not pose any risk to the participants in terms of their privacy, interests, or well-being.

Such an approach was required to gather important information about the value of naturally produced eye movements in human-computer interaction and computer-mediated collaboration. To answer the research questions and to maintain the validity of the research study, it was critical that the participants were not aware of the experimental details. For instance, in Study I, if participants were made aware that the system would attempt to calibrate the gaze tracker while they were answering the survey questions, it is possible they would have altered their gaze behaviour to enable the calibration (e.g. by looking at the interaction area for an unnaturally longer time). In Study III, if the participants were made aware that, in the later phase of the study, other participants would view their gaze-overlaid egocentric video in order to predict the turn direction of the car, they might have consciously or subconsciously made eye movements to guide this prediction or tried to suppress the naturally occurring guiding eye movements. Such a behaviour would have biased the collected data and exaggerated or reduced the value of gaze. In Study VI, if the participants knew about their shared gaze in remote collaboration, they would have used it explicitly as a means of communication. This would have limited the insights we could have gathered about the communicative value of implicitly produced gaze.

In the research reported in this thesis, we used video recording of the user study session. We either used the videos as viewing materials for the second phase of the study (e.g. Studies III and IV) or to analyse the interaction between participants by transcribing the speech and relevant actions (e.g. Studies II, V, and VI). We safely archived all of the data, including the video data, and all of the results were reported while maintaining participants' anonymity, as per the guidelines provided by the Finnish Advisory Board on Research Integrity¹² and the practices followed at the Tampere University at the time when the research was undertaken.

¹² <https://www.tenk.fi/en>

Ubiquitous gaze-based interaction and shared-gaze interfaces in particular are not without their share of privacy issues. Shared-gaze interfaces that transfer a user's gaze information of remote locations may have privacy implications (Zhang et al., 2017) and may not be desirable in all situations. The exploratory research undertaken as part of this thesis to understand potential users' expectations of everyday gaze interaction technology also highlights the potential users' privacy concerns regarding the technology. Even though the focus of the research is to understand the costs and benefits of shared gaze in a collaborative physical task, it is also in the interest of this thesis to initiate a discussion on the potential privacy issues of such systems, as well as their implications in a future world, where gaze tracking may be ubiquitous.

Summary

- *The studies reported as part of this thesis were grounded on previous literature on shared gaze and influenced by the learnings from the early pilot evaluations.*
- *Delay in video communication and quality of gaze tracking influence the value of shared gaze.*
- *Value of shared gaze is amplified in linguistically complex situations.*



7 Introduction to the Studies

The publications produced as part of this thesis can be categorised into three different themes. Figure 13 presents a visual representation of how the individual studies are situated in the different themes.

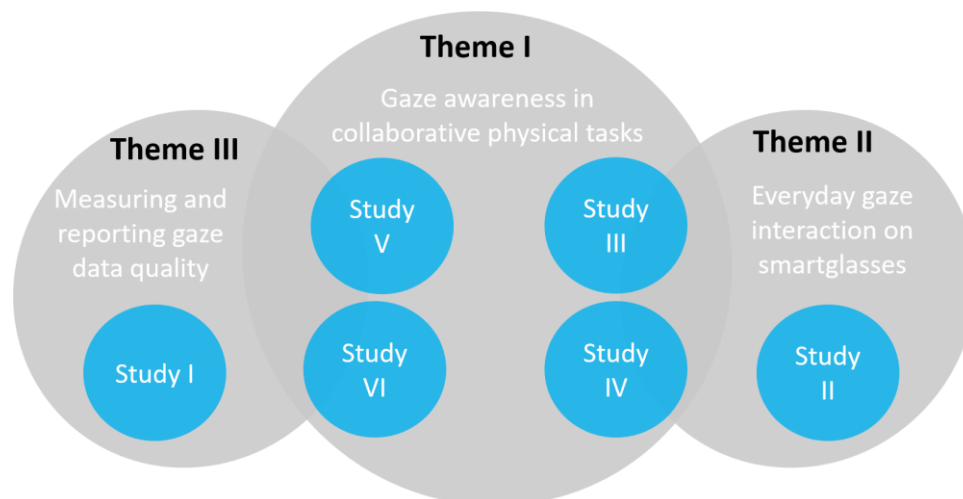


Figure 12. The publications and research themes

The primary focus of the thesis was to understand the value of shared gaze in collaborative physical task (Theme 1). Four (Studies III, IV, V, and VI) out of six publications reported in the thesis directly enabled this research objective, whereas the remaining studies (Themes II and III) supported the primary research objective indirectly.

Study 1 provided a tool to measure and report gaze data quality. This was used in the subsequent studies to report the gaze data quality, and in understanding how gaze-tracking accuracy influences shared gaze collaboration. Study II laid the required groundwork for exploring smartglasses as a platform for collaborative physical tasks. In Study II, we focused on understanding the user's requirements, concerns, and preferences regarding smartglasses with gaze tracking. Smartglasses with gaze tracking capability are particularly suited for collaborative physical tasks, as they normally have a world-facing camera (e.g. Google Glass, Epson Moverio). The wearable form factor enables the user's hands to be free to perform the physical actions. Collaboration using gaze-augmented egocentric video from a wearable gaze tracker or smartglasses was the focus of two of the publications included in the thesis (Studies III and IV). Even though Study II did not directly explore collaborative use cases, it can be considered a step towards human-centred design and development of everyday gaze interaction on smartglasses and a prerequisite to exploring shared gaze for collaborative physical tasks using smartglasses. In the following section, I introduce the six publications.

7.1 STUDY I: MEASURING AND REPORTING GAZE-TRACKING QUALITY

Reference

Deepak Akkil, Poika Isokoski, Jari Kangas, Jussi Rantala, and Roope Raisamo. "TraQuMe: a tool for measuring the gaze tracking quality." In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 327-330. ACM, 2014. DOI: [10.1145/2578153.2578192](https://doi.org/10.1145/2578153.2578192)

Objective and Methods

In this paper, I presented TraQuMe, a flexible gaze-tracking quality-measurement software. TraQuMe is built on the ETUDriver platform (Bates & Spakov, 2006) and is thus tracker independent and can connect to multiple gaze trackers. TraQuMe shows fixation points such as a conventional gaze-tracker calibration routine. Based on the fixation points and the collected gaze data, it outputs an easy-to-understand visualisation along with numeric values for accuracy, precision, and robustness of tracking. TraQuMe can be used during experiments to ensure gaze data quality, as objective grounds to recalibrate the user or to omit the data from analysis, and to easily report the data quality values in the publication.

Since TraQuMe may need to be run multiple times in an experiment with diverse tracking needs, the speed of measurement and flexibility of use were two of the central design criteria. Figure 14 below shows TraQuMe visualisation displaying good and bad gaze data quality.

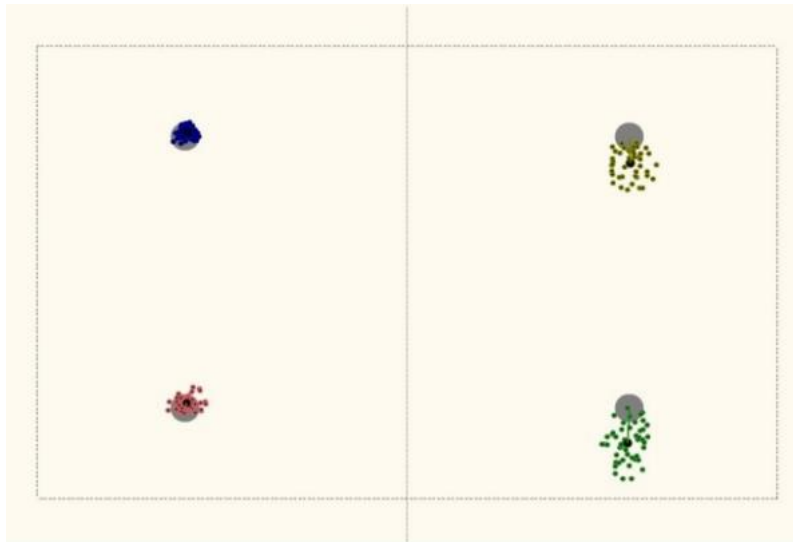


Figure 13. An example visualisation presented in TraQuMe showing good and bad gaze data quality

To evaluate TraQuMe and to provide an example of how to use the software, we conducted a controlled user study comparing a form-based hidden calibration method that we developed to the standard 2-, 5-, and 9-point calibration routines. After each of the calibration routines, we evaluated the tracking quality using a 4-point validation method using TraQuMe.

Results and Discussion

Our evaluations showed noticeable difference in gaze-tracking quality only between the 9-point calibration and the hidden calibration mechanism. The 9-point calibration was the most consistent and accurate, and form-based calibration led to noticeably higher variability in gaze-tracking accuracy across different participants. None of the differences were statistically significant. The results suggest that the hidden calibration method is almost as good as the calibration where the participants are knowingly cooperating.

In this paper, we studied two interlinked issues in the research involving gaze tracking. We presented a tool for measuring and reporting gaze data quality. We further presented a hidden gaze-tracker calibration mechanism to be used in experiments where users' awareness of being gaze tracked can potentially influence their gaze behaviour. We used TraQuMe in the context of comparing the new calibration routine to the standard 2-, 5-, and 9-point calibrations. Both TraQuMe and the hidden gaze-tracker calibration mechanism are functional, and we warmly recommend both to the research community.

7.2 STUDY II: USER EXPECTATIONS OF EVERYDAY GAZE INTERACTION

Reference

Deepak Akkil, Andrés Lucero, Jari Kangas, Tero Jokela, Marja Salmimaa, and Roope Raisamo. 2016. User Expectations of Everyday Gaze Interaction on Smartglasses. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 24, 10 pages. DOI: [10.1145/2971485.2971496](https://doi.org/10.1145/2971485.2971496)

Objective and Method

Gaze-tracking technology is increasingly available at prices cheaper than ever before. The technology is a feasible, practical, and beneficial input modality in different everyday contexts, such as interacting with large displays (Melodie Vidal et al., 2013), mobile phones (Drewes et al., 2007), and wearables such as smartwatches (Akkil et al., 2015) and smartglasses (Baldauf et al., 2010).

There has been a lot of promising research on gaze-based interaction for the mainstream consumer market in the past decade. However, most of them are limited to technology development or evaluation of new interaction techniques to be used in a very specific context of use. Such studies, while extremely useful, provide limited insights into a user's holistic perception and expectations of this promising technology. They do not answer the larger yet fundamental questions, such as *What do potential users feel about an environment where gaze interaction is ubiquitous?*, *In what contexts would users want to use gaze interaction?*, *When would gaze-based interaction not be acceptable?*, and *What are the social and personal implications of everyday use of this technology?* We chose smartglasses capable of gaze tracking as a platform to explore these fundamental questions.

We conducted six exploratory focus group sessions with heterogenous groups of participants. We used five carefully crafted scenarios, each giving an abstract "ideal-world" narration of a distant future with gaze-tracking smartglasses. All scenarios were inspired by previous research on gaze-based interaction and included a mixture of different use contexts (indoor/outdoor, individual/social, private/public). These scenarios were used as the probing material for the focus group sessions. The moderator asked several open-ended questions to understand how users felt about using the technology in the context of the scenario and their requirements, concerns, and preferences.

We first transcribed the focus group sessions and analysed them using affinity diagramming (Holtzblatt & Beyer, 2014). We arranged different affinity notes hierarchically based on their content, allowing themes to emerge organically.

Results and Discussion

Below, I describe the three prominent themes that emerged from the study.

Social Aspects of the Technology

The context of use influenced how participants felt about gaze-tracking technology. Generally, gaze interaction on smartglasses was perceived positively only when alone in a private or public context and was perceived negatively in social situations. There were three specific concerns regarding the social aspects of the technology. First, participants felt that performing unnatural eye movements to interact with the smartglasses or the environment in a social situation may be noticeable to an onlooker and thus may look “weird.” The second concern was about how easy it would be to covertly interact with the device while pretending to be in a social situation. Participants were of the opinion that wearing gaze-tracking smartglasses in a social situation may be perceived negatively. Third, participants recognised the importance of eye contact in social situations and how using eyes to interact may reduce the natural communicative use of eyes in social situations. Generally, the technology was perceived to be not conducive to sociability.

Concerns about the Technology

Participants raised concerns about several aspects of the technology, such as personal safety and health, privacy, and trust issues of the technology. Participants were concerned about the safety implications of the long-term use of the system and the health implications of performing unnatural eye movements to interact with the system. Another major concern about the technology concerned the privacy aspects of its use, specifically pertaining to collecting personal gaze data. Third, the technology was still considered nascent and not to be trusted in replacing other mature technologies, such as mobile devices. Also, users expressed concern about not always being in control with such a technology in terms of the potential ease of identifying when the device is not working properly, troubleshooting issues, and recovering from errors.

Interaction Preferences

The most promising use of gaze-tracking smartglasses was considered to be intuitive interaction with distant objects. Participants expressed the need for subtle feedback when there were interactive objects or information at the point where they were looking. *“Glasses should be polite; it should ask if the user wants to know more information about the item”* [P18]. Generally, the preferred interaction technique was to dwell on the items, whereas gaze gestures were considered suitable for short and infrequent interactions. Users may not want to use gaze interaction in all use contexts. Future gaze-tracking smartglasses should not rely on gaze as the main, or sole, input modality. The device should support complementary input (e.g. mobile device, voice input, etc.) and nonvisual output modalities to enable

flexibility of use (e.g. by allowing users to disable or provide optional vibrotactile feedback to communicate subtle information without distracting the user).

7.3 STUDY III: GAZE AUGMENTATION AND AWARENESS OF INTENTION

Reference

Deepak Akkil and Poika Isokoski. 2016. Gaze Augmentation in Egocentric Video Improves Awareness of Intention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1573-1584. DOI: [10.1145/2858036.2858127](https://doi.org/10.1145/2858036.2858127)

Objective and Method

During recent years, there has been growing interest in wearable cameras (e.g. the GoPro camera) and smartglasses (e.g. Epson Moverio). Such devices can capture the egocentric (first-person view) video of a person, enabling users to stream the video in real time over the Internet or use it as a medium to facilitate remote collaboration.

We identified three potential benefits of overlaying gaze information in egocentric video for remote collaboration: i) aid deictic referencing using gaze as an explicit pointing mechanism, ii) improve situational awareness and enable grounding in communication, and iii) enable collaborators to predict intention. However, prior studies have failed to show any clear and measurable practical benefits of gaze overlay in egocentric video (S. R. Fussell et al., 2003). The focus of the paper was to validate that overlaying gaze information can indeed be helpful. So, we chose to focus on one of the three hypothesised benefits: intention prediction.

We deconstructed a potentially collaborative car navigation scenario (i.e. a scenario where a remote person is guiding a car driver to a specific location) to the sub-tasks, where the ability to predict the driver's intention will be clearly visible in turn-taking behaviour at road intersections.

The study was conducted in two phases. In the first phase, we invited two actors to record their gaze behaviour while driving in a car simulator. We recorded 15 videos of each actor driving through a four-way intersection on a road with a single lane in each direction in moderate traffic conditions. In the second phase, we recruited 12 volunteer participants to view the video and predict the drivers' turn direction.

Our study followed a within-subject design with one independent variable (i.e. availability of gaze pointer). There were three dependent measures: i) accuracy of predicting the turn direction, ii) subjective confidence in their

prediction, and iii) the average synchronous gaze distance between the driver and the participant.

Results and Discussion

Our results show that gaze augmentation in egocentric video did enable viewers to predict the intention, not only more accurately but also more confidently. The viewers predicted the turn direction of the driver up to 26% more accurately in the presence of the gaze overlay.

Further, our analysis provides preliminary indication that task-relevant expertise may be a key modulator of the usefulness of the gaze overlay. Participants who rated themselves to be frequent drivers were more confident with their predictions in the presence of gaze overlay than participants who did not drive often. In addition, viewers of the gaze-augmented video exhibited gaze behaviour comparable to that of the driver.

Before this study, we did not know if gaze augmentation in egocentric video could provide any value in remote collaboration. The results of this study are encouraging and provide a platform for investigating the value of gaze augmentation in egocentric video in real-time, real-world remote collaborations.

7.4 STUDY IV: SHARED GAZE FOR SPATIAL REFERENCING

Reference

Deepak Akkil and Poika Isokoski. 2016. Accuracy of interpreting pointing gestures in egocentric view. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 262-273. DOI: [10.1145/2971648.2971687](https://doi.org/10.1145/2971648.2971687)

Objective and Method

Pointing is one of the foundational building blocks of human-human and human-computer interactions. In remote collaboration involving head-mounted cameras, there are four different ways of communicating spatial information. The first is pointing by hand. The user's hand is visible in the video feed and can be used to communicate spatial points of interest. The second is pointing by head orientation by bringing the object of interest to the centre of the video feed. A third option becomes available if the user's gaze can be visualised in real time. If the gaze pointer is overlaid on the video feed, the gaze pointer can be used to communicate spatial information. A fourth option is to show the gaze point, even while pointing with the hand. The purpose of this study was to compare the four different pointing mechanisms in terms of successful communication of the spatial information.

This study was conducted in two phases. In the first phase, we prepared a room with pointing targets attached on the walls (5 rows of pointing target placed on 3 walls, with 24 targets in a row, each separated by 60 cm, i.e., $5 \times 24 = 120$ pointing targets). Next, we invited two actors to perform the pointing task while wearing the head-mounted camera. The actors were chosen such that they had different self-reported ocular dominance (left eye dominant and right eye dominant). After a round of practice, we recorded 30 videos of the actors pointing at different pointing targets for each pointing condition.

In the second phase, we recruited 16 volunteer participants to view the recorded videos and estimate the pointing target. Our experiment followed a within-subject design, with four experimental conditions (i.e. hand, head, gaze, hand+gaze). In addition to the pointing target, we also gathered the participant's subjective confidence in the estimation.

Results and Discussion

Our results indicate that gaze augmentation (i.e. both the gaze and hand + gaze conditions) enabled more accurate and confident estimation of pointing targets than hand-only and head-based pointing conditions. The differences between gaze and hand + gaze were not statistically significant. Further, interpreting targets with gaze augmentation was not influenced by the eccentricity or density of the targets. On the other hand, hand pointing was more difficult to interpret when the targets were closer together, and head-pointing was more difficult to interpret when the targets had large eccentricity.

Our results indicate that all four of the conditions are feasible pointing options in the egocentric view. Supporting gaze modality needs to be considered based on the accuracy requirements of the task. Second, when gaze is available, additional hand pointing (i.e. hand + gaze) does not lead to significant improvement in accuracy over gaze-only pointing. Interpreting hand pointing is most accurate when targets are straight ahead and may be influenced by the ocular dominance of the collaborator performing the pointing.

7.5 STUDY V: SHARED GAZE FOR STATIONARY COLLABORATIVE PHYSICAL TASKS

Reference

Deepak Akkil, and Poika Isokoski. 2018. Comparison of Gaze and Mouse Pointers for Video-based Collaborative Physical Task. In *Interacting with Computers* (2018).DOI: [10.1093/iwc/iwy026](https://doi.org/10.1093/iwc/iwy026)

Objective and Method

Previous studies on remote collaboration have shown that sharing gaze and sharing mouse movement between collaborators can be useful. However, no studies exist comparing these two mechanisms in the context of a collaborative physical task. This publication presents the first experimental study comparing shared gaze and shared mouse cursor for collaborative physical tasks.

Our study involved remote collaboration between a remote desktop computer user (expert) and a worker in a puzzle block-arrangement task. The worker had access to physical puzzle blocks but did not know the structure to build. The remote expert knew the structure to build and was required to guide the worker to accomplish the task. Our collaboration system used overhead cameras and projector systems at the worker end. The expert saw the video feed from the overhead camera, with their gaze or mouse cursor projected directly on the task space of the worker.

Both gaze and mouse cursors are continuous pointing mechanisms that can be useful for explicit spatial referencing. In contrast, gaze also implicitly conveys attention and cognitive processes, whereas mouse always requires explicit user action. The automaticity provided by gaze may be useful in situations where the user is distracted or is multitasking (Schneider & Pea, 2014). Further, remote experts may have differing strategies for using the shared mouse. Some may use it frequently, some rarely, and others not at all, despite being always available (Müller et al., 2013). On the other hand, shared gaze ensures a level of consistency of use between collaborators.

Further, to understand the difference between the shared gaze and shared mouse cursors, we designed an experiment with two independent variables: the pointer used by the expert (gaze, mouse, none) and the expert's level of distraction of (distraction, no distraction). The experiment followed a within-subject design and involved 24 participants (12 pairs). We analysed the effect of the pointing modality on task-completion times, perceived quality of collaboration, and characteristics of the conversation that ensued between the collaborators.

Results and Discussion

Our results suggest that both shared gaze and shared mouse pointer can be useful for video-based collaborative physical tasks compared to having no pointer at all. When comparing gaze and mouse cursor, both performed equally well in sub-tasks that required only pointing. However, mouse cursor outperformed gaze in sub-tasks that required communicating procedural instructions (e.g. *"turn the block like this,"* while making a clockwise movement with the mouse). Analysis of the verbal effort shows that collaborators required more verbal effort with shared gaze than with shared mouse cursor. This can be attributed to larger

verbal effort required to communicate procedural instructions and a larger number of verbal acknowledgements in the gaze condition. Our results indicate the need for analysing the task characteristics (e.g. pointing vs. procedural tasks) before deciding the optimal remote-gesturing mechanism.

We also noticed that gaze-tracking accuracy influences the usefulness of the shared-gaze cursor. An increase in the gaze-tracking offset is correlated with an increase in the task-completion times, as well as with an increase in the total number of phrases required to complete the task.

Our workers appreciated the awareness of where the expert was paying attention when the expert was multitasking. The combination of gaze and mouse pointer (i.e. a mouse pointer that is only visible when the expert is attending to the collaboration) may be useful in such scenarios.

7.6 STUDY VI: SHARED GAZE FOR MOBILE COLLABORATIVE PHYSICAL TASKS

Reference

Deepak Akkil, Biju Thankachan, and Poika Isokoski. 2018. I see what you see: gaze awareness in mobile video collaboration. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 32, 9 pages. DOI: [10.1145/3204493.3204542](https://doi.org/10.1145/3204493.3204542)

Objective and Method

After careful review of the previous literature, I identified that previous work on shared-gaze interfaces have relied on three slightly different shared-gaze configurations. Some of the studies shared the gaze of a person who was not aware of the gaze sharing (i.e., Gaze-Unaware). Thus, the shared gaze is implicitly produced without the intention to communicate (e.g., (C. Liu et al., 2011; Qvarfordt et al., 2005; Stein & Brennan, 2004)). On the other hand, the gaze producer was aware of the shared gaze in some studies, however, could not see their own gaze point was being transferred (i.e., Gaze-Invisible). Some other studies relied on another configuration in which the gaze producer was not only aware of the shared gaze but also viewed the exact point being transferred (i.e., Gaze-Visible). In these situations, participants were more aware of their own eye movements and gaze-tracking quality (e.g., (Akkil et al., 2016; Higuch et al., 2016; Zhang et al., 2017)). However, no previous studies experimentally compared the three configurations to understand if all are equally effective. The objective of this study was to compare the three different gaze configurations (i.e., Gaze-Unaware, Gaze-Invisible, and Gaze-Visible).

We designed a controlled user study comparing the three different shared-gaze configurations, to a baseline of shared mouse in mobile video-based collaborative physical tasks. The task was to arrange 3D puzzle blocks in a predefined form. One of the collaborators used a mobile phone to share the video to a remote instructor, who saw the video on a stationary computer display. The mobile user had access to the puzzle blocks but did not know the arrangement to make. The remote stationary user was aware of the final arrangement, however, could not directly access the puzzle blocks. The video feed from the mobile camera was also presented on the mobile phone display with the gaze or mouse of the remote stationary instructor overlaid. The task for the pairs was to collaborate over mobile video telephony to arrange the blocks correctly.

Our study followed a within-subject design. We recruited 24 participants (12 pairs) to take part in the study. There were four experimental conditions (Gaze-Unaware, Gaze-Visible, Gaze-Invisible, and Mouse). The dependent variables were task completion times, number of utterances required to complete the task, and subjective perception of the collaboration.

Results and Discussion

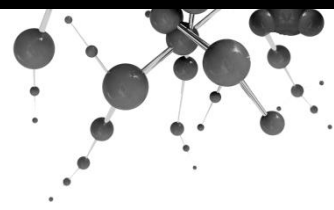
In the Gaze-Unaware condition, pairs took significantly more time to complete the task than with the mouse. Similarly, in the Gaze-Unaware, pairs required significantly more verbal effort to complete the task than all the other conditions. Other differences were not statistically significant.

In terms of subjective evaluation of the condition, the differences between the conditions were clear at the instructor's end and less evident at the mobile worker's end. The instructors overwhelmingly preferred using mouse to gaze sharing, due to its accuracy and flexibility. On the other hand, most of the workers preferred one of the three shared gaze conditions over the mouse. Mobile workers highlighted multiple benefits of shared gaze. Shared gaze was less susceptible to the wrong interpretation when the mobile device moved. While using the mouse, any small movement of the device often led to the wrong interpretation of the pointing target. Furthermore, shared gaze allowed the rough prediction of the target location, even before any verbal instructions.

Our results suggest that implicitly produced gaze may not be as beneficial as explicitly produced gaze in collaborative physical tasks. Thus, designers should provide awareness cues to enable the instructor's awareness of the shared gaze. In addition, it would be best to provide a feature to toggle gaze ON and OFF. This would enable the instructors to check the gaze-tracking accuracy, when they find this information useful. Also, while instructors prefer using the mouse, workers find value in shared gaze. Gaze can be an alternative or complementary awareness mechanism, when the mouse is either not available or not being actively used

.

.....



8 Discussion

This thesis focused on three interlinked themes. The primary theme was to investigate the value of gaze awareness in computer-mediated remote collaborative physical tasks. The second theme was to understand the potential users' expectations in the everyday use of gaze interaction technologies. The third theme focused on the development of a flexible and easy-to-use software tool to enable gaze-tracking researchers to take a more objective view of gaze data quality at different phases of research.

The first and second themes were dissected into three research questions (RQ1-3 below). The tertiary theme had a broad research objective associated with it. In the following section, I will discuss the key findings of my thesis in relation to the initial research questions and research objectives.

RQ1: Can gaze sharing between collaborators lead to measurable benefits in video-based collaborative physical tasks? If yes, what benefits does it provide?

This research question forms the foundation of this thesis. The only study on the topic prior to the start of this work, conducted by Fussell et al. (2003), could not find any measurable benefit of gaze-augmented video for collaborative physical tasks. They attributed their results to the various technical challenges associated with their study set-up. After more than 15 years of technological progress, this thesis revisited this topic.

Measuring the benefits of shared gaze in a complex collaborative environment can be challenging from a methodological perspective. Collaboration itself is a very complex process, and it is often the case that

benefits of subtle interventions, such as shared gaze, are not always evident in high-level measures, such as the efficiency of collaboration and accuracy of the collaborative task (Qvarfordt et al., 2005). The complexities involved in collaboration introduce large subjective variability in the high-level measures. This may make the effect of the intervention hidden in experimental studies, even with reasonable sample sizes. In addition, human behaviour and cognitive processes are highly adaptable. When certain cues are missing, users can extract the same information from other contextual information available or rely more on other available channels to compensate for them (e.g. (Nüssli & Jermann, 2012; van Marlen et al., 2016). These subtle details may not be adequately captured in high-level performance measures and would require low-level analysis of the collaboration to uncover. Furthermore, it is also plausible that certain interventions are beneficial in certain parts of the collaborative process, while being harmful or distracting in other parts, thereby effectively negating the overall benefit.

In order to overcome this methodological challenge, this thesis answers the research question from two different but complementary perspectives. In Studies III and IV, I deconstructed a collaborative task to find subtasks where gaze sharing could be beneficial. Such a deconstructed approach enabled the elimination of additional complexities associated with real-time collaboration. In contrast, Studies V and VI took a more holistic approach by focusing on the value of shared gaze in a real-world, real-time collaborative physical task.

This thesis answered RQ1, based on the findings of Studies III, IV, V and VI. Sharing gaze between collaborators enabled communicating task-relevant information, such as spatial references and intentions, more accurately. Furthermore, it allowed a more confident interpretation of this information at the receiver's end. For example, in our Studies III and IV, viewers of gaze-augmented egocentric video could much more accurately and confidently interpret pointing acts at distant targets and predict the intention of a driver in terms of turn direction at intersections than when gaze information was unavailable. In addition, shared gaze also enabled improved situational awareness between the collaborators (Study III).

Shared Gaze Aids Communication of Spatial Information

Shared gaze is remarkably good at communicating spatial information. Gaze information overlaid on the shared video or directly projected on the task space intuitively communicates the user's visual attention.

Sharing gaze in a collaborative physical task enables the use of the gaze cursor for implicit and explicit spatial references. We naturally tend to look at the objects that we are talking about (e.g., pick the small screwdriver, while naturally looking at it) (Z M Griffin & Bock, 2000) and look at semantically related objects in the scene when we hear their

references (e.g., the partner looks at the screwdriver when they hear “*pick the small screwdriver*”) (Cooper, 1974), thereby implicitly communicating spatial references and comprehension of spatial references. In addition, when the collaborators are aware of shared gaze, it enables the use of gaze information as an explicit mechanism to convey spatial information (e.g., can you see the building that I am looking at).

Our results from Study IV suggest, in the context of collaboration using head-mounted cameras, shared gaze is not just good at communicating spatial information, it may, in fact, be superior to other modalities, such as hand or head pointing in terms of accuracy of communication and confidence of the receiver’s interpretation. Furthermore, our results from Studies V and VI suggest, in the context of collaborative physical tasks involving stationary cameras or mobile cameras, shared gaze is comparable to shared mouse for spatial referencing.

Shared Gaze Helps to Predict Task-Relevant Intentions

Another benefit provided by shared gaze is in enabling the collaborators viewing the gaze-augmented video to interpret the intentions of the gaze producer. Predicting intentions of the collaboration partner can be useful, as it enables the collaborators to be aware of and prepare for what is coming next (Sebanz, Bekkering, & Knoblich, 2006). Furthermore, it also enables them to proactively repair the communication if the upcoming action is not acceptable, even before it happens.

Shared gaze enables the prediction of the partner’s intention in two related ways. First, viewers of the gaze-augmented video can interpret the gaze signal, combine the information communicated by the gaze with other contextual cues, and effectively predict the future physical actions of the user (i.e., what my partner will do next). In Study III, viewers of the gaze-augmented video could more accurately and confidently predict the turn direction of a driver at a four-way intersection, in comparison to viewing the same video without the gaze augmentation. Overlaying gaze information on the egocentric video helped the viewers to interpret what physical actions the gaze producer would perform in the next few seconds. Second, the ability to predict intentions can also manifest in an ability to anticipate the upcoming verbal instruction. In Study VI, which involved an object-assembly task, we observed scenarios where our participants could anticipate which object the instructor would ask them to pick next.

Gaze behaviour is highly intertwined with physical actions (Land, 2006). In everyday physical tasks, the eyes are responsible for not only gathering visual information relevant to the current action but also in gathering the relevant visual information required for executing future motor actions. For example, when approaching a sink to wash the hands, the eyes may already fixate at the soap dispenser to locate its position and plan the future motor action. Such guiding fixations are called “look-ahead”

fixations (Pelz & Canosa, 2001). Look-ahead fixations are very reliable predictors of an upcoming action (Mennie, Hayhoe, & Sullivan, 2007). The time window before the look-ahead fixation occurs depends on the task and can typically occur several seconds before the upcoming motor action. In a typical reach operation to pick an object, it occurs up to three seconds before the reach operation (Mennie et al., 2007). Our results from Study III suggest that viewers of gaze-augmented video can detect patterns of look-ahead fixations from the video and combine it with other contextual information to infer the task-related intention of the partner.

Gaze and speech production are also time locked. Griffin and Bock (2000) note that people look at objects prior to describing them. The eye movements during speech production are closely related to the linguistic complexity (Zenzi M. Griffin, 399AD). Normally, people speaking extemporaneously look at an object roughly one second before its verbal description (Z M Griffin & Bock, 2000). The duration of the gaze on the object prior to the verbal utterances is related to the linguistic complexity. For example, the gaze durations are longer when there are either none (A. S. Meyer, Sleiderink, & Levelt, 1998) or more than one (Zenzi M. Griffin, 206AD) common names to appropriately refer to the object of interest. In Study VI, we observed that in gaze sharing with a stationary instructor and a mobile worker in an assembly task the mobile worker could occasionally predict the correct object to select and the spatial locations to place the object, even before the instructor's verbal point of disambiguation. Such an ability was not available when a more explicit pointer, such as a mouse, was shared between the remote instructor and the mobile worker.

Shared Gaze Enables Situational Awareness and Aids Conversational Grounding

Another benefit of showing the collaborator's gaze on the video is the situational awareness it provides. In Study III, we noticed that gaze-augmentation enabled viewers of the video to synchronise eye movement with the eye movement of the partner (i.e., viewers of the video looked at same parts of the scene synchronously with the gaze producer). Such a coordinated gaze behaviour between collaborators helped establish a state of continual joint attention, improved the awareness of the partner's cognitive processes, and enabled effortless grounding in communication (Richardson & Dale, 2005). The improved situational awareness facilitated by this coordinated gaze behaviour may explain why the viewers of the gaze-augmented video in Study III were better at predicting the task-related intention of the partner.

The importance of coordinated gaze behaviour has been highlighted in previous studies. For example, Richardson and Dale (2005) showed that a closer coordination in gaze behaviour between a speaker and listener is a marker of improved language comprehension. Similarly, Cherubini et al.

(2008) showed that it is possible to automatically detect misunderstanding in collaboration during an event-planning task simply by analysing the distance in gaze patterns of the collaborators.

When gaze of the collaborator is presented, it enables the viewer to utilise it flexibly based on the requirements of the task. For example, Brennan et al. (2008b) found gaze sharing allowed the collaborators involved in a collaborative visual search task to efficiently divide the task space by following a *I look where you are not looking* strategy. In contrast, where the task was to predict a car driver's turn directions, Study III viewers followed an *I look where you are looking* strategy that enabled them to attain the situational awareness from the driver's perspective. Taken together, users followed a flexible approach on how to utilise the shared gaze, depending on the task requirements.

Based on Studies III, IV, V, and VI reported in this thesis, we can conclude that gaze sharing of collaborators can be beneficial by enabling effortless spatial referencing, improving prediction of task-relevant intention, and enhancing situational awareness. These individual benefits can potentially add up to higher-level benefits in collaboration, such as improved efficiency (Studies V and VI), improved accuracy of communication (Studies III and IV), and improved subjective perception of the collaboration (Studies V and VI).

RQ2: What contextual factors influence the usability of gaze-based HCI and, more specifically, shared gaze for collaboration?

In Study II, we explored how the context of use influences the acceptability of gaze-based interaction in general. Gaze as a means to interact with computing devices was generally preferred in an individual use context (i.e., when the user is not involved in collocated social interactions) or not in the presence of other unfamiliar collocated individuals. Human eyes help gather visual information about the environment and naturally communicate visual attention to onlookers, playing an important role in non-verbal communication. Using gaze as an explicit mechanism to interact with computing devices introduces an additional function for gaze. This was considered as problematic in social situations and not conducive to sociability.

Computer-mediated collaborative physical tasks present a scenario where gaze is used as a means of communication between two or more geographically separated collaborators. This introduces additional challenges in the usability of shared gaze. In Studies III, IV, V, and VI, I explored how the contextual factors influence the usability and user preference of shared gaze in computer-mediated collaborative physical tasks. In our series of studies, we noticed three distinct categories of the contextual factors that can modulate the value of shared gaze: task context,

technological context, and the user context. I will briefly describe the different contextual factors below.

Task Context

Everyday computing tasks are pointing intensive. In contrast, many of our everyday physical tasks involve both pointing and complex procedural manipulations (S. R. Fussell et al., 2003). Fussell et al. (2003) note that collocated individuals collaborating to perform a complex physical task use two types of gestures to support the communication: pointing gestures and representational gestures. Pointing gestures are used to communicate objects and locations (e.g., *put that object there*). Representational gestures communicate the form of an object and the nature of action to be performed on it (e.g., *turn the knob like this*, while turning hands clockwise).

In remote collaborative physical tasks that are pointing intensive (e.g., a tourist showing important city landmarks to a remote partner), gaze provided a substantial benefit by allowing effortless and accurate pointing by the user and confident interpretation by the remote partner (Study IV). On the other hand, in tasks that require communicating representational gestures gaze provides less benefit in the collaboration (Study V).

In scenarios involving communicating extensive procedural manipulations, the task requires a more flexible and expressive gesturing mechanism to efficiently communicate, for example, a pen-based annotation system (Kirk & Fraser, 2006), a representation of a hand (Alem & Li, 2011), or a shared mouse cursor (S. Fussell et al., 2004). The flexibility offered by these gesturing mechanisms allows for showing complex physical manipulations (e.g., *turn it like this* while making a clockwise movement of the mouse, or *place it like this* while making a Z gesture with the mouse). Gaze provides little flexibility to communicate such complex instructions.

Our results suggests that shared gaze may be more useful in physical tasks that are pointing intensive than tasks that involve communication of complex representational gestures.

Technical Context

One of the critical factors that affects the usefulness of shared gaze in remote task-based collaboration is the accuracy of gaze tracking. When gaze tracking is not accurate, it can often lead to misinterpretation of the gaze signal, increase the ambiguity in communication, and increase the verbal effort for coordination. Previous research on shared gaze has observed the gaze-tracking accuracy can influence its usefulness (D'Angelo & Gergle, 2016; van Rheden et al., 2017). In a remote guidance task that involves many different task objects and when the shared gaze cursor is misaligned, it may appear to the viewer that the remote user is talking about a different object, requiring elaborate verbal instruction to

clarify the task object and repair the communication (D'Angelo & Gergle, 2016).

Our results extend these observations to collaborative physical tasks and show a correlation between the gaze-tracking accuracy achieved in the study and high-level measures, such as the task completion times and the verbal effort to complete the task (Studies V and VI). Gaze-tracking accuracy can influence the high-level quantitative measures of the collaboration. As the gaze-tracking accuracy reduces, collaborators take more time and verbal effort to complete the task.

User Context

Three factors in the user context may influence the usability of shared gaze in collaborative physical tasks. First, the task-related expertise of the collaborator may influence how confidently they interpret and use the gaze signal to benefit the collaboration. In Study III, participants who reported to be frequent drivers were also more confident in their prediction of turn direction when shared gaze was available. Frequent drivers are more familiar with gaze-allocation strategies that may be associated with driving and, thus, may be better at utilizing the gaze channel. Even though the difference was not statistically significant, it is indicative of a task-based expertise component in how viewers make use of the partner's gaze information.

The task-related expertise of the viewer may be a factor in tasks that have easily distinguishable characteristic gaze patterns associated with them and a meaning behind the different gaze allocation strategies. In the driving scenario, there were different gaze patterns of the driver that may have been easier to decode for a viewer who drives frequently. For example, when turning left in the presence of an oncoming car, the driver looked at the oncoming car to determine if it was slowing down, in order to safely navigate a left turn; or, when going straight ahead, the driver looked at both the left and right sides to ensure there were no other vehicles approaching the intersection and then fixated straight ahead.

Second, the awareness of the producer of the gaze of the gaze sharing can influence how well they use it to aid communication. In Study V, we found that the collaborator's awareness of gaze sharing can mediate the utility of the shared gaze cursor. Collaborating pairs in which the producer of gaze was unaware of the gaze sharing relied more on extensive verbal instructions and were less efficient than pairs who were mutually aware of the gaze sharing. Pairs who were aware of the gaze sharing used it as an explicit medium for communication, and relied less on long verbal utterances, with an increased shift towards clearer deictic references.

A third factor in the user context that could influence the utility of shared gaze is the collaborator's role. Remote collaborative physical tasks, by nature, introduce an asymmetry in collaborator roles. In Study VI, there was a clear difference in user preferences, depending on the roles of the collaborators. Instructors situated in a stationary environment, who were also the producer of the shared gaze, preferred to use the mouse for remote gesturing, due to the accuracy and flexibility offered by the mouse to convey complex instructions. On the other hand, mobile workers, who received the shared gaze, found value in gaze sharing due to its intuitiveness, consistency of use, and ability to predict the upcoming instructions.

RQ3: How does shared gaze compare against other remote gesturing mechanisms available for collaborative physical tasks?

Most computer-mediated collaborative contexts enable different modalities to communicate task-relevant information. For example, shared gaze provides fine-grained awareness of visual attention between collaborators. On the other hand, communicating head-orientation of the partner provides a rough awareness of the attention. In addition to different levels of attention awareness, other remote gesturing mechanisms, such as a shared mouse, hand representations through video, and touchscreen-based annotation systems are also available in different collaborative scenarios.

The different remote gesturing mechanisms have their own unique affordances and limitations. The choice of which communication cue to use for remote collaboration should be made based on the characteristics of the task and the context in which the collaboration is taking place. It should be noted that all the viable communication cues may not compete and some may be complementary to each other. For example, Higguch et al. (2016) shows that gaze sharing along with presenting collaboration partner's hand representation can enable more efficient collaboration in construction tasks than simply presenting the hand representation.

Most of the previous experimental research involving shared gaze for task-based collaboration has compared shared-gaze interfaces, with interfaces that do not offer any additional communication mechanisms. Notable exceptions are studies by Muller et al. (2011, 2013a, 2014). Previous studies that compared shared gaze with a no pointer baseline have found several benefits of shared gaze. Such comparisons are theoretically interesting when understanding the implicit and explicit benefits of collaborative shared gaze. From a practical standpoint, however, the studies provide little insight into the value of gaze in computer-mediated collaboration where more than one remote gesturing cue is available. When more than one remote gesturing mechanism is available, it is important to compare the combinations against each other to understand which works best in a given circumstance. Another

criticism with respect to the previous studies is they tend to paint an overly positive picture of the value of shared gaze, since they compare gaze with a “weaker” experimental baseline where there are no other remote gesturing mechanisms.

Three of the studies included in this thesis present experiments where shared gaze was compared with other feasible gesturing mechanisms in the context of the collaboration (Study IV, V, and VI). Study IV compared the different ways of communicating spatial information in collaboration involving a head-mounted camera. We compared shared gaze with hand-based, head-based and a combination of hand and gaze -based remote gesturing. In Study V, we compared the shared gaze with a shared mouse cursor in real-time remote collaboration involving a 3D object arrangement task. In Study VI, we compared shared gaze with a shared mouse in mobile collaborative physical tasks.

Our results suggests that sharing gaze information is more accurate in communicating spatial information in egocentric view compared to using hands or head-based pointing. The superiority of shared gaze for communicating spatial information in computer-mediated collaboration involving head-mounted cameras makes it a compelling option for video-based collaboration involving pointing-intensive tasks.

In collaboration involving a stationary remote user, gaze sharing and mouse position of the stationary user are both feasible. When such collaborations involved stationary cameras, using a shared mouse was noticeably faster than shared gaze, and was preferred by both the collaborators (Study V). A more detailed analysis of the sub-tasks showed that shared gaze and shared mouse were not statistically significantly different in subtasks that involved extensive pointing, while the mouse outperformed gaze in the subtask requiring communication of complex representational information.

However, the differences between shared gaze and shared mouse are not so straight forward in the collaborative physical tasks using mobile video. The frequent movement of the device in the hands of the mobile user makes pointing at task objects with a mouse more difficult. Additionally, the smaller mobile display makes interpretation of complex mouse expressions challenging for the mobile worker. Compared to a shared mouse, shared gaze is less affected by the frequent movement of the mobile camera. In mobile collaborative physical tasks, shared gaze was comparable to shared mouse in terms of efficiency of collaboration, and the preference of modality depended largely on the roles of the collaborators (Study VI). The mouse is a remarkably flexible and expressive tool to communicate complex spatial and procedural instructions (Gutwin & Penner, 2002). Shared mouse was generally

preferred by the stationary expert. Shared gaze, on the other hand, was found useful by the mobile workers.

In summary, shared gaze is remarkably good at communicating spatial information. This is especially true in collaborative physical tasks involving egocentric view. However, when the task requires communicating extensive procedural instructions, mouse outperforms gaze. The differences between shared mouse and shared gaze are more pronounced in collaboration involving stationary cameras than mobile cameras.

Research Objective: Enabling Researchers to Objectively View Gaze Data Quality

The discussion on gaze data quality and its impact on research findings is not new. The pioneering work done by Holmqvist et al. (2012) and Nyström et al. (2013) have helped us understand to what extent individual and environmental factors influence the gaze-tracking quality, and the impact it may have on research findings. The focus of this thesis was instead on equipping researchers with the tools and recommendations to view the gaze data quality in a more objective way in different phases of the research.

In practice, this means enabling researchers to do the following (Please note that this is only an indicative list, and the relevance of these factors may vary based on the research):

1. Setting an objective threshold of when to recalibrate a participant (e.g., “The participants were recalibrated when gaze tracking offset was greater than 3 cm at the centre of the screen”.)
2. Setting an objective threshold of when to exclude data from the analysis (e.g., “The data from P1 was excluded because the participant could not be tracked robustly and more than 50% of the gaze samples were missing”.)
3. Reporting the objective gaze data quality obtained in the studies. (e.g., “Gaze tracking accuracy obtained in the study varied from 0.5 degrees to 2 degrees of visual angle, mean = 0.75 degrees, SD = 0.3 degrees”.)

One of the core works done as part of the thesis was to develop TraQuMe, an open-source, tracker-independent tool to measure gaze data quality. It should be noted that some of the gaze tracker manufacturers already provide numeric values indicating the tracking quality. For example, Tobii Pro Lab (Tobii AB, 2017) analysis software outputs the accuracy and precision of gaze tracking soon after the calibration procedure. However, the gaze data quality evaluation cannot be performed independently of

the calibration nor in any other point other than those used for calibration. One of the advantages of TraQuMe is its flexibility of use. Furthermore, TraQuMe can be used with eye trackers from different manufacturers, enabling a direct comparison of results.

TraQuMe was utilised in this thesis in order to enable the understanding of how gaze data quality influences the use of shared gaze in collaborative physical tasks. In addition, one of the objectives in making the tool freely available to the research community was to enable and encourage researchers to make use of the tool in their research and build on it, if necessary. This thesis also presented recommendations on how to use TraQuMe and report the gaze data quality in research involving gaze trackers.

Some significant publications in the field of gaze-based HCI have already made use of TraQuMe in research. For example, Raiha et al. (2013) used TraQuMe to evaluate how gaze-based text entry is influenced by the gaze data quality. Spakov et al. (2018) used the system to compare different unsupervised gaze-tracker calibration techniques for school children. Li et al. (Z. Li, Akkil, & Raisamo, 2019) used TraQuMe to exclude data from analysis where gaze tracking offset was more than a certain limit. These examples show the utility and flexibility of TraQuMe for research involving gaze trackers in HCI.

In this thesis, I also set a precedent, by reporting the gaze-tracking accuracy measures obtained in the studies. None of the previous research on shared gaze for collaboration had reported numeric values for the gaze data quality obtained, making the direct comparison of studies difficult. I believe reporting the actual gaze data quality measures is a small, yet significant step towards methodological consistency in research involving shared gaze for collaboration.

Limitations and Future Work

The work presented as part of this thesis contributes only a part towards the vision of pervasive gaze-based interaction, in general, and shared gaze interfaces for remote collaboration, specifically. In this section, I detail some of the limitations of the work presented in this thesis and present avenues for future research.

First, the work presented in this thesis focused on a very specific type of collaboration involving physical tasks. The different collaborative scenarios we looked at involved collaborative car navigation (Study III), spatial referencing (Study IV), and remote guidance (Studies V and VI). We did not cover the length and breadth of all collaborative activity within the realm of collaborative physical tasks. It is very likely that other collaborative activity (e.g., collaborative learning involving real-world physical objects or collaborative visual search in the physical world) may

show different costs and benefits of shared gaze. We should be cautious when generalising the results for other collaborative scenarios. I leave this aspect for future works to investigate.

Furthermore, the methodologies, experimental tasks, and the contexts of evaluation used in this thesis were influenced by the current state of the technology, in terms of accuracy and ergonomics of tracking, and network communication delays associated with video telephony. As technology is rapidly advancing, it is likely technology may soon support “in the wild” investigations, allowing evaluation of the value of gaze sharing in more naturalistic tasks and contexts. This is especially interesting, since our results from Study II regarding users’ expectations of gaze-based interaction suggest that gaze-based interaction may be preferred in an individual use context compared to its use in social situations. Several remote collaborative scenarios may involve other onlookers in addition to the collaborators (e.g., a tourist in a busy city showing interesting landmarks to a remote partner). Social conventions regarding video-recording in public and eye movements in social situations may influence the collaboration strategies in such a scenario. Due to the controlled nature of the studies presented in this thesis, our work provides limited insights into the role of acceptability of shared gaze interfaces “in the wild”.

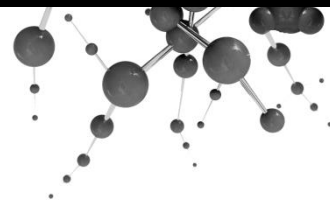
All of the participants were novice users of gaze tracking and, more specifically, of shared gaze interfaces for remote collaboration. All the studies that were conducted as part of the thesis and all the previous work in the area of gaze awareness in remote collaboration have focused on short-term evaluations. We do not know how learning influences the gaze producer’s use for communication and how easily users viewing the shared gaze learn to interpret the communication cues presented by the gaze cursor. Future longitudinal studies are required to answer this question.

The focus of this thesis was finding how gaze awareness improved task efficiency, accuracy, and the extent of the verbal effort required for collaboration (i.e., pragmatic or utilitarian benefits of gaze sharing). One could argue there may be hedonistic benefits in knowing where others are looking in certain scenarios (e.g., playful use of shared gaze (Akkil et al., 2016), the joy of a parent in seeing the reading strategy of their child). We did not investigate the affective characteristic of shared gaze systems nor design choices to make shared gaze interfaces more playful, joyful, and emotionally stimulating. We leave this aspect for future research.

Our work focused on collaboration scenarios where the gaze of one collaborator is shared. Different collaborative physical tasks may benefit from involvement of more than two collaborators (e.g., two geographically separated experts assisting a field worker to accomplish a complex task, or one instructor communicating to multiple workers simultaneously) and

multi-directional gaze sharing. We did not explore how the benefits of shared gaze may translate to such scenarios. We also leave this part for future work.

Lastly, our early attempt to inquire about potential users' concerns and preferences regarding pervasive gaze-based interaction uncovered social and privacy issues associated with the technology. Future research is required to investigate different approaches to address these issues. Furthermore, we hope the early attempt to inquire the user's preferences and concerns regarding gaze tracking technology (Study II) will pave the way to many future works towards human-centred design of pervasive gaze interactive systems.



9 Conclusion

Collaboration is at the heart of human interactions. Humans have evolved to what we are today, as a result of our innate ability to understand and collaborate with each other. With the geographically dispersed nature of workplaces and social circles, technologies that allow remote individuals to communicate and collaborate is a necessity of our times. Despite years of technological advancement, however, distributed collaboration is still a challenge (Bjorn, Esbensen, Jensen, & Matthiesen, 2014).

Today, there is an ever-increasing need to develop improved remote collaboration technologies. Collaboration technologies could, potentially, reduce the need for people to travel to remote locations to conduct different personal and professional tasks. An increased adoption of which could lead to a substantial reduction of the carbon footprint. Development of efficient and natural remote collaboration technologies is more urgent than ever.

Video-based collaboration technology is an important part of computer-mediated social interactions and increasingly relevant in the domains of education, telemedicine, law enforcement, and different industrial and consumer workflows, such as technical support.

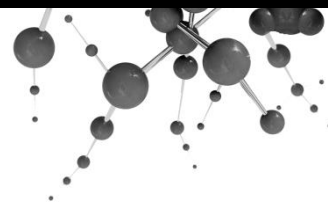
The research presented in the thesis investigated the value of gaze awareness in the increasingly important collaborative scenario involving physical tasks. The results of this thesis will be useful to design improved collaboration technologies to support physical tasks. Furthermore, the results contribute to the theoretical understanding of

gaze awareness in mediated communication. In addition to the focus on collaboration, the methodological and empirical contribution of this thesis extends to gaze-based HCI and research involving the use of gaze trackers.

The key implications of this thesis are as follows.

- Gaze awareness can be beneficial in remote collaborative physical tasks. It enables collaborators to communicate intentions, spatial references and effortlessly establish joint attention. However, contextual details, such as expertise of the collaborators, task characteristics, and accuracy of gaze tracking, can influence the utility of shared gaze.
- It is important to compare shared gaze with other remote gesturing mechanisms to understand the costs and benefits of the different task contexts, in order to design the optimal collaborative interfaces. Shared gaze may not always be beneficial nor a superior communicative cue for collaborative physical tasks. Future remote collaboration systems should support complementary remote-gesturing mechanisms for effortless collaboration.
- Shared gaze is remarkably good for communicating spatial references in an egocentric view, compared to hand and head-based pointing.
- When shared-gaze is used for collaboration, it is important to let the users continuously know their gaze is being shared. In addition to making the users more efficient at using the modality for collaboration, it can potentially reduce the privacy issues that may stem from shared gaze.
- Gaze is considered a useful interaction modality in smartglasses by the potential users. However, it is preferred in an individual use context, as opposed to a social use context. Providing flexibility of use through multiple input and output interaction modalities is key to meeting the end user expectations regarding everyday gaze-based interaction on smartglasses.
- Gaze tracking quality can influence the value of shared gaze and, more generally, the findings in research that make use of gaze trackers. Researchers should take a more objective approach to deal with the gaze data quality in different phases of the research.

The finding of this thesis can contribute towards designing future remote collaboration systems, towards the vision of pervasive gaze-based interaction, and towards the validity, repeatability, and comparability of research involving gaze trackers.



10 References

- Acker SR, & Levitt SR (1987) Designing Videoconference Facilities for Improved Eye Contact. *Journal of Broadcasting & Electronic Media* 31:181-191. DOI: 10.1080/08838158709386656
- Adolphe RM, Vickers JN, & Laplante G (1997) The Effects of Training Visual Attention on Gaze Behavior and Accuracy: A Pilot Study. *International Journal of Sports Vision* 4:28-33
- Akkil D, & Isokoski P (2016a) Accuracy of interpreting pointing gestures in egocentric view. In: UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp 262-273
- Akkil D, & Isokoski P (2019) Comparison of Gaze and Mouse Pointers for Video-based Collaborative Physical Task. *Interacting with Computers*. DOI: 10.1093/iwc/iwy026
- Akkil D, & Isokoski P (2016b) Gaze Augmentation in Egocentric Video Improves Awareness of Intention. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* 1573-1584. DOI: 10.1145/2858036.2858127
- Akkil D, Isokoski P, Kangas J, et al (2014) TraQuMe: a tool for measuring the gaze tracking quality. *Eye Tracking Research and Applications Symposium (ETRA)* 978-981. DOI: 10.1145/2578153.2578192
- Akkil D, James JM, Isokoski P, & Kangas J (2016) GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. *Proceedings of the*

2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16 07-12-May-:1151-1158. DOI: 10.1145/2851581.2892459

Akkil D, Kangas J, Isokoski P, et al (2015) Glance Awareness and Gaze Interaction in Smartwatches. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* 1271-1276. DOI: 10.1145/2702613.2732816

Akkil D, Thankachan B, & Isokoski P (2018) I see what you see. In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications - ETRA '18*. ACM Press, New York, New York, USA, pp 1-9

Alem L, & Li J (2011) A Study of Gestures in a Video-Mediated Collaborative Assembly Task. *Advances in Human-Computer Interaction* 2011:1-7. DOI: 10.1155/2011/987830

Bahill a T, & Laritz T (1984) Why Can't Batters Keep Their Eyes on the Ball? *American Scientist* 249-253

Bahill AT, Clark MR, & Stark L (1975) The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24:191-204. DOI: 10.1016/0025-5564(75)90075-9

Baldauf M, Fröhlich P, & Hutter S (2010) KIBITZER: a wearable system for eye-gaze-based mobile urban exploration. *Proceedings of the 1st Augmented Human International Conference* 9. DOI: 10.1145/1785455.1785464

Bard E, Hill R, Arai M, & Foster ME (2009) Referring and gaze alignment: accessibility is alive and well in situated dialogue. In: *Proceedings of CogSci 2009*

Bard EG, Hill RL, Foster ME, & Arai M (2014) Tuning accessibility of referring expressions in situated dialogue. *Language, Cognition and Neuroscience* 29:928-949. DOI: 10.1080/23273798.2014.895845

Bates R, & Spakov O (2006) D2 . 3 Implementation of COGAIN Gaze Tracking Standards

Bayliss AP, Frischen A, Fenske MJ, & Tipper SP (2007) Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition* 104:644-653. DOI: 10.1016/j.cognition.2006.07.012

Bednarik R, & Shipilov A (2011) Gaze cursor during distant collaborative programming: A preliminary analysis. *Proceedings of the DUET*

- Bednarik R, & Shipilov A (2012) Usability of gaze-transfer in collaborative programming : How and when could it work , and some implications for research agenda. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion ACM
- Berndtsson G, Folkesson M, & Kulyk V (2012) Subjective quality assessment of video conferences and telemeetings. *2012 19th International Packet Video Workshop, PV 2012* 25–30. DOI: 10.1109/PV.2012.6229740
- Billinghurst M, Gupta K, Katsutoshi M, et al (2017) Is It in Your Eyes? Explorations in Using Gaze Cues for Remote Collaboration Mark. *Collaboration Meets Interactive Spaces* 1–483. DOI: 10.1007/978-3-319-45853-3
- Bjorn P, Esbensen M, Jensen RE, & Matthiesen S (2014) Does Distance Still Matter? Revisiting the CSCW Fundamentals. *ACM Transactions on Computer-Human Interaction* 21:1–26. DOI: 10.1145/2670534
- Blattgerste J, Renner P, & Pfeiffer T (2018) Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. *Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18* 1–9. DOI: 10.1145/3206343.3206349
- Blignaut P, & Wium D (2014) Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods* 46:67–80. DOI: 10.3758/s13428-013-0343-0
- Bock SW, Dicke P, & Thier P (2008) How precise is gaze following in humans? *Vision Research* 48:946–957. DOI: 10.1016/j.visres.2008.01.011
- Boucher JD, Pattacini U, Lelong A, et al (2012) I reach faster when i see you look: Gaze effects in human-human and human-robot face-to-face cooperation. *Frontiers in Neurorobotics* 6:1–11. DOI: 10.3389/fnbot.2012.00003
- Brennan SE, Chen X, Dickinson CA, et al (2008) Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition* 106:1465–77. DOI: 10.1016/j.cognition.2007.05.012
- Brennan SE, Hanna JEJ, Zelinsky G, & Savietta K (2012) Eye gaze cues for coordination in collaborative tasks. DUET Workshop, CSCW'12, February 11-15, 2012, Seattle, Washington, USA
- Broz F, Lehmann H, Nehaniv CL, & Dautenhahn K (2012) Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation. In: Proceedings - IEEE International Workshop on Robot and Human Interactive Communication. pp 858–864

- Bulling A, & Gellersen H (2010) Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing* 9:. DOI: 10.1109/MPRV.2010.86
- Carletta J, Hill RL, Nicol C, et al (2010) Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods* 42:254–265. DOI: 10.3758/BRM.42.1.254
- Cary MS (2006) The Role of Gaze in the Initiation of Conversation. *Social Psychology* 41:269. DOI: 10.2307/3033565
- Causer J, Harvey A, Snelgrove R, et al (2014) Quiet eye training improves surgical knot tying more than traditional technical training: A randomized controlled study. *American Journal of Surgery* 208:171–177. DOI: 10.1016/j.amjsurg.2013.12.042
- Causer J, Holmes PS, & Williams AM (2011) Quiet eye training in a visuomotor control task. *Medicine and Science in Sports and Exercise* 43:1042–1049. DOI: 10.1249/MSS.0b013e3182035de6
- Cheng S, Sun Z, Sun L, et al (2015) Gaze-Based Annotations for Reading Comprehension. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* 1:1569–1572. DOI: 10.1145/2702123.2702271
- Cherubini M, Nüssli M-A, & Dillenbourg P (2008) Deixis and gaze in collaborative work at a distance (over a shared map). *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08* 1:173. DOI: 10.1145/1344471.1344515
- Chitty N (2013) User Fatigue and Eye Controlled Technology. OCAD University
- Clark HH, & Krych MA (2004) Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50:62–81. DOI: 10.1016/j.jml.2003.08.004
- Cline MG (1967) The perception of where a person is looking. *The American journal of psychology* 80:41–50. DOI: 10.2307/1420539
- Cook M (1977) Gaze and Mutual Gaze in Social Encounters: How long – and when – we look others" in the eye. *American Scientist* 65:328–333
- Cooper RM (1974) The Control of of Eye Spoken Fixation by the Meaning Language. *Cognitive Psychology* 107:84–107
- Corkum V, & Moore C (1998) The origins of joint visual attention in infants. *Developmental psychology*. DOI: 10.1037/0012-1649.34.1.28

- D'Angelo S, & Begel A (2017) Improving Communication Between Pair Programmers Using Shared Gaze Awareness. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* 6245–6290. DOI: 10.1145/3025453.3025573
- D'Angelo S, & Gergle D (2016) Gazed and Confused : Understanding and Designing Shared Gaze for Remote Collaboration. *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems - CHI '15* 2492–2496
- D'Angelo S, & Gergle D (2018) An Eye For Design: Gaze Visualizations for Remote Collaborative Work. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. pp 1–12
- Drewes H, De Luca A, & Schmidt A (2007) Eye-gaze interaction for mobile phones. *Proceedings of the international conference on mobile technology, applications* 07:364–371. DOI: 10.1145/1378063.1378122
- Duchowski AT (2007) *Eye tracking methodology: theory and practice*. Springer
- Duchowski AT, Cournia N, Cumming B, et al (2004a) Visual deictic reference in a collaborative virtual environment. *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004* 1:35–40. DOI: 10.1145/968363.968369
- Duchowski AT, Shivashankaraiah V, Gramopadhye AK, et al (2004b) Binocular eye tracking in virtual reality for inspection training. *Proceedings of the 2000 symposium on Eye tracking research & applications* 89–96. DOI: 10.1145/355017.355031
- Foddy M (1978) Patterns of Gaze in Cooperative and Competitive Negotiation. *Human Relations Research* 31:925–938. DOI: 10.1177/001872677803101101
- Foulsham T, Cheng JT, Tracy JL, et al (2010) Gaze allocation in a dynamic situation : Effects of social status and speaking. *Cognition* 117:319–331. DOI: 10.1016/j.cognition.2010.09.003
- Foulsham T, & Lock M (2015) How the Eyes Tell Lies: Social Gaze During a Preference Task. *Cognitive Science* 39:1704–1726. DOI: 10.1111/cogs.12211
- Fussell S, Setlock L, Yang J, et al (2004) Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction* 19:273–309. DOI: 10.1207/s15327051hci1903_3

- Fussell SR, Setlock LD, & Kraut RE (2003) Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of the conference on Human factors in computing systems - CHI '03* 513. DOI: 10.1145/642611.642701
- Gallagher-Mitchell T, Simms V, & Litchfield D (2017) a. *The Quarterly Journal of Experimental Psychology* 0218:1-30. DOI: 10.1080/17470218.2017.1335335
- Gallup AC, Chong A, & Couzin ID (2012) The directional flow of visual information transfer between pedestrians. *Biology Letters* 8:520-522
- García P, Olvido J, Ehlers KR, & Tylén K (2017) Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto-declaratives. *Language Sciences* 54:73-81. DOI: 10.1016/j.langcom.2016.10.007
- George N, & Conty L (2008) Facing the gaze of others. *Neurophysiologie Clinique* 38:197-207
- Gergle D, Kraut R, & Fussell S (2013) Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction* 28:1-39. DOI: 10.1080/07370024.2012.678246
- Gibaldi A, Vanegas M, Bex PJ, & Maiello G (2017) Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. 923-946. DOI: 10.3758/s13428-016-0762-9
- Gibson J, & Pick A (1963) Perception of another's looking behavior. *The American Journal of Psychology* 76:386-394
- Gonzalez CC, Causer J, Miall RC, et al (2017) Identifying the causal mechanisms of the quiet eye. *European Journal of Sport Science* 17:74-84. DOI: 10.1080/17461391.2015.1075595
- Griffin ZM (399AD) Why Look? Reasons for Eye Movements Related to Language Production. *The interface of language, vision, and action: Eye movements and the visual world* Ferreira,:Eye-247
- Griffin ZM (206AD) Gaze durations during speech reflect word selection and phonological encoding. *Cognition* 82:1-16
- Griffin ZM, & Bock K (2000) What the eyes say about speaking. *Psychological science: a journal of the American Psychological Society / APS* 11:274-279. DOI: 10.1111/1467-9280.00255
- Gullberg M (2003) Eye movements and gestures in human interaction. In: *The Mind's Eye*. pp 685-703

- Guo J, & Feng G (2013) How Eye Gaze Feedback Changes Parent-Child Joint Attention in Shared Storybook Reading? An Eye-Tracking Intervention Study. *Eye Gaze in Intelligent User Interfaces: Gaze-based Analyses, Models and Applications* 9–21. DOI: 10.1007/978-1-4471-4784-8_2
- Gupta K, Lee GA., & Billingham M (2016) Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 22:2413–2422. DOI: 10.1109/TVCG.2016.2593778
- Gutwin C, & Penner R (2002) Improving interpretation of remote gestures with telepointer traces. *Proceedings of the 2002 ACM conference on Computer supported cooperative work - CSCW '02* 49. DOI: 10.1145/587078.587086
- Hains SMJ, & Muir DW (1996) Infant Sensitivity to Adult Eye Direction. *Child Development*. DOI: 10.1111/j.1467-8624.1996.tb01836.x
- Hanna JE, & Brennan SE (2007) Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57:596–615. DOI: 10.1016/j.jml.2007.01.008
- Hansen DW, & Ji Q (2010) In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32:478–500. DOI: 10.1109/TPAMI.2009.30
- Harrer A, Schlosser C, Schlieker-Steens P, & Kienle A (2015) Here's looking at you, kid - Can gaze awareness help to learn to learn together in collaborative problem solving? *Proceedings - IEEE 15th International Conference on Advanced Learning Technologies* 190–194. DOI: 10.1109/ICALT.2015.135
- Hartridge H, & Thomson LC (1942) Method of investigating eye movements. *British Journal of Ophthalmology* 581–591
- Higuch K, Yonetani R, & Sato Y (2016) Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* 5180–5190. DOI: 10.1145/2858036.2858438
- Holmqvist K, Nyström M, & Mulvey F (2012) Eye tracker data quality: What it is and how to measure it. *Eye Tracking Research and Applications Symposium (ETRA)* 1:45–52
- Holtzblatt K, & Beyer H (2014) Contextual Design: Evolved. Synthesis Lectures on Human-Centered Informatics. DOI:

10.2200/S00597ED1V01Y201409HCI024

- Holtzblatt K, Wendell JB, & Wood S (2005) Rapid Contextual Design. *Ubiquity* 2005:3-3. DOI: 10.1145/1066348.1066325
- Houben MMJ, Goumans J, & Van Der Steen J (2006) Recording three-dimensional eye movements: Scleral search coils versus video oculography. *Investigative Ophthalmology and Visual Science* 47:179-187. DOI: 10.1167/iovs.05-0234
- Hyrskykari A, Majaranta P, Aaltonen A, & Raiha K-J (2000) Design issues of iDICT: a gaze-assisted translation aid. Proceedings of the 2000 symposium on Eye tracking research & applications. DOI: 10.1145/355017.355019
- Ishii H, & Kobayashi M (1992) ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92* 525-532. DOI: 10.1145/142750.142977
- ISO 9241-11 (2018). Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts.
- Itier RJ, & Batty M (2009) Neural basis of eye and gaze processing: the core of social cognition. *Neurosci Biobehav Rev* 33:843-863. DOI: 10.1167/iovs.07-1072.Complement-Associated
- Jack RE, Scheepers C, Fiset D, et al (2008) Culture Shapes How We Look at Faces. *PLoS ONE* 3:e3022. DOI: 10.1371/journal.pone.0003022
- Jarodzka H, Balslev T, Holmqvist K, et al (2012) Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science* 40:813-827. DOI: 10.1007/s11251-012-9218-5
- Jarodzka H, Holmqvist K, & Gruber H (2017) Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research* 10:1-18. DOI: 10.16910/jemr.10.1.3
- Jarodzka H, Scheiter K, Gerjets P, et al (2009) How to Convey Perceptual Skills by Displaying Experts' Gaze Data. *Cogsci* 2920-2925. DOI: 10.1016/j.learninstruc.2009.02.019.Lindsey
- Jarodzka H, Van Gog T, Dorr M, et al (2013) Learning to see: Guiding students' attention via a Model's eye movements fosters learning. *Learning and Instruction* 25:62-70. DOI: 10.1016/j.learninstruc.2012.11.004

- Jenkins J, & Langton SRH (2003) Configural processing in the perception of eye-gaze direction. *Perception* 32:1181-1188. DOI: 10.1068/p3398
- Jermann P, Nüssli M, & Li W (2010) Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris. *BCS '10 Proceedings of the 24th BCS Interaction Specialist Group Conference* 36-44
- Johansen R (1988) Groupware: computer support for business teams. *Series in communication technology and society* xviii, 205
- John B, Sridharan S, & Bailey R (2014) Collaborative eye tracking for image analysis. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '14* 239-242. DOI: 10.1145/2578153.2578215
- Kangas J, Akkil D, Rantala J, et al (2014) Gaze gestures and haptic feedback in mobile devices. *Proceedings of the SIGCHI conference on Human Factors in computing systems* 435-438. DOI: 10.1145/2556288.2557040
- Kendon A (1967) Some functions of gaze direction in social interaction. *Acta psychologica* 16:22-63
- Kirk D, & Fraser D (2006) Comparing remote gesture technologies for supporting collaborative physical tasks. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* 1191-1200. DOI: 10.1145/1124772.1124951
- Kobayashi H, & Kohshima S (2001) Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of human evolution* 40:419-435. DOI: 10.1006/jhev.2001.0468
- Kok EM, Aizenman AM, Võ ML-H, & Wolfe JM (2017) Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision* 17:2. DOI: 10.1167/17.12.2
- Kumar M, Paepcke A, & Winograd T (2007) EyePoint. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* 421. DOI: 10.1145/1240624.1240692
- Kushalnagar RS, Kushalnagar P, & Manganelli G (2012) Collaborative Gaze Cues for Deaf Students. Dual Eye Tracking Workshop. DOI: 10.13140/2.1.3415.1364
- Laidlaw KEW, Foulsham T, Kuhn G, & Kingstone A (2011) Potential social interactions are important to social attention. 108:. DOI: 10.1073/pnas.1017022108

- Lambert RH, Monty RA, & Hall RJ (1974) High-speed data processing and unobtrusive monitoring of eye movements. *Behavior Research Methods & Instrumentation* 6:525–530. DOI: 10.3758/BF03201340
- Land MF (2006) Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research* 25:296–324. DOI: 10.1016/j.preteyeres.2006.01.002
- Land MF, & McLeod P (2000) From eye movements to actions: Batsmen hit the ball. *Nature Neuroscience* 3:1340–1345. DOI: 10.1038/81887
- Lankes M, Maurer B, & Stiglbauer B (2016) An Eye for an Eye : Gaze Input in Competitive Online Games and its Effects on Social Presence. Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology
- Lankes M, Rammer D, & Maurer B (2017) Eye Contact: Gaze as a Connector Between Spectators and Players in Online Games. *Lecture Notes in Computer Science* 10507 LNCS:507–509. DOI: 10.1007/978-3-319-66715-7
- Lee J, Park H, Lee S, et al (2011) Design and Implementation of an Augmented Reality System Using Gaze Interaction. *2011 International Conference on Information Science and Applications* 1–8. DOI: 10.1109/ICISA.2011.5772406
- Li HZ (2004) Culture and gaze direction in conversation. *Rask* 20:3–26
- Li J, Manavalan ME, D’Angelo S, & Gergle D (2016) Designing Shared Gaze Awareness for Remote Collaboration. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* 325–328. DOI: 10.1145/2818052.2869097
- Li W, Nüssli M-A, & Jermann P (2010) Gaze quality assisted automatic recognition of social contexts in collaborative Tetris. *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on - ICMI-MLMI ’10* 1. DOI: 10.1145/1891903.1891914
- Li Z, Akkil D, & Raisamo R (2019) Gaze Augmented Hand-Based Kinesthetic Interaction: What You See Is What You Feel. *IEEE Transactions on Haptics*. DOI: 10.1109/toh.2019.2896027
- Litchfield D, & Ball LJ (2011) Rapid communication using another’s gaze as an explicit aid to insight problem solving. *Quarterly Journal of Experimental Psychology* 64:649–656. DOI: 10.1080/17470218.2011.558628

- Litchfield D, Ball LJ, Donovan T, et al (2010) Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied* 16:251–262. DOI: 10.1037/a0020082
- Litchfield D, Ball LJ, Donovan T, et al (2008) Learning from others : effects of viewing another person ' s eye movements while searching for chest nodules. *Imaging* 9:1–9. DOI: 10.1117/12.768812
- Liu C, Kay DL, & Chai JC (2011) Awareness of partners eye gaze in situated referential grounding: An empirical study. In: 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction. pp 33–38
- Liu Y, Hsueh P, & Lai J (2009) Who is the Expert? Analyzing Gaze Data to Predict Expertise Level in Collaborative Applications. *Training* 0:0–3
- Loomis JM, Kellyô JW, Bailenson MPJN, & Beall AC (2008) Psychophysics of perceiving eye-gaze and head direction with peripheral vision : Implications for the dynamics of eye-gaze behavior. 37:1443–1458. DOI: 10.1068/p5896
- Macdonald RG, & Tatler BW (2012) The effect of social roles on gaze cue utilisation in a real - world collaboration. 942–947
- Macdonald RG, & Tatler BW (2017) Do as eye say : Gaze cueing and language in a real-world social interaction. 13:1–12. DOI: 10.1167/13.4.6.doi
- Majaranta P, & Bulling A (2014) Eye Tracking and Eye-Based Human – Computer Interaction. *Advances in physiological computing* 39–65. DOI: 10.1007/978-1-4471-6392-3
- Majaranta P, & R ih a K (2002) Twenty years of eye typing: systems and design issues. *Proceedings of the 2002 symposium on Eye tracking research & applications* 1:15–22. DOI: 10.1145/507072.507076
- Marti S, Bayet L, & Dehaene S (2015) Subjective report of eye fixations during serial search. *Consciousness and Cognition* 33:1–15. DOI: 10.1016/j.concog.2014.11.007
- Martinez-Conde S, Macknik SL, & Hubel DH (2004) The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience* 5:229–240. DOI: 10.1038/nrn1348
- Mason L, Pluchino P, & Tornatora MC (2015) Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology* 41:172–187. DOI: 10.1016/j.cedpsych.2015.01.004

- Mason L, Pluchino P, & Tornatora MC (2016) Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology* 47:1083–1095. DOI: 10.1111/bjet.12271
- Matin E (1974) Saccadic suppression: A review and an analysis. *Psychological Bulletin* 81:899–917. DOI: 10.1037/h0037368
- Mattessich PW, Monsey BR, & Murray-Close M (2001) Collaboration: what makes it work. A review of research literature on factors influencing successful collaboration. For the collaboration handbook
- Maurer B, Aslan I, Wuchse M, et al (2015) Gaze-Based Onlooker Integration: Exploring the In-Between of Active Player and Passive Spectator in Co-Located Gaming. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15. ACM Press, New York, New York, USA, pp 163–173. DOI: 10.1145/2793107.2793126
- Maurer B, Lankes M, Stiglbauer B, & Tscheligi M (2014a) EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game. Springer Berlin Heidelberg, Berlin, Heidelberg. DOI: 10.1007/978-3-319-46100-7_9
- Maurer B, Trösterer S, Gärtner M, et al (2014b) Shared Gaze in the Car. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14* 1–6. DOI: 10.1145/2667239.2667274
- McDonnell GP, Mills M, Marshall JE, et al (2017) You detect while I search: examining visual search efficiency in a joint search task. *Visual Cognition* 0:1–18. DOI: 10.1080/13506285.2017.1386748
- Mennie N, Hayhoe M, & Sullivan B (2007) Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research* 179:427–442. DOI: 10.1007/s00221-006-0804-0
- Messmer N, Leggett N, Prince M, & McCarley JS (2017) Gaze Linking in Visual Search: A Help or a Hindrance? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61:1376–1379. DOI: 10.1177/1541931213601828
- Meyer AS, Sleiderink AM, & Levelt WJM (1998) Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66:25–33. DOI: 10.1016/S0010-0277(98)00009-2
- Meyer CH, Lasker AG, & Robinson DA (1985) The upper limit of human smooth pursuit velocity. *Vision Research* 25:561–563. DOI:

10.1016/0042-6989(85)90160-9

- Monk AF, & Gale C (2002) A Look Is Worth a Thousand Words : Full Gaze Awareness in Video- Mediated Conversation. *Discourse Processes* 33:257–278. DOI: 10.1207/S15326950DP3303
- Moore LJ, Vine SJ, Smith AN, et al (2014) Quiet eye training improves small arms maritime marksmanship. *Military Psychology* 26:355–365. DOI: 10.1037/mil0000039
- Mowrer OH, Ruch TC, & Miller NE (2017) The corneo-retinal potential difference as the basis of the galvanometric method of recording eye movements. *American Journal of Physiology-Legacy Content* 114:423–428. DOI: 10.1152/ajplegacy.1935.114.2.423
- Müller R, Helmert JR, & Pannasch S (2014) Limitations of gaze transfer: Without visual context, eye movements do not help to coordinate joint action, whereas mouse movements do. *Acta Psychologica* 152:19–28. DOI: 10.1016/j.actpsy.2014.07.013
- Müller R, Helmert JR, Pannasch S, & Velichkovsky BM (2013) Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *Quarterly Journal of Experimental Psychology* 66:1302–1316. DOI: 10.1080/17470218.2012.737813
- Müller R, Helmert JR, Pannasch S, & Velichkovsky BM (2011) Following closely? The effects of viewing conditions on gaze versus mouse transfer in remote cooperation. *Proceedings of the Dual Eye-Tracking in CSCW (DUET)* 212–222
- Nalanagula D, Greenstein JS, & Gramopadhye AK (2006) Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics* 36:289–300. DOI: 10.1016/j.ergon.2005.11.008
- Nardi BA, Kuchinsky A, Whittaker S, et al (1996) Video-as-Data : Technical and Social Aspects of a Collaborative Multimedia Application. 73–100
- Neale DC, Carroll JM, & Rosson MB (2004) Evaluating computer-supported cooperative work. *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04* 112. DOI: 10.1145/1031607.1031626
- Neider MB, Chen X, Dickinson CA, et al (2010) Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review* 17:718–724. DOI: 10.3758/PBR.17.5.718

- Newn J (2018) Enabling Intent Recognition Through Gaze Awareness in User Interfaces Exploring Social Gaze in Human-Computer Interaction View project. DOI: 10.1145/3170427.3173028
- Newn J, Allison F, Velloso E, & Vetere F (2018) Looks Can Be Deceiving : Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'18* 1-12. DOI: 10.1145/3173574.3173835
- Newn J, Velloso E, Allison F, et al (2017) Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games
- Niehorster D, Cornelissen T, Hooge I, & Holmqvist K (2017) Searching with and against each other. *Journal of Vision* 17:222. DOI: 10.1167/17.10.222
- Nüssli M-A, & Jermann P (2012) Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12* 1125. DOI: 10.1145/2145204.2145371
- Nyström M, Andersson R, Holmqvist K, & van de Weijer J (2013) The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods* 45:272-88. DOI: 10.3758/s13428-012-0247-4
- Nyström M, Niehorster DC, Cornelissen T, & Garde H (2017) Real-time sharing of gaze data between multiple eye trackers—evaluation, tools, and advice. *Behavior Research Methods* 49:1310-1322. DOI: 10.3758/s13428-016-0806-1
- O'Hara K, Black A, & Lipson M (2006) Everyday practices with mobile video telephony. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* 871-880. DOI: 10.1145/1124772.1124900
- O'Shea RP (1991) Thumb's rule tested: visual angle of thumb's width is about 2 deg. *Perception* 20:415-418. DOI: 10.1068/p200415
- Olson GM, & Olson JS (2000) Distance matters. *Human-Computer Interaction* 15:139-178. DOI: 10.1207/S15327051HCI1523_4
- Olson JS, & Olson GM (2006) Bridging Distance: Empirical Studies of Distributed Teams. *Advances in Management Information Systems*
- Panchuk D, Vickers JN, & Hopkins WG (2017) Quiet eye predicts goaltender success in deflected ice hockey shots†. *European Journal of*

Sport Science 17:93–99. DOI: 10.1080/17461391.2016.1156160

- Pelz JB, & Canosa R (2001) Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research* 41:3587–3596. DOI: 10.1016/S0042-6989(01)00245-0
- Pfeuffer K, Alexander J, & Gellersen H (2016) GazeArchers: Playing with Individual and Shared Attention in a Two-Player Look&Shoot Tabletop Game. *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia - MUM '16* 213–216. DOI: 10.1145/3012709.3012717
- Qian YY, & Teather RJ (2017) The eyes don't have it. *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17* 91–98. DOI: 10.1145/3131277.3132182
- Qvarfordt P, Beymer D, & Zhai S (2005) RealTourist—A Study of Augmenting Human-Human and Human-Computer Dialogue with Eye-Gaze Overlay. *Human-Computer Interaction-INTERACT* 767–780
- Qvarfordt P, & Zhai S (2005) Conversing with the user based on eye-gaze patterns. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05* 221. DOI: 10.1145/1054972.1055004
- Räihä K-J (2013) Life in the Fast Lane: Effect of Language and Calibration Accuracy on the Speed of Text Entry by Gaze. *8119*:402–417. DOI: 10.1007/978-3-642-40498-6
- Reingold EM, & Sheridan H (2012) Eye movements and visual expertise in chess and medicine. In: *The Oxford Handbook of Eye Movements*
- Ricciardelli P, Baylis G, & Driver J (2000) The positive and negative of human expertise in gaze perception. *Cognition* 77:1–14. DOI: 10.1016/S0010-0277(00)00092-5
- Richardson DC, & Dale R (2005) Looking To Understand : The Coupling Between Speakers ' and Listeners ' Eye Movements and Its Relationship to Discourse Comprehension. *Science* 29:1045–1060. DOI: 10.1207/s15516709cog0000
- Roberts D, Wolff R, Rae J, et al (2009) Communicating Eye-Gaze across a distance: comparing an Eye-Gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. *Proceedings - IEEE Virtual Reality* 135–142. DOI: 10.1109/VR.2009.4811013
- Robinson D (1963) Movement Using a Scieral Search in a Magnetic Field. *IEEE Transactions on Bio-Medical Electronics* 10:137–145. DOI:

10.1109/TBMEL.1963.4322822

- Sadasivan S, Greenstein JS, Gramopadhye AK, & Duchowski AT (2005) Use of eye movements as feedforward training for a synthetic aircraft inspection task. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05* 141. DOI: 10.1145/1054972.1054993
- Schlösser C, Schlieker-steens P, & Kienle A (2015) Using Real-Time Gaze Based Awareness Methods to Enhance Collaboration. In: Baloian N, Zorian Y, Taslakian P, Shoukouryan S (eds) *Collaboration and Technology*. Springer International Publishing, Cham, pp 19-27
- Schlösser C, Schröder B, Cedli L, & Kienle A (2018) Beyond Gaze Cursor: Exploring Information-based Gaze Sharing in Chat. In: *Proceedings of the Workshop on Communication by Gaze Interaction*. p 10:1--10:5
- Schneider B, & Pea R (2013) Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning* 8:375-397. DOI: 10.1007/s11412-013-9181-4
- Schneider B, & Pea R (2014) The Effect of Mutual Gaze Perception on Students ' Verbal Coordination. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)* 138-144
- Sebanz N, Bekkering H, & Knoblich G (2006) Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10:70-76. DOI: 10.1016/j.tics.2005.12.009
- Seppänen M, & Gegenfurtner A (2012) Seeing through a teacher's eyes improves students' imaging interpretation. *Medical Education* 46:1112-1113. DOI: 10.1111/medu.12029
- Sharma K, D'Angelo S, Gergle D, & Dillenbourg P (2016) Visual augmentation of deictic gestures in MOOC videos. *Proceedings of International Conference of the Learning Sciences, ICLS* 1:202-209
- Sharma K, Jermann P, & Dillenbourg P (2015a) Displaying teacher's gaze in a MOOC: Effects on students' video navigation patterns. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp 325-338
- Sharma K, Jermann P, & Dillenbourg P (2015b) Displaying teacher's gaze in a MOOC: Effects on students' video navigation patterns. *Lecture Notes in Computer Science* 9307:325-338. DOI: 10.1007/978-3-319-24258-3_24

- Shikida M (2016) Conveying Gaze-Awareness by Using a Faint Light with a Video-Conferencing System. *International Journal of Future Computer and Communication* 5:38–42. DOI: 10.18178/ijfcc.2016.5.1.440
- Shimojo S, Simion C, Shimojo E, & Scheier C (2003) Gaze bias both reflects and influences preference. *Nature Neuroscience* 6:1317–1322. DOI: 10.1038/nn1150
- Siirtola H, Špakov O, Istance H, & Rähä K-J (2019) Shared Gaze in Collaborative Visual Search. *International Journal of Human-Computer Interaction* 1–13. DOI: 10.1080/10447318.2019.1565746
- Špakov O, Istance H, Siirtola H, & Rähä K-J (2016) GazeLaser : A Hands-Free Highlighting Technique for Presentations. *CHI Extended Abstracts on Human Factors in Computing Systems* 2648–2654. DOI: 10.1145/2851581.2892504
- Špakov O, Istance H, Viitanen T, et al (2018) Enabling Unsupervised Eye Tracker Calibration by School Children through Games. DOI: 10.1145/3204493.3204534
- Špakov O, Siirtola H, Istance H, & Rähä K-J (2017) Visualizing the Reading Activity of People Learning to Read. *Journal of Eye Movement Research* 10:1–12. DOI: 10.16910/JEMR.10.5.5
- Sridharan S, McNamara A, & Grimm C (2012) Subtle gaze manipulation for improved mammography training. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12* 1:75. DOI: 10.1145/2168556.2168568
- Stein R, & Brennan SE (2004) Another person's eye gaze as a cue in solving programming problems. *Proceedings of the 6th international conference on Multimodal interfaces* 9–15. DOI: <http://doi.acm.org/10.1145/1027933.1027936>
- Stellmach S, & Dachselt R (2012) Look & Touch: Gaze-supported Target Acquisition. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)* 2981–2990. DOI: 10.1145/2207676.2208709
- Symons LA, Lee K, Cedrone CC, & Nishimura M (2008) What are you looking at? Acuity for triadic eye gaze. *Journal of general psychology* 131:451–69.
- Theeuwes J, Kramer F, Hahn S, & Irwin DE (1998) Our eyes do not always go where we want them to go: capture of gaze by new objects. *Psychological science* 9:379–385. DOI: 10.2307/40063323

- Tobii Technology (2017) Tobii Pro Lab User Manual. 1-40
- Tobii Technology (2016) Tobii TX300 Technical Specification. Tobii Technology
- Tobii Technology (2011) Accuracy and precision test method for remote eye trackers. Technology
- Trösterer S, Gärtner M, Wuchse M, et al (2015a) Four eyes see more than two: Shared gaze in the car. *Lecture Notes in Computer Science* 9297:331-348
- Trösterer S, Wuchse M, Döttlinger C, et al (2015b) Light my way: visualizing shared gaze in the car. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15* 196-203. DOI: 10.1145/2799250.2799258
- Tufft MRA, Gobel MS, & Richardson DC (2015) Social Eye Cue: How Knowledge Of Another Person 's Attention Changes Your Own. *Proceedings of the Cognitive Science Society* 2440-2445
- van Gog T, Jarodzka H, Scheiter K, et al (2009) Attention guidance during example study via the model's eye movements. *Computers in Human Behavior* 25:785-791. DOI: 10.1016/j.chb.2009.02.007
- van Marlen T, van Wermeskerken M, Jarodzka H, & van Gog T (2016) Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior* 65:448-459. DOI: 10.1016/j.chb.2016.08.041
- van Rheden V, Maurer B, Smit D, et al (2017) LaserViz: Shared Gaze in the Co-Located Physical World. *Proceedings of the Tenth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '17* 191-196. DOI: 10.1145/3024969.3025010
- Velichkovsky B, Pomplun M, & Rieser J (1996) Attention and communication: Eye-movement-based research paradigms. *Advances in Psychology* 116:125-154. DOI: 10.1016/S0166-4115(96)80074-4
- Velichkovsky BM (1995) Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition* 3:199-223. DOI: 10.1075/pc.3.2.02vel
- Vertegaal R (1999) The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems ACM* 294-301. DOI: 10.1145/302979.303065

- Vickers JN (1992) Gaze control in golf putting. *Perception* 21:117–132. DOI: 10.1068/p210117
- Vickers JN (1996) Visual control when aiming at a far target. *Journal of Experimental Psychology: Human Perception and Performance* 22:342–354. DOI: 10.1037/0096-1523.22.2.342
- Vickers JN (2016) Origins and current issues in Quiet Eye research. *Current Issues in Sport Science* 1:3. DOI: 10.15203/CISS
- Vickers JN (2007) Perception, cognition, and decision training: The quiet eye in action.
- Vickers JN, & Adolphe R (1997) Gaze Behavior During a Ball Tacking and Aiming Skill. *International journal of sports vision* 4:18–27
- Vickers JN, Vandervies B, Kohut C, & Ryley B (2017) Quiet eye training improves accuracy in basketball field goal shooting, 1st edn. Elsevier B.V.
- Vidal M, Bulling A, & Gellersen H (2015) The Royal Corgi: Exploring Social Gaze Interaction for Immersive Gameplay. *Proceedings of the SIGCHI conference on Human Factors in computing systems* 115–124
- Vidal M, Bulling A, & Gellersen H (2013a) Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. ... of the 2013 ACM international joint ... 439–448. DOI: 10.1145/2493432.2493477
- Vidal M, Pfeuffer K, Bulling A, & Gellersen HW (2013b) Pursuits: eye-based interaction with moving targets. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* 3147. DOI: 10.1145/2468356.2479632
- Vidal M, Turner J, Bulling A, & Gellersen H (2012) Wearable eye tracking for mental health monitoring. *Computer Communications* 35:1306–1311. DOI: 10.1016/j.comcom.2011.11.002
- Vine SJ, Moore LJ, & Wilson MR (2011) Quiet eye training facilitates competitive putting performance in elite golfers. *Frontiers in Psychology* 2:1–9. DOI: 10.3389/fpsyg.2011.00008
- Vine SJ, Moore LJ, & Wilson MR (2014) Quiet eye training: The acquisition, refinement and resilient performance of targeting skills. *European Journal of Sport Science* 14:. DOI: 10.1080/17461391.2012.683815
- Vine SJ, & Wilson MR (2010) Quiet eye training: Effects on learning and performance under pressure. *Journal of Applied Sport Psychology*

22:361–376. DOI: 10.1080/10413200.2010.495106

Vine SJ, & Wilson MR (2011) The influence of quiet eye training and pressure on attention and visuo-motor control. *Acta Psychologica* 136:340–346. DOI: 10.1016/j.actpsy.2010.12.008

Wade NJ (2015) How Were Eye Movements Recorded Before Yarbus? *Perception* 44:851–883. DOI: 10.1177/0301006615594947

Wahn B, Schwandt J, Krüger M, et al (2016) Multisensory teamwork: using a tactile or an auditory display to exchange gaze information improves performance in joint visual search. *Ergonomics* 59:781–795. DOI: 10.1080/00140139.2015.1099742

Walker-Smith GJ, Gale AG, & Findlay JM (1977) Eye Movement Strategies Involved in Face Perception. *Perception* 6:313–326. DOI: 10.1068/p060313

Walls GL (1962) The Evolutionary History of Eye Movements. *Vision Research* 2:69–80

Whitmire E, Trutoiu L, Cavin R, et al (2016) EyeContact: Scleral Coil Eye Tracking for Virtual Reality. In: Proceedings of the ACM International Symposium on Wearable Computers. pp 184–191

Williams AM, Singer RN, & Frehlich SG (2002) Quiet eye duration, expertise, and task complexity in near and far aiming tasks. *Journal of Motor Behavior* 34:197–207. DOI: 10.1080/00222890209601941

Wilson MR, Vine SJ, Bright E, et al (2011) Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study. *Surgical Endoscopy and Other Interventional Techniques* 25:3731–3739. DOI: 10.1007/s00464-011-1802-2

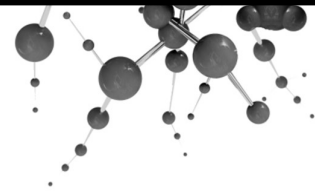
Wood G, & Wilson MR (2011) Quiet-eye training for soccer penalty kicks. *Cognitive Processing* 12:257–266. DOI: 10.1007/s10339-011-0393-0

Wood G, & Wilson MR (2012) Quiet-eye training, perceived control and performing under pressure. *Psychology of Sport and Exercise* 13:721–728. DOI: 10.1016/j.psychsport.2012.05.003

Wu D-A, Shimojo S, Wang SW, & Camerer CF (2012) Shared Visual Attention Reduces Hindsight Bias. *Psychological Science* 23:1524–1533. DOI: 10.1177/0956797612447817

Yamani Y, Neider MB, Kramer AF, & McCarley JS (2017) Characterizing the efficiency of collaborative visual search with systems factorial

- technology. *Archives of Scientific Psychology* 5:1–9. DOI: 10.1037/arc0000030
- Young LR, & Sheena D (1970) Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation* 7:397–429. DOI: 10.3758/BF03201553
- Yu C, Xu Y, Liu B, & Liu Y (2014) Can you SEE me now?: A measurement study of mobile video calls. *INFOCOM, 2014 Proceedings IEEE* 1456–1464. DOI: 10.1109/INFOCOM.2014.6848080
- Zhang Y, Pfeuffer K, Chong MK, et al (2017) Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21:173–186. DOI: 10.1007/s00779-016-0969-x
- Zuckerman M, Miserandino M, & Bernieri F (2007) Civil Inattention Exists – in Elevators. *Personality and Social Psychology Bulletin* 9:578–586. DOI: 10.1177/0146167283094007



Paper 1

Deepak Akkil, Poika Isokoski, Jari Kangas, Jussi Rantala, and Roope Raisamo. 2014. TraQuMe: a tool for measuring the gaze tracking quality. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA '14). 327-330. DOI: [10.1145/2578153.2578192](https://doi.org/10.1145/2578153.2578192)

© ACM 2014, Reprinted with permission.

TraQuMe: A Tool for Measuring the Gaze Tracking Quality

Deepak Akkil, Poika Isokoski, Jari Kangas, Jussi Rantala, Roope Raisamo

Tampere Unit for Computer-Human Interaction

School of Information Sciences

University of Tampere, Finland

{deepak.akkil, poika.isokoski, jari.kangas, jussi.e.rantala, roope.raisamo}@uta.fi

Abstract

Consistent measuring and reporting of gaze data quality is important in research that involves eye trackers. We have developed TraQuMe: a generic system to evaluate the gaze data quality. The quality measurement is fast and the interpretation of the results is aided by graphical output. Numeric data is saved for reporting of aggregate metrics for the whole experiment. We tested TraQuMe in the context of a novel hidden calibration procedure that we developed to aid in experiments where participants should not know that their gaze is being tracked. The quality of tracking data after the hidden calibration procedure was very close to that obtained with the Tobii's T60 trackers built-in 2 point, 5 point and 9 point calibrations.

CR Categories: H.5.2. User interfaces: Evaluation.

Keywords: Gaze tracking; gaze interaction.

1 Introduction

The utility of the gaze tracker is dependent on the quality of the gaze data that it can generate. However, little emphasis has been put on measuring and reporting the gaze tracker data quality. The issue has been discussed by Holmqvist *et al.* [2012] and Nyström *et al.* [2013]. Tracker manufacturers have also given recommendations for tracker performance measurement [Tobii, 2011]. Yet, no manufacturer independent measurement tools are available. Some tracker manufacturers offer calibration verification tools as a part of their software bundle [SMI, 2012]. Unfortunately, the results may not be compatible across tracker manufacturers.

Until now, the most common way to report gaze data quality has been to refer to the numbers reported by the tracker manufacturer [Holmqvist *et al.* 2012]. However, the manufacturer specifications may deviate from results achievable in practical lab environment. For example, Morgante *et al.* tested temporal and spatial accuracy of Tobii T60XL gaze tracker and found that the practical accuracy was worse than what was mentioned in device specification sheet [2012]. Practical tracking quality also shows person to person variation. Hence it is important to measure the actual tracking quality for each participant in the experiment.

Spatial accuracy and precision are the two most important measures of the quality of gaze data. Accuracy is defined as the closeness of the measured gaze point to the point that the tracked eye is looking at. Precision is defined as the ability of the tracker to re-produce the measurement [Nystrom *et al.* 2013].

There are many goals that eye tracking quality metrics could be used for: 1) to document the random variables that vary in experiments due to participant selection and calibration errors, 2) as an exclusion criterion for “bad” data, and 3) to compare different calibration methods.

Our tool, TraQuMe¹ (short for Tracking Quality Measurement), is a light-weight, generic and tracker independent data quality measurement software. Since the measurements are intended to be run multiple times during a lab session, the speed of measurement and ease of use of the software were central design criteria. The measurement should happen in a few seconds so that the participants would not be burdened by too much extra work. Trackstick [Blignaut & Beelders 2012] is a similar quality measurement tool but it is compatible only with the Tobii gaze trackers. We are not aware of other tracker independent tools.

2 System Description

TraQuMe works by showing fixation targets on various screen locations one after the other just like in a typical system controlled calibration routine. User's gaze samples are collected when he or she fixates on the targets. TraQuMe then displays a visualization of the samples collected (Figure 2) and computes the spatial accuracy and precision for both individual eyes and the average binocular gaze point.

2.1 Technical details

TraQuMe is a Microsoft Windows form application built using .NET 4.5 framework and is developed on top of the COGAIN ETU-Driver platform [Bates and Spakov 2006]. It allows the operator to configure the number of validation targets, their on-screen positions, data collection duration etc. TraQuMe also provides an option to configure the background screen color and the fixation point color.

TraQuMe fixation stimulus is shown in Figure 1. The data collection and the associated animation begin 500ms after the point is in position. This delay was to ensure that eyes have arrived at the target before data recording starts. At the time of data collection, the larger circle shrinks to the center and expands back.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
ETRA 2014, March 26 – 28, 2014, Safety Harbor, Florida, USA.
2014 Copyright held by the Owner/Author.
ACM 978-1-4503-2751-0/14/03

¹TraQuMe can be downloaded from:
<http://www.uta.fi/sis/tauchi/virg/traqueme.html>

For individual eyes the accuracy is calculated as

$$accuracy = \sqrt[2]{(X_{true} - x_{mean})^2 + (Y_{true} - y_{mean})^2} \quad (1)$$

X_{true} , Y_{true} are the X and Y coordinates of the screen fixation point and x_{mean} , y_{mean} are the mean gaze point in the collected data [Tobii, 2011].

The standard deviation precision is calculated as

$$precision = \sqrt[2]{(SD(Gaze.X))^2 + (SD(Gaze.Y))^2} \quad (2)$$

$SD(Gaze.X)$ and $SD(Gaze.Y)$ are the standard deviations in the collected gaze data along X and Y direction respectively [Tobii, 2011]. The binocular quality measures are based on the means of X and Y coordinates for both eyes. When only a single eye is visible, that data is used in the computation of binocular metrics.

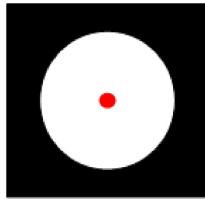


Figure 1. Fixation target (colors are user adjustable).

3 How to use TraQuMe

The recommended way to use TraQuMe is soon after the initial calibration process, or in between blocks in an experiment to check if the data quality is high enough to continue. Running it at the end of the experiment gives an additional data point for estimating the average data quality during the experiment.

Accuracy and precision are computed in centimeters and screen pixels. TraQuMe can also do the conversion to degrees of visual angle if the experimenter enters the distance between the display and the eyes. Ideally the conversion should be automatic, but all trackers do not offer the distance data.

Different experiments and applications have different quality requirements. Mean and maximum values of the quality metrics of a set of stimuli spread over the whole trackable area are a good indication of the overall quality. However, some applications utilize only a part of the trackable area. In such applications even one verification point may be sufficient.

A 9-point gaze quality measurement process with a data collection duration of 1.5 seconds per point takes approximately 30 seconds to complete. For easy visual detection of outliers a gaze data visualization is shown. If outliers are present, they should be taken into consideration when making decisions based on the statistics that TraQuMe computes. The tool calculates spatial accuracy, precision, and number of samples for each target. The number of samples should be related to the data collection duration and the tracker's sample rate. Missing samples indicate periods when the tracker could not see the eyes. This is a third quality metric that is independent of accuracy and precision.

As an example of how TraQuMe output can be used in practice, we evaluated a new calibration procedure for the Tobii T60 eye tracker as described in the next section.

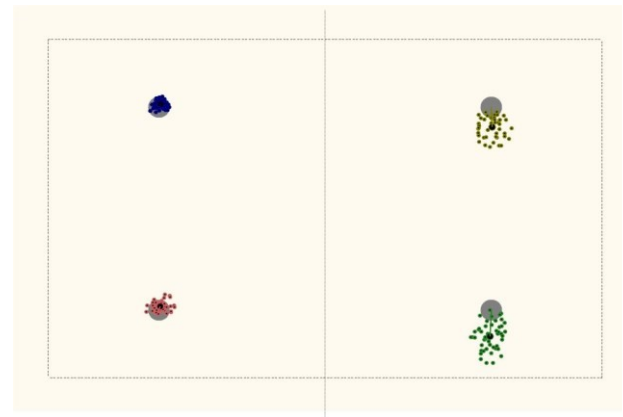


Figure 2. Visualization of good (LEFT) and bad (RIGHT) quality gaze data. The dashed frame indicates the edges of the display.

4 Form Based Calibration Routine

Nyström *et al.* note that in many experiments it is desirable not to make the participants aware that their eyes are being tracked when the experiment is being conducted [2013]. It is possible that knowledge of being tracked could affect the gaze behavior of the participants. Obviously, secretly tracking the gaze is ethically problematic. However, benign deception is sometimes justified, and allowed by ethical review boards, if valuable information is expected as a result. We recommend careful consideration and consultation of the local research ethics review board before undertaking such experiments.

In our experiment the deception was so benign that we considered it justified. Only calibration took place before we revealed the situation to the participants. All actual eye tracking took place after the participants knew what was going on.

Our hidden calibration routine utilizes the knowledge from previous research regarding temporal and spatial coupling of point of gaze (POG) and mouse movement in aiming tasks [Smith *et al.* 2000; Hornof and Halverson 2002]. The main finding was that in tasks that require mouse pointing at small targets, we can isolate the moment when the mouse cursor makes the final approach to the target as the time when the user is looking at the target with high probability.

Our calibration procedure was based on an on-screen questionnaire that has “radio button” widgets that the participant must click to record his or her responses. Such procedure is suitable in many experiments because from the participant's point of view it is plausible that the experimenters collect participant demographics with a computerized questionnaire. At the end, of course, the participant must be debriefed carefully explaining what data were recorded and what the researcher intends to do with them. Good practice also requires that the participant is given the opportunity to withdraw from the experiment.

5 Comparison of Calibration Methods

Apparatus and Participants

We used the Tobii T60 eye tracker. It was calibrated using our form based calibration routine and the default 2, 5, and 9 point calibration techniques in Tobii's SDK. The data quality was

measured using TraQuMe with four points (Figure 2). Each of the 4 fixation points was 20 percent of the screen dimension away from the corners. For quality measurement, gaze data was collected for 1.5 seconds per point.

We recruited 12 volunteer participants from the university community (2 females and 10 males) aged between 19 to 50 years. Seven participants had normal vision while the vision of the remaining five was corrected to normal.

Design

Our main interest was in comparing the accuracy and precision of conventional calibration techniques to the form based technique. The ideal result in this comparison would be that there is no difference in tracker data quality between the normal calibration procedures and the form-based “hidden” procedure. All calibration techniques were compared statistically to each other in pairwise randomization test. In these tests the null hypothesis is that the pairwise differences were just as likely to end up positive as negative. Repeated random assignment (n=10,000) of the sign of the differences gives us a sampling distribution of the mean difference. The observed mean difference is then compared to this distribution to see how likely it was to occur by chance. We could have done the same with pairwise t-tests, but in the absence of good assumptions about the nature of the distributions in TraQuMe data, we opted for the side of caution and used the non-parametric randomization tests.

Procedure

Before the start of the experiment the participants were seated in front of the tracker at a distance of 60-70 cm. In order to reduce the operator effect on the calibration, the experimenters had a script (on paper) that they followed when giving instructions. The participants were told that the study is related to gaze tracking and they will be briefed in detail after filling an electronic background questionnaire.

The background questionnaire layout used for the experiment is shown in Figure 3. Only one question was active at a time and the other questions were greyed out. Participants used a mouse

Figure 3. The questionnaire used for hidden calibration (fading effect removed for clarity).

to answer the questions. The number of samples collected for each calibration point was set to the default value in the Tobii analytics SDK. Normally, data were collected for roughly 0.5 seconds. To verify whether the participants knew what was going on, they were asked if they found anything peculiar about the questionnaire. After they answered, the calibration process was disclosed. Then, the gaze data quality for the calibration was measured using TraQuMe. After the first TraQuMe measurement, the participant completed Tobii’s built in 2 point, 5 point, and 9 point calibration processes each followed by a TraQuMe measurement. In the Tobii calibration set ups we used a white background for the screen to match the background color and screen illumination of the form based calibration routine. The order of the built-in Tobii calibrations was counter-balanced between participants.

6 Results

Data considerations

Early on in our measurements the form-based calibration failed for two participants due to a programming error that crashed the system under certain conditions. The software was fixed and new participants were recruited to replace these two. Later, due to an error in the Tobii 2 point calibration routine, which caused a failure in calibration when data for one eye was missing for one point, another participant had to be replaced (the nature of the failure was found later in debugging, not immediately).

Comparison of the calibration techniques

The binocular mean and maximum offset for the four points were used to compare the accuracy of the four calibration techniques. Figure 4 shows the mean offset for different calibration techniques.

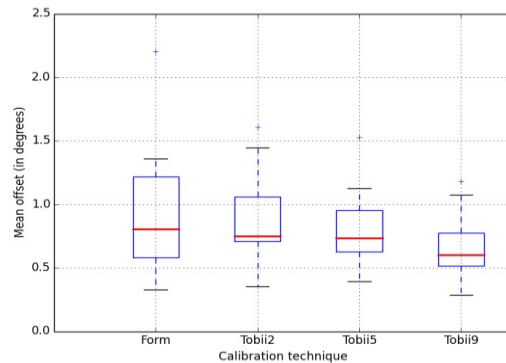


Figure 4. Mean offset for different calibration techniques.

However, randomization test revealed no statistical significance in these differences. The maximum offset across the four points followed the same pattern with Tobii’s 9 point method showing the smallest median value (Table 1). Again, the differences were not statistically significant.

Form	2 point	5 Point	9 Point
1.176	1.179	1.098	0.980

Table 1. Median value of maximum offset (in degrees of visual angle) for different calibration techniques

As expected, there were no noticeable (or statistically significant) differences in the binocular precision for the different calibration techniques (Table 2).

Form	2 point	5 Point	9 Point
0.246	0.271	0.283	0.286

Table 2. Median value of mean precision measurement (in degrees of visual angle) for different calibration techniques

Questions on the successfulness of the hidden calibration

When asked if the participants found anything peculiar about the questionnaire, 10 out of 12 felt that the questionnaire was “normal” to them while 2 felt that the layout of the questions in the form caught their attention. None of the participants had realized that the tracker was calibrated with the data collected during form filling.

7 Discussion

For repeatability and comparability of experimental work it is important to have detailed records for the quality of gaze data used in experiments. Furthermore, Holmqvist *et al.* report that poor gaze data quality may lead to incorrect findings in gaze research [2012]. Evaluating and reporting the gaze data quality leads to greater confidence in the findings.

In our experiment, we used TraQuMe to compare the four different calibration methods. The main finding was that the hidden calibration procedure worked very well. Some variability in calibration results is to be expected when participants are not explicitly cooperating in calibration. For example, they may be more likely to blink their eyes at a critical moment leading to a shortage of valid gaze points for the calibration. They may also utilize eye-hand coordination strategies that are not optimal for calibration. E.g. they may not focus their gaze exactly on the radio button when the mouse enters it. However, the calibration results with the hidden calibration were generally almost as good as they were when the participants were cooperating. When the participants’ answers are useful for the research, the form based calibration serves a dual purpose of providing the form data and the calibration data. This saves time and makes running experiments more efficient.

Our findings on TraQuMe itself were also encouraging. With TraQuMe we can report that the median offset with the hidden calibration routine was just below 0.81° (range 0.32° - 2.20°), with a median precision of 0.25° (range 0.11°-1.05°). Thanks to TraQuMe the measurements can be collected rapidly, compared and reported easily and fairly.

Our study has a few shortcomings. We only used 4 points to compare the different calibration techniques. A more comprehensive comparison with detailed data on all screen areas would require more measurement points spanning the entire screen. Another weakness in our measurement setup was that the form-based calibration needed to be completed first to make sure that the participants did not have any knowledge regarding the nature of the test while filling the questionnaire. The price we paid was that if fatigue or boredom played a role, it was not completely counterbalanced. It is possible that the participants were more alert in the beginning and thus performed the calibration quality measurement after the form-based calibration better than with

some of the subsequent calibration methods. Further studies will show whether our surprisingly good results can be replicated.

8 Conclusions

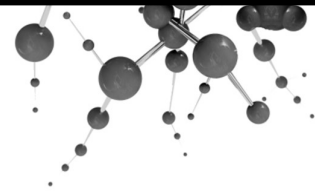
We have investigated two interlinked issues in doing research with eye trackers. First, we described a tool for verifying eye tracking data quality in about 30 seconds. Second, we described and evaluated a technique for calibrating the eye tracker in a laboratory setting without letting the participant know of the calibration. The data quality measurement tool and the calibration procedure were both found functional and we can warmly recommend using them in experimental work.

9 Acknowledges

We thank the members of TAUCHI unit who provided helpful comments on different versions of this paper. This work was funded by the Academy of Finland, project Haptic Gaze Interaction (decisions numbers 260026 and 260179).

10 References

- BATES, R. AND SPAKOV, O. (2006) D2.3 Implementation of COGAIN Gaze Tracking Standards. COGAIN, IST-2003-511598: Deliverable 2.3. Retrieved from <http://wiki.cogain.org/>
- BLIGNAUT, P, AND BEELDERS, T. "TrackStick: a data quality measuring tool for Tobii eye trackers." *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012.
- HOLMQVIST, K., NYSTRÖM, M., AND MULVEYZ, F. 2012. Eye tracker data quality: what it is and how to measure it. *ETRA'12*, 45-52.
- HORNOF, A.J., AND HALVERSON, T. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers* 34, no. 4 (2002): 592-604.
- MORGANTE, J,D., ZOLFAGHARI, R., AND JOHNSON, S.P. 2012. A Critical Test of Temporal and Spatial Accuracy of the Tobii T60XL Eye Tracker. *Infancy*, 17, 1, 9–32.
- NYSTRÖM, M., ANDERSSON, R., HOLMQVIST, K., AND VAN DE WEIJER, J. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45, 1, 272-288.
- SMI 2012, Sensometric instruments iView X™ SDK 3.0 manual. Retrieved from <http://www.smivision.com>.
- SMITH, B.A., HO, J., ARK, W., AND ZHAI, S.. "Hand eye coordination patterns in target selection." In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 117-122. ACM, 2000.
- TOBII TECHNOLOGY 2011. Accuracy and Precision Test Method for Remote Eyetrackers, rev. 2.1.1, 7 February 2011. Tobii Technology AB. Retrieved from <http://www.tobii.com>.



Paper 2

Deepak Akkil, Andrés Lucero, Jari Kangas, Tero Jokela, Marja Salmimaa, and Roope Raisamo. 2016. User Expectations of Everyday Gaze Interaction on Smartglasses. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. Article 24, 10 pages. DOI: [10.1145/2971485.2971496](https://doi.org/10.1145/2971485.2971496)

© ACM 2016, Reprinted with permission.

User Expectations of Everyday Gaze Interaction on Smartglasses

Deepak Akkil¹, Andrés Lucero², Jari Kangas¹, Tero Jokela³,
Marja Salmimaa³ and Roope Raisamo¹

¹ Tampere Unit for Computer-
Human Interaction
University of Tampere, Finland
firstname.lastname@uta.fi

² University of Southern
Denmark
Kolding, Denmark
lucero@acm.org

³ Nokia Technologies
Tampere, Finland
firstname.lastname@nokia.com

ABSTRACT

Gaze tracking technology is increasingly seen as a viable and practical input modality in a variety of everyday contexts, such as interacting with computers, mobile devices, public displays and wearables (e.g. smartglasses). We conducted an exploratory study consisting of six focus group sessions to understand people's expectations towards everyday gaze interaction on smartglasses. Our results provide novel insights into the role of use-context and social conventions regarding gaze behavior in acceptance of gaze interaction, various social and personal issues that need to be considered while designing gaze-based applications and user preferences of various gaze-based interaction techniques. Our results have many practical design implications and serve towards human-centric design and development of everyday gaze interaction technologies.

Author Keywords

Everyday gaze interaction; gaze tracking; head-mounted displays; interactive eyewear;

ACM Classification Keywords

I.3.6 Methodology and Techniques: Interaction techniques.

INTRODUCTION

Gaze-based human-computer interaction has been available for decades. However, until recently its use has been limited to a desktop-based assistive technology catering for motor-disabled user groups. Recent advancements in both software and hardware technology have made gaze-tracking cheaper, more accurate and ergonomic to use. The technology is increasingly seen as a viable and practical input modality for able-bodied users in a variety of everyday contexts such as

interacting with distant displays [30,33], mobile phones [16] and wearables such as smartwatches [2] and smartglasses [18].

Previous studies on gaze interaction targeting able-bodied users have mainly focused on the development of enabling technologies (e.g. developing gaze tracking sensors and algorithms to be used in various devices) [14,15] and experimental evaluations of specific interaction techniques and applications [9,16,30,33]. E.g., Vidal et al. [33] studied spontaneous smooth-pursuit gaze interaction on public displays and report the usability of the technique based on success of the interaction and other time-based measures. Similarly, Stellmach and Dachselt [30] studied the combination of gaze and touch to interact with computers and report both qualitative and quantitative findings. One should note that, all these insights are specific to the interaction technique in question and the context in which the study was conducted.

While very important for technology and research development, such studies provide limited insights into people's holistic perceptions and expectation of the future technology [25]. They do not answer questions like "What are the users' impressions about an environment where gaze interaction is ubiquitous?", "In what contexts would users prefer to use gaze interaction if the technology was perfect?", "In what contexts would such a technology not be acceptable?" and "What are the social and personal implications of everyday use of this technology?". The ideal research method to answer these questions would be to conduct observational studies of how people use gaze tracking technologies in everyday scenarios. However, such studies are difficult to conduct now because gaze tracking technology still requires further research and development to work seamlessly in all the contexts and environments [5].

Another promising approach to get insights regarding a future technology, is to enquire about user's expectations of using the technology [25]. Olsson [25] notes that knowing people's technology expectations helps us to understand how a technology should function in varying contexts, providing both general and specific insights to channel its design and development. In this paper, we present a study that aims to

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.
NordiCHI '16, October 23-27, 2016, Gothenburg, Sweden
© 2016 ACM. ISBN 978-1-4503-4763-1/16/10...\$15.00
DOI: <http://dx.doi.org/10.1145/2971485.2971496>

understand the expectations, needs and concerns of future users of gaze-tracking technology.

While there are many potential form-factors that a future gaze-tracking capable device could take (e.g. displays with gaze-tracking sensors, smart contact lenses, smartglasses), we chose smartglasses as the platform for investigation. Within the scope of this study, we define smartglasses as eyewear computers with gaze-tracking capability and a binocular see-through display that enables augmenting virtual content on the real-world. Smartglasses are gaining popularity with the advent of commercial devices like Google Glass and Microsoft HoloLens. Gaze tracking is an input technology with large potential in such devices [5]. Unlike other form factors, smartglasses enable a use case in which gaze is tracked continuously and used in varied contexts, where people use gaze to interact with different objects in the environment, instead of confining the interaction to a display. Selecting smartglasses as the platform in our study allowed us to focus on a single form-factor, while broadening the investigation to a variety of use-context, providing richer understanding about suitability and acceptability of gaze interaction.

We conducted six focus group sessions with heterogeneous participant groups, using scenarios of gaze-tracking smartglasses as probing materials to enquire users' expectations. Our focus was to understand if the context of use (individual/social, public/private, indoor/outdoor) has an influence on the acceptability of the technology and to elicit specific needs and concerns of the users regarding the use of gaze interaction on smartglasses.

The rest of this paper is structured as follows. We begin by reviewing relevant related work. Then, we describe our study and the five scenarios for gaze interaction on smartglasses used as the introductory material in the focus group. Next, we report the results of our focus group study followed by discussion and conclusions.

RELATED WORK

Gaze-based Interaction techniques

There are multiple ways of using gaze in human-computer interaction. Gaze can be used as implicit input, where the system identifies user's interests based on the gaze pattern and modifies the system behavior accordingly. Alternately, gaze can also be used to provide explicit commands. There are three common ways of explicitly using gaze: dwell-time based interaction, gaze gestures and smooth-pursuit based interactions. Dwell-time based interaction requires the user to stare at items on a screen or in the real-world for a pre-defined time to select them. Gaze gestures are predefined eye movements that map to some specific user command [8]. Smooth pursuit-based interaction relies on correlation between trajectory of eye movement and on-screen object [33]. Gaze gestures and smooth-pursuit based interactions are known to be less sensitive to tracking inaccuracies and suitable for mobile gaze interaction.

Gaze Interaction on Smartglasses

Lee et al. [18] developed an augmented reality annotation system, by integrating an optical see-through head-mounted display device with a gaze tracker. The user could receive augmented information of real-world objects on their display, by selecting the object using gaze. They used a two-stage selection process using dwell and half-blink to avoid accidental invocation of actions. Baldauf et al. [3] studied the use of gaze-input and audio output for retrieving annotated digital information from the surroundings. In our study, we use smartglasses as the platform to further investigate users' expectation towards gaze interaction.

Challenges to Gaze Interaction in the Wild

The Midas-Touch problem (distinguishing eye movement for interaction from normal eye movement) and reduced gaze data quality are two of the classic problems in gaze-based interaction [21]. Bulling and Gellersen [5] note that for wearable trackers, the tracking accuracy is further reduced due to calibration drift during operation induced by mobility. Many different approaches are proposed to improve tracking quality using re-calibration procedures hidden from the user based on task characteristics [1] or visual saliency [31]. Another challenge in mobile video-based gaze tracking is the battery consumption. Most wearable trackers only work for a limited duration of 2-4 hours [5]. This has led research in the direction of light-weight eye movement measurement techniques based on electrooculography (EOG).

Many technical and interaction-level challenges still exist in the vision of ubiquitous gaze-based interaction. Our study complements the previous work in this area and aims to look at everyday gaze interaction, not from a technological perspective, but by enquiring the expectations and needs of potential users of this promising technology.

User Expectation and User Experience

Hassenzahl and Tractinsky [10] define user experience as *"consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system and the context within which the interaction occurs."* This definition emphasizes the role of temporality and context on experience. Michalco et al. [23] notes that people form expectations of an interactive product even before using it and these expectations influence their attitude towards the product. McCarthy and Wright [22] note that only when experience meets or exceeds the expectation, users identify positively with the experience. Expectation disconfirmation is a strong factor in the user's experience with the product.

There is wealth of literature that confirms the role of user expectation in shaping user experience. Gaze interaction is a promising future technology for the consumer market. In our study, we aim to understand and reflect the expectations of the potential users of this technology to further channel the research, design and development. In the following section, we explain the focus group study we conducted.

FOCUS GROUP STUDY

We conducted six exploratory focus group sessions with heterogeneous groups of participants. Focus groups were selected as the data collection method because it is suitable for early exploratory studies providing concentrated amounts of data on the specific topic of interest efficiently. Focus group sessions followed a scenario-driven approach. We created five scenarios presenting an “ideal-world” narration of a future with gaze-tracking smartglasses, which was used as probing material in the focus groups. The scenarios provided the participants a common ground to reflect upon their needs, preferences and expectations, without giving too much detail about the technology or the interactions. Each focus group session had 3-4 participants and lasted approximately 2 hours.

Five Scenarios

There were many potential ways of designing the scenarios, e.g. deriving it from mobile phone usage trends or surveying studies on applications of smartglasses. Our scenarios were mostly inspired from previous work on mobile gaze-based interaction, covered a variety of contexts of use and were all potential smartglasses applications. The scenarios were developed with the following considerations:

- Mix of indoor/outdoor, individual/social, private/public contexts.
- Mix of different gaze interaction techniques implicit/explicit, gaze gestures/dwell-time based.
- Plausible future real-world use case based on current trends and research.
- Each scenario highlighted a specific advantage of using gaze.

Handsfree interaction

It is the month of December and it has been a harsh winter so far. James is walking to the University of Tampere to attend the morning lecture. He is wearing his smartglasses with gaze-tracking capability. While on his way, James realizes that he had agreed to call Susan. Without taking his hands out of his pockets, James makes a ‘Z’ gesture with his eyes to launch the contact list. He uses his eyes to browse through the contacts one by one on his glasses and proceeds to call Susan. They decide to meet in the evening for coffee.

This scenario focuses on outdoor usage of the device in an individual context. The scenario further introduces the concept of using gaze gestures for mobile interaction [7]. The scenario was inspired by previous work by Kangas et al. [16].

Private interaction

Laura has decided to go watch the local ice hockey game with her friends. They gather at the city center and wait for others to join them. Laura suddenly notices a notification on her glass display. She quickly looks at the notification to open the message. It is Laura’s boyfriend from Germany. The message says: ‘It’s a beautiful evening, wish you were

here with me’. Her face glows and she cannot help but smile. She gazes at the ‘Reply’ option for a short while and selects a ‘Kiss’ symbol. She responds to the message with her eyes and then joins her friends in the conversation.

This scenario focuses on outdoor usage of the device in a social context. The scenario was inspired by earlier work on the use of smartglasses to receive and read mobile notifications [19,20] and using gaze to interact with notifications on smartwatches [2].

Implicit interaction

Martin loves to travel and has just arrived in Helsinki. The weather is nice, and the place is full of tourists. Martin likes to explore a new place on his own and decides to take a walking tour of the city. Wearing his smartglasses, Martin walks down the street along the park and sees a beautiful and royal-looking building to his right. Intrigued by the architecture, Martin starts looking at it more carefully. He wishes he knew more about the building. As if they could read his mind, the smartglasses recognize Martin’s interest based on the long staring. They then display that the building is the Royal Museum built in 1887. When Martin finishes reading the information, it shows more information and a brief history of the building.

This scenario focuses on outdoor usage of gaze-tracking capable smartglasses in an individual context. This scenario was motivated by two previous studies. First, the work of Qvarfordt et al. [28] on the use of eye gaze to detect user interest and proactively adapt output information in a desktop-based tourist information system. Second, the work by Baldauf et al. [3] on the use of mobile gaze trackers to retrieve georeferenced information for urban exploration.

Unobtrusive interaction

Mark is a student at the University of Tampere. He is a fun-loving person and loves to keep himself engaged. Mark wants to travel Helsinki to meet a friend. He boards a bus and sits next to an elderly person who is sleeping. While looking around, Mark finds out that the bus offers onboard entertainment similar to that in airplanes. It includes entertainment eye glasses with gaze-tracking capability and a display on the glasses. Mark switches on the glass and wears it. Mark can see a menu with options like ‘News’, ‘Music’, ‘Games’ and ‘Movies’. Mark realizes that the glass is responding to what he looks at. He swiftly scrolls to the ‘Movies’ section and selects one of the latest movies from the list with his eyes.

This scenario focuses on indoor usage in a (semi) public social context. The scenario is inspired by the previous work on gaze as attentive interfaces [4] and use of smartglasses for entertainment applications [26]. Unlike the other scenarios, the smartglasses are not a personal device but part of the bus’s onboard entertainment system.

Social interaction

Anne is at a business conference. She knows a few of the other participants but not all. She realizes that it's a great networking opportunity. Anne looks at different people around her one by one. Her glass identifies them and displays their name and interests on the display. She slowly changes her gaze from one person to another and soon finds someone with similar business interests. She decides to go say hi and to discuss some ideas. Anne is ecstatic about making the most out of this networking opportunity.

This scenario focuses on the use of the device in an indoor, social context. The scenario is motivated by previous work on using gaze input on smartglasses for networking [29] and using smartglasses as a name-tag application by facially recognizing collocated individuals [32].

Technology Demonstration

We felt it was critical to give participants concrete examples of the potential of the technology before the start of the discussions. We prepared four demonstrations to convey the capabilities of smartglasses with binocular see-through display and gaze-based interaction.

Remote Gaze-Tracking

We used an EyeTribe gaze tracker connected to a Windows 7 tablet for the gaze interaction demonstrations. We developed a messaging application (see Figure 1a), which could be navigated horizontally or vertically by either dwelling at the corresponding red arrows for 750ms, or by using simple two-stroke gaze gestures. The first stroke of the

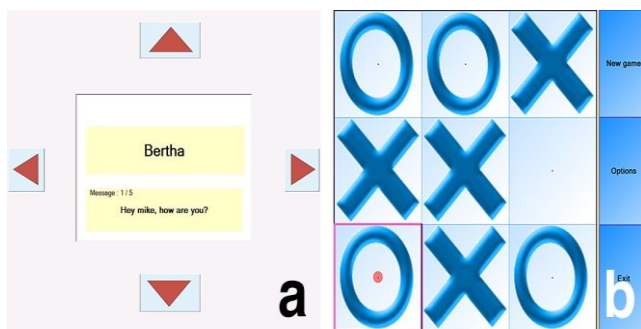


Figure 1. Technology demonstrations: a) messaging application that uses dwell and simple two stroke gaze gestures, and b) dwell-time based TicTacToe game.

gaze gesture started from the center of the box towards any of the four cardinal directions and the second stroke returned the gaze back to the box. The gaze gestures were the same as used by Kangas et al. [16]. We used a time-out of one second between strokes to differentiate between normal eye movements and an intentional gaze gesture. Secondly, we used a gaze controllable version of the TicTacToe board

game. In the game, each cell could be selected by dwelling at it for 750ms (see Figure 1b).

Smartglasses Demonstration

We used Epson Moverio BT-100 binocular see-through smartglasses for demonstration. The built-in gallery application showed various 2-D and 3-D images, which could be browsed using the handheld touchpad.

Mobile Gaze Interaction

Further, we developed an application using the Ergoneer Dikablis head-worn monocular gaze tracker. Several visual markers were placed in different parts of the room and the application could recognize when the person was looking at the visual markers and gave auditory feedback (i.e. a short beep) and visual feedback (i.e. color of a corresponding GUI object turned blue) when the user fixated upon the markers for longer than 300ms.

Video Demonstration

We selected a video developed by Nokia Research Center¹, depicting a concept of gaze-based interaction on smartglasses along with other smart technologies. The video was freely available on the internet.

Participants

A total of 23 participants from the local university were recruited using noticeboard advertisements and mailing lists. Participants varied in age (19-52 years, median 24), gender (10 male and 13 female) and study background (e.g. computer science, business, health-science, literature and education). Eight participants had prior experience in gaze interaction as part of previous experiments and two participants had earlier used head-mounted display devices. In the background questionnaire, on a scale of 1 to 7 (where 1 is strongly disagree and 7 is strongly agree), participants stated that personal devices were an important part of their lives (Mean=5.9, StDev=0.92) and that they are interested in trying new technological devices (Mean=5.4, StDev=1.07).



Figure 2. Seating arrangement of participants and moderator (rightmost) during the focus group session

¹ <https://www.youtube.com/watch?v=A4pDf7m2UPE>

Procedure

The study consisted of four main parts: introduction, technology demonstration, scenario discussion and debriefing.

Introduction

The moderator welcomed all the participants to the focus group discussion. The participants and the moderator were then seated on a couch in a semi-circle around a coffee table (see Figure 2) and then they were asked to introduce themselves. The moderator described the purpose of the study, and then participants signed an informed consent form and completed a short background questionnaire.

Technology Demonstration

Participants took turns trying the remote gaze-tracking demonstration, while the rest watched. Participants sat comfortably on a chair in front of the tablet connected to the EyeTribe gaze tracker that was set up on a table. After a brief 9-point calibration procedure, participants first played 3-5 rounds of the dwell-time based TicTacToe game, followed by the messaging application. The participants used the messaging application using both gaze gestures and dwell-time based input. Next, all participants tried the smartglasses demonstration. The participants were instructed to walk around the room wearing the glasses and asked to imagine wearing such a device while walking in an outdoor environment. This was required to give the participants perception of a real-world mobile scenario. Further, one participant per focus group session demonstrated the mobile gaze interaction system. Again, following a 4-point calibration routine, they were asked to gaze at the different visual markers placed nearby. The other focus group participants watched the demonstration. Finally, the participants viewed the video of gaze-based interaction on smartglasses. This part lasted for approximately 25 minutes.

Scenario Discussion

After a brief general discussion on the demonstrations and the technologies, the five scenario descriptions were handed out to the participants on paper. The moderator then instructed the participants to read a specific scenario. For each scenario, the participants were encouraged to imagine an idealistic world where the different technologies would work seamlessly. The participants discussed their general impression of using gaze in the specific context. This was followed by several open-ended questions relating to the use of gaze interaction on smartglasses. The scenarios were presented to all the focus groups in the same order. After approximately 1 hour, there was a 10-minute coffee break. The discussion for each scenario lasted approximately 15 minutes, for a total of 75 minutes.

Debriefing

Following the scenario discussion, the moderator asked a few closing questions, to elicit any concluding remarks. The moderator then thanked the participants for their participation. Participants were compensated with a movie ticket for their time. The focus-group sessions were video recorded for later analysis.

Analysis

The focus group sessions were first transcribed and later analyzed using affinity diagramming [11]. Four researchers involved in the study individually analyzed the transcripts of three different sessions each, creating 40-50 affinity notes per session. The affinity notes were then hierarchically organized and grouped into common themes, while relevant user quotes were preserved.

RESULTS

In the following sections, we describe our main results. Figure 3 gives an overview of the thematic structure of the focus group data.

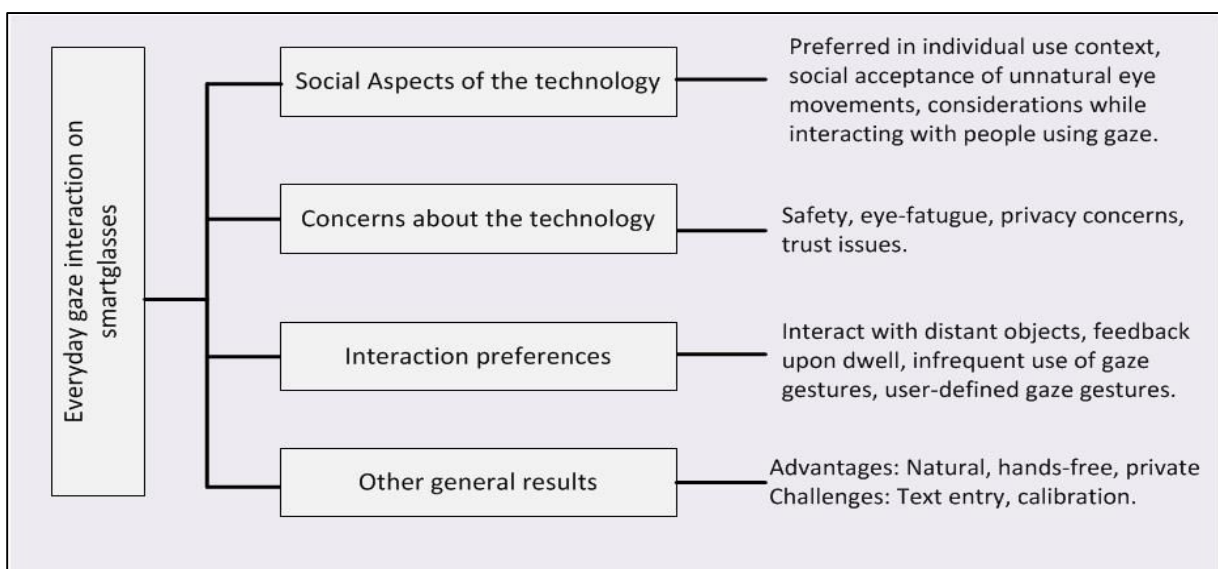


Figure 3. Thematic structure of focus group data

Social Aspects

Context of use had a strong influence on how participants perceived the technology. Participants generally felt positive about the use of gaze interaction on smartglasses in an individual context in both private and public environments, but not in social scenarios. *“I think that this technology is better used when you are alone, not when you are with other people”* (P3). Participants had three distinct concerns about use of such technology in social situations.

Gaze Interaction in the Presence of Onlookers

Participant expressed that watching a collocated person performing unnatural eye movements like gaze gestures in a public environment will be *“noticeable”*, *“little weird”* and *“take some getting used to”*. Many participants compared it to the *“talking to yourself”* feeling when Bluetooth headsets were launched. *“You might think they are looking at you or making some gestures to you. It is the same, sometimes I think someone is talking to me when they are talking to their headsets.”* (P12). Interacting with the device may give the impression that the person is performing the eye movements looking at another person. For the same reason, few participants felt that it would be more comfortable for the user if the glasses are tinted, so that onlookers cannot see the users' eye movements. *“I would use it, if there is some shades or something. So that it is not clear glass”* (P21). *“It (tinting) could help so that people cannot see that you are [makes sequence of eye movements]. You are going to be comfortable doing that on the streets (P15)”*.

Gaze Interaction on People

Participants also felt strongly about using gaze to interact with collocated people, i.e. dwelling at people to get more information about them (as in the *conference* scenario) or something worn by them (e.g. dwelling at the shirt or shoe to know its brand). Participants felt that even though it is natural to glance at people in an environment, it is disturbing to look at people for a longer duration. *“It is quite disturbing to stare at some people, especially strangers. I think it is invasive in general.”* (P20). *“People are not products. I am not interested in using it on people”* (P3). Though some participants felt such interactions may be acceptable in a controlled environment, where the user already knows about the purpose of the technology and knows what the *“staring”* means.

Another interesting difference emerged about the visualization of information when they were related to a person and an object or product. In case of interacting with an object using gaze, the participants preferred the extra information be shown on display and visually linked to the object (e.g. by placing the information above the object). However, while interacting with people (or clothes and accessories worn by them), participants suggested that the user could glance at the person or the object worn and read more about them on glasses later without requiring to dwell

at the person for long or appear to be staring in the person's vicinity while reading the information on the display.

Gaze Interaction in Social Situations

Participants recognized that eyes, and especially eye contact, are important elements in everyday social interactions and hence our participants felt such technology may be disruptive, distracting and not socially engaging. *“I would not like to use this in a social environment, because the way you initiate social contact is through eye contact. If you are interacting with something using the eyes, you may miss the other person's eye contact. It is not conducive to sociability in my opinion.”* (P10)

A majority of the participants also felt that unlike using other modalities like touching the device or using voice commands to interact with smartglasses, gaze makes it easy to covertly interact with the device, or pretend to attend to a situation while acting on the glasses. Few participants felt strongly about wearing such gaze tracking capable smartglasses in social scenarios.

[P6] *I personally hate it when I communicate with somebody and he uses mobile phone or is thinking something else. That is why I would not use it in social situations.*

[P8] *Maybe in black sunglasses. Then other person would not see your eyes.*

[P6] *It is the same. I will just feel that I am talking to a wall.*

Some participants were of the opinion that when gaze tracking becomes common in smartglasses, wearing a smartglasses in conversations could be perceived negatively. *“People usually appreciate if others listen to them. When you have the glasses on, and everybody also knows that you can be doing stuff there with your eyes, it can be unnerving”* (P16). While few others thought that people may get used to others wearing such glasses while in a conversation. If the glasses are tinted, they proposed that there could be some visual indicator of the activity, so that the conversation partner can know if the person is interacting with the glasses or listening to the conversation. *“If someone is talking to you, it might be a good thing that they know you are doing something on your smartglasses. It might be a good idea to have some light showing that (P13)”*.

Safety, Health and Privacy Concerns

Personal Safety and Health

Many of the participants also raised personal safety and health-related concerns. Participants raised concerns about the safety aspect of long-term use of gaze-tracking technology. *“Is it (gaze tracking) safe to use for long durations?”* (P6). Earlier work has investigated health issues with desktop-based eye gaze interaction for disabled user groups [6,24]. Most current day commercial wearable gaze trackers use artificial infrared lighting close to the eyes for tracking the pupil. Long-term exposure of the eye to strong infrared (IR) lighting may have health implications [24]. Considering that people could wear smartglasses for long

durations every day, and that the infra-red source is closer to the eyes than remote trackers, extensive research should go into the safety aspect of the system.

Participants felt that using eyes to control such glasses, especially using frequent gaze gestures, may be unhealthy or lead to eye fatigue. *“I can see eye strain happening really easily, trying to move your eyes that much.”* (P18). Chitty [6] investigated eye fatigue using assistive eye gaze interaction on desktop computers. Novice users may feel eye fatigue due to unnatural eye movement. However, most experienced users do not normally report any fatigue in use of gaze interaction in desktop computers.

Privacy

Participants also raised privacy concerns of using gaze-tracking smartglasses in everyday life. The privacy issues associated with the video capability of such devices and its covert use in public places was discussed. However, another important concern raised was about the ease of collecting personal gaze data and the potential misuse of it. Information about what a person is looking at and for how long, or how carefully, can provide a wealth of sensitive information about the person’s interests and preferences *“Somebody is probably going to collect that data of what you are looking at and start recognizing certain patterns. It is like a very effective data collection tool.”* (P9).

Trust

Participants in general did not feel gaze tracking smartglasses, can be trusted to replace more mature technologies like mobile phones. *“I still do not think I can trust such a device (P6)”*. *“I would probably lose my nerves if the glasses did not obey me automatically. I look there and nothing happens! Then, I am not going to use this ever again (P12)”*. Unlike familiar devices like mobile phones, users expressed concern about potential ease of identifying when the device is not working properly, troubleshooting issues and recovering from errors. *“It would be very frustrating if it did not work. I will not know if it is my mistake or the system’s mistake.”* (P12).

Interaction Preferences

Most of the participants felt that interacting with distant objects or retrieving information about objects in the environment as a key application for gaze tracking smartglasses. *“This is one application the glasses would be really good for. If glasses are on your eyes and (its display) overlaid on your vision and then you could see that there is a tag to a hotel, there is a tag to a museum and there is a tag to a subway station, you could then look at the tags and get more information.”* (P16).

Dwell-time based Interaction

Dwell was considered the most natural method for selecting an item, using gaze on smartglasses. Participants felt that in scenarios of dwelling at a real-world object or glasses implicitly identifying user interest (as in the *tourist* scenario),

the glasses should provide some gentle feedback when there is more information available about the real-world object that is glanced at and it should be under the user’s request that more information be displayed. *“Glasses should be polite, it should ask if the user wants to know more information about the item.”* (P18).

Gaze Gestures

Participants preferred dwell-time based interaction over gaze gestures for frequent interactions, as gestures require unnatural eye movements. Many of the participants felt that gaze gestures are better suited to short and infrequent interactions as they were clear and less likely to be misinterpreted by the system (e.g. simple distinct commands like ‘Yes’, ‘No’, ‘Ok’, shortcuts to different applications, unlock the device). Most earlier works on gaze gesture in desktop computing scenarios use the technique for frequent interactions, like scrolling text entry [12], or as discrete input in games [13]. Our results suggest that gaze gestures may be more suited for clear but infrequent interactions.

Participants thought that it is important to let users define the gestures that they find comfortable. *“If the user has the ability to custom define the gesture. A ‘Z’ gesture might not be easy for me but, might be easy for someone else. If I can make my gesture that will make it easier.”* (P10). *“I might prefer an ‘N’ gesture (P21)”*. While earlier work has investigated the usefulness of user-defined hand gestures for smartglasses [27], most work on gaze gestures has used predefined gestures for interaction. Our results suggest that allowing users to customize the gaze gestures to suit their preferences may be advantageous.

Participants felt that another drawback of gaze gestures is that the user may forget the gesture or may not be aware of it during first time use. It could hence be beneficial if the glasses reminded the users of some of the possible gestures. *“If I do not remember all the gestures, it could remind me some of the gestures”* (P1). Participants also felt that the system should provide adequate feedback to aid performing the gestures, this is in-line with work by Kangas et al [16].

Other General Results

Our participants also highlighted many positive aspects and challenges of using gaze interaction on smartglasses. Unlike in handheld devices that can be easily touched to interact, gaze was considered to be a natural method for interaction in smartglasses. Our participants felt the main advantage of gaze is that it is hands-free and the interactions are more private and unobtrusive. *“The most important thing is to free the hands. If we use other methods to interact, it defeats the purpose.”* (P22).

Participants also identified few interaction challenges. Most participants considered entering text (e.g. to respond to a message or search for music) by eyes to be complex, strenuous and slow. *“Entering text using eyes will be very difficult and unnecessarily time-consuming, I would not want to use it”* (P9). This is in-line with previous work on dwell-

time based text entry by Majaranta et al. [21]. Another challenge recognized by the participants was the need for calibration. Our participants had only knowledge of the conventional methods for calibrating the trackers using multiple on-screen or real-world fixation points from the technology demonstrations. They considered this technique not suitable for smartglasses as it is slow and expected more flexible calibration procedures, in-line with previous work on automatic recalibration of tracker by Sugano et al. [31].

In general, participants felt that combining smartglasses with mobile phones could be desirable. The glasses were not considered a device that the user would wear at all times. Also, mobile phones were considered to complement smartglasses in functionalities in which glasses are lacking (e.g. text entry). Participants also observed the need for different output modalities to support the interaction effectively. While mobile, voice was the preferred output modality over visually presenting information, in-line with previous work by Baldauf et al. [3] that combined gaze events with audio output in mobile scenarios.

DISCUSSION

Enquiring user expectation towards everyday gaze interaction on smartglasses is important, considering that gaze tracking is soon expected to be a mainstream technology and also the social acceptability issues that are known to be associated with smart glasses (e.g. Google Glass). Our study was designed to be exploratory in nature and provides practical user-expectation insights and design guidelines that could serve as the basis for designing future gaze interaction applications. In the following section, we discuss the design implications of our results.

Design Implications

Our results suggest that context of use has a strong influence on how people feel about gaze technologies. Wearers of gaze tracking glasses may not be always comfortable performing unnatural eye movements in public scenarios and such gestures may also have an influence on the onlookers. Designers and application developers should consider the usage context of the system and attention should be given to social norms concerning eye-contact and unnatural eye movement. Eye contact is critical in face-to-face communication. Applications for smartglasses to be used in social environments, or to facilitate collaboration between collocated users, should hence consider approaches to minimize the use of eyes for interaction and free them for their face-to-face conversational functions.

Human eyes naturally support visual exploration of an environment and participants felt that eyes are a powerful modality to find and interact with objects in the environment. However, designers should be careful while developing applications where eyes are used as a medium to “select or point at” other collocated individuals. Careful design should be employed to use natural glancing as the interaction mechanics and reduce staring at the individual or their

vicinity while pointing at them or reading information about them on the display of the glasses.

Special attention should be taken while using gaze gestures for interaction on smartglasses. Gaze gestures have the advantage that they are clear and not invoked by accident. However, our results suggest that gestures are more suited for short and infrequent interactions. While using gestures, the system should support options to remind the users of the possible gestures and also allow users to define their own gestures for flexibility and comfort of use.

Participants raised concerns regarding eye-fatigue while using gaze interaction. Designers of everyday gaze interaction applications should strive to reduce the unnatural eye movements or design to provide adequate rest for people’s eyes. These approaches are especially important for early stage users, as experienced users do not report eye fatigue [6]. Ensuring a positive user experience for novice users is critical for technology adoption. Gaze interaction application could keep track of the experience of the user and employ interactions that require complex unnatural eye movements only for more experienced users.

Further, technology manufacturers and designers should consider the perceived safety and privacy concerns of potential users of the technology. These concerns could also be dealt with at a design level. Considerations like relying on visible spectrum gaze-tracking when possible and automatically turning off the IR light source when no eye movement is detected, may greatly reduce the adverse effects of long-term use of gaze-tracking technology and the perceived safety issues with the device. Such approaches will also help reduce the power consumption, which is a major problem in such wearable devices.

Participants voiced privacy concerns regarding storing and sharing gaze data. The device should support options to disable gaze tracking in specific environments. Providing other flexible input methods like combining the smartglasses with mobile devices or voice-based input would mean that users can continue to use the device, even in scenarios where gaze tracking is disabled. Designers should also employ a transparent privacy policy. Allowing the users to control the data recorded and transmitted online will be critical to reduce the privacy concerns of the potential users.

Our participants felt that gaze-tracking technologies cannot be “trusted” to replace other established devices. Participants also raised the need for ways to easily identify and troubleshoot problems with the device. In order for everyday gaze interaction technologies to be widely adopted by consumers, it is important that the technology instills a feeling of reliability and confidence in the minds of the users. Some desktop-based gaze-tracking systems (e.g. Tobii EyeX) provide users a continuous indication of visibility of the eye and tracking robustness. This continuous feedback allows users to ascertain when the device may not function (e.g. because eyes are not visible) and take corrective

measures. For wearable systems, dynamic situations like lighting, vibrations in the environment and movement of the device may affect robustness and accuracy of tracking. One should note that the accuracy required depends on the task (e.g. accurate tracking is required to precisely point with gaze a distant landmark from a high rise building but not necessarily to point at a large object near the user). Feedback options should also be employed in wearable gaze tracking systems, allowing users to easily ascertain the robustness of tracking and to assess if the device can be efficiently used in the specific context for the task at hand. There should be hence ways of not just automatically (re)-calibrating the tracker (e.g. [31]), but also keeping the users continuously aware of the tracking status and enabling them to take flexible and intuitive corrective measures when tracking quality is not enough for the current task.

Our results suggest that participants may not want to use gaze interaction in all use contexts. It would hence be important to support complimentary input modalities (e.g. mobile device, voice input *etc.*). Different output modalities should also be provided to enable flexible use cases (e.g. by allowing users to disable the display and use the device with voice output while outdoors, supporting haptics to convey subtle information without distracting the user *etc.*).

Limitations and Future Work

Our study has a few limitations. First, our participants were educated and technically-oriented. While we tried to have a heterogeneous mix of participants in terms of gender and study background, it should be noted that our participants were predominantly from Europe. It is likely that culture has an effect on people's attitudes and preference towards technology. Culture is also known to have an effect on the social gaze behavior [17]. Further research is required to understand the effect of participant selection on our results.

Second, our participants were unfamiliar with gaze tracking technology and smart glasses. The technology demonstration before the start of the discussion helped them get a fair understanding of the technology. However, it may have also influenced the participants' perception and opinion about the technology.

Third, we had to focus on one specific form factor for the smartglasses, i.e. smartglasses with binocular see-through displays, to reduce the scope of the study and not confuse the participants with different options. We think, however, that many of the results could also be extended to other everyday gaze interaction technologies (e.g. on a mobile phone). Future work could investigate if that is really the case.

Fourth, our study focused on understanding user expectation of gaze-based interaction on smartglasses. One could imagine that a combination of modalities (gaze, touch, voice, body gestures *etc.*) could be beneficial in many scenarios to interact with smartglasses. The focus of the work was not to compare the user preferences of using gaze interaction with other plausible combinations. Future work should look into

how users would prefer to combine these modalities to interact with smartglasses. Also, while we tried to cover a wide range of gaze interaction techniques, our study did not focus on smooth-pursuit based interaction, a calibration-free gaze interaction technique that has been gaining popularity recently. Future work should investigate user expectations and preferences of using smooth pursuits for everyday interactions.

Inquiring about needs and expectations of users of a future technology is challenging, especially without tangible prototypes to test the interactions. The intention of this study was to inform the design of future gaze-based technologies and increase awareness of some of the social and personal issues that needs to be taken into account while designing such systems. The goal of this study is not to replace an actual field observation of people using gaze-tracking capable smartglasses, when ubiquitous gaze interaction becomes technically feasible. Rather, this research contributes as a significant step towards gaining understanding of users' expectations towards everyday gaze interaction.

CONCLUSION

Our study was designed to be broad and exploratory in nature. It presents many new insights regarding expectation of potential users (e.g. social aspects of gaze interaction, need for flexible and complementary supporting modalities, concerns of the potential user group, and expectations regarding gaze gestures). In future, we plan to continue this line of research and develop applications for gaze-tracking capable smartglasses using other user-centric methods, focusing on the various social and personal issues that was revealed in this study.

ACKNOWLEDGEMENTS

We thank Nokia Technologies for their support during this work. The work was partly funded by Academy of Finland, projects Haptic Gaze Interaction (decisions 260026 and 260179).

REFERENCES

1. Akkil, D., Isokoski, P., Kangas, J., Rantala, J., and Raisamo, R. TraQuMe: a tool for measuring the gaze tracking quality. In *Proc. ETRA'14*, ACM Press (2014), 327-330.
2. Akkil, D., Kangas, J., Rantala, J., Isokoski, P., Spakov, O., and Raisamo, R. Glance Awareness and Gaze Interaction in Smartwatches. In *Proc. CHI EA'15*, ACM Press (2014), 1271-1276.
3. Baldauf, M., Fröhlich, P. and Hutter, S. KIBITZER: a wearable system for eye-gaze-based mobile urban exploration. In *Proc. of AH'10*, ACM Press (2010), p.9.
4. Biedert, R., Buscher, G., Schwarz, S., Hees, J. and Dengel, A. Text 2.0. In *Proc. of CHI EA'10*, ACM Press (2010), 4003-4008.

5. Bulling, A. and Gellersen, H. Toward mobile eye-based human-computer interaction. *Pervasive Computing*, IEEE, 9(4), 8-12.
6. Chitty, N. User Fatigue and Eye Controlled Technology. OCAD University (2013).
7. Drewes, H., De Luca, A. and Schmidt, A. Eye-gaze interaction for mobile phones. In *Proc. Mobility'07*, ACM Press (2007), 64-371.
8. Dybdal, M.L., Agustin, J.S. and Hansen, J.P. Gaze input for mobile devices by dwell and gestures. In *Proc. of ETRA'12*, ACM Press (2012), 225-228.
9. Esteves, A., Velloso, E., Bulling, A. and Gellersen, H. Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements. In *Proc. of UIST'15*, ACM Press (2015), 457-466.
10. Hassenzahl, M. and Tractinsky, N. User experience-a research agenda. *Behaviour & information technology* (2006), 25(2), 91-97.
11. Holtzblatt, K., Wendell, J.B. and Wood, S. Rapid contextual design: a how-to guide to key techniques for user-centered design. *Ubiquity* (2015), 3-3.
12. Isokoski, P. Text input methods for eye trackers using off-screen targets. In *Proc. of ETRA'00*, ACM Press (2000), 15-21.
13. Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S. and Vickers, S. Designing gaze gestures for gaming: an investigation of performance. In *Proc. ETRA'10*, ACM Press (2010), 323-330.
14. Järvenpää, T. and Aaltonen, V. Compact near-to-eye display with integrated gaze tracker. In *Proc. SPIE Photonics Europe* (2008), 700106.
15. Järvenpää, T. and Äyräs, P. Highly integrated near-to-eye display and gaze tracker. In *Proc. SPIE Photonics Europe* (2010), 77230.
16. Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P. and Raisamo, R. Gaze gestures and haptic feedback in mobile devices. In *Proc. of CHI'14*, ACM Press (2014), 435-438.
17. LaFrance, M. and Mayo, C., 1978. Cultural aspects of nonverbal communication. *International Journal of Intercultural Relations*, (1978) 2(1), 71-89.
18. Lee, J.Y., Park, H.M., Lee, S.H., Shin, S.H., Kim, T.E. and Choi, J.S. Design and implementation of an augmented reality system using gaze interaction. *Multimedia Tools and Applications* (2014), 265-280.
19. Lucero, A., Lyons, K., Vetek, A., Järvenpää, T., White, S. and Salmimaa, M. Exploring the interaction design space for interactive glasses. In *Proc. of CHI EA'13*, ACM Press (2013), 1341-1346.
20. Lucero, A. and Vetek, A. NotifEye: using interactive glasses to deal with notifications while walking in public. In *Proc. of ACE'14*, ACM Press (2014), 17.
21. Majaranta, P. and Rähkä, K.J. Twenty years of eye typing: systems and design issues. In *Proc. of ETRA'02*, ACM Press (2002), 15-22.
22. McCarthy, J. and Wright, P. Technology as experience. *Interactions* (2004), 11(5), 42-43.
23. Michalco, J., Simonsen, J.G. and Hornbæk, K. An Exploration of the Relation Between Expectations and User Experience. *International Journal of Human-Computer Interaction* (2015), 31(9), 603-617.
24. Mulvey, F., Villanueva, A., Sliney, D., Lange, R., and Donegan, M. Safety issues and infrared light. *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies* (2011). 336 - 358.
25. Olsson, T., Lagerstam, E., Kärkkäinen, T. and Väänänen-Vainio-Mattila, K., Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. *Personal and ubiquitous computing* (2013), 17(2), 287-304.
26. Pierce, J.S., Pausch, R., Sturgill, C.B. and Christiansen, K.D. Designing a successful HMD-based experience. *Presence* (1999), 8(4), 469-473.
27. Piumsomboon, T., Clark, A., Billingham, M. and Cockburn, A. User-defined gestures for augmented reality. In *Proc. INTERACT'13*, Springer (2013), 282-299.
28. Qvarfordt, P. and Zhai, S. Conversing with the user based on eye-gaze patterns. In *Proc. of CHI'05*, ACM Press (2005), 221-230.
29. Selker, T., Lockerd, A. and Martinez, J. Eye-R, a glasses-mounted eye motion detection interface. In *Proc. of CHI EA'01*, ACM Press (2001), 179-180.
30. Stellmach, S. and Dachsel, R., Look & touch: gaze-supported target acquisition. In *Proc. of CHI'12*, ACM press (2012), 2981-2990.
31. Sugano, Y. and Bulling, A., Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency. In *Proc. of UIST '15* (2015), 363-372.
32. Utsumi, Y., Kato, Y., Kunze, K., Iwamura, M. and Kise, K. Who are you?: A wearable face recognition system to support human memory. In *Proc. of AH'13*, ACM Press (2013), 150-153.
33. Vidal, M., Bulling, A. and Gellersen, H. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. of Ubicomp'13*, ACM Press (2013), 439-448.



Paper 3

Deepak Akkil and Poika Isokoski. 2016. Gaze Augmentation in Egocentric Video Improves Awareness of Intention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1573-1584. DOI: [10.1145/2858036.2858127](https://doi.org/10.1145/2858036.2858127)

© ACM 2016, Reprinted with permission.

Gaze Augmentation in Egocentric Video Improves Awareness of Intention

Deepak Akkil

Tampere Unit for Computer-Human
Interaction
University of Tampere, Finland
deepak.akkil@uta.fi

Poika Isokoski

Tampere Unit for Computer-Human
Interaction
University of Tampere, Finland
poika.isokoski@uta.fi

ABSTRACT

Video communication using head-mounted cameras could be useful to mediate shared activities and support collaboration. Growing popularity of wearable gaze trackers presents an opportunity to add gaze information on the egocentric video. We hypothesized three potential benefits of gaze-augmented egocentric video to support collaborative scenarios: support deictic referencing, enable grounding in communication, and enable better awareness of the collaborator's intentions. Previous research on using egocentric videos for real-world collaborative tasks has failed to show clear benefits of gaze point visualization. We designed a study, deconstructing a collaborative car navigation scenario, to specifically target the value of gaze-augmented video for intention prediction. Our results show that viewers of gaze-augmented video could predict the direction taken by a driver at a four-way intersection more accurately and more confidently than a viewer of the same video without the superimposed gaze point. Our study demonstrates that gaze augmentation can be useful and encourages further study in real-world collaborative scenarios.

Author Keywords

Video-based collaboration; Gaze tracking; Wearable computing; Intention prediction.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

Recent technological advancements in mobile hardware and network connectivity enable easy and seamless video communication almost anywhere. Mobility offered by the present day video communication systems like mobile

phones and tablets enables new forms of communication and collaboration between remote participants. There is a trend to move beyond conversation-only video communication towards re-purposing video to share activities and experiences [2,15,27].

People use video communication systems to show things to talk about in the environment (e.g., an interesting person, a new device or tour of a flat) [22,28], to co-ordinate a joint activity (e.g., discuss what to buy while in a store) [17] or to collaborate to achieve a specific goal (e.g., to help in repairing a complex machine) [11,17]. Such uses of video communication introduce a new set of challenges to efficiently support video as a collaborative activity space [2].

Further, there has been growing interest in wearable cameras (e.g. the Go Pro cameras) and head-mounted display devices (e.g., the Google glass) that can capture egocentric (first-person view) videos. Video communication through such devices can be useful to mediate shared activities and support physical collaboration which requires using hands [15,17].

Physical co-presence during collaboration provides many different sources of information (e.g. eye gaze, facial expression and body orientation) to help establish joint focus of attention between collaborators, monitor comprehension, and proactively help and repair the conversation [10,11]. Previous research on video-based collaboration has shown the need to provide cues in addition to shared visual information to indicate the user's focus of attention, so as to improve the awareness of the remote partner [11,12]. Gaze tracking could be used in egocentric video-based collaboration to provide accurate awareness of the collaboration partner's visual attention.

Gaze-tracking technology has been maturing from a desktop-based assistive technology to easy-to-use wearable solutions (e.g., FOVE, a commercial virtual reality headset with gaze-tracking [8]). Using gaze-tracking capable head-mounted devices in a video-mediated collaborative task could enable the collaborator to see the egocentric video superimposed with information of the gaze. Such gaze-

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.

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

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

ACM 978-1-4503-3362-7/16/05...\$15.00

DOI: <http://dx.doi.org/10.1145/2858036.2858127>

augmented video could potentially be beneficial in the collaboration. For example:

- Help deictic referencing by using gaze as a pointing mechanism.
- Improve situational awareness and help enable grounding in communication (e.g., “*he understood because he is now looking at it*”).
- Enable collaborators to infer each other’s intention, based on the context and gaze behaviour.

Previous work has investigated the use of gaze augmentation in egocentric videos for collaborative tasks [9,11]. However, the study did not show measurable benefits of augmenting gaze information to the video, possibly because the collaboration as a whole is a complex system where the effects of gaze may have been masked by other factors. However, the absence of clear benefits also casts doubt to the whole notion that gaze information would be useful. To verify that benefits do exist, we deconstructed a collaborative navigation scenario (e.g. a scenario where a remote partner guides a car driver to a specific location), to find sub-tasks where the effect of gaze information might be beneficial. In this process we found that the awareness of the intention of the remote collaborator may show in the ability to predict turns at road intersections.

Thus, to validate the hypothesis of improved awareness of intention with gaze information, we conducted a lab study where the participants viewed a gaze-augmented egocentric video of another person in a simulated car driving task. The task of the viewer was to guess the direction the driver would continue on to at four-way intersections. This paper presents the results of the study and discusses the key findings. The next section introduces the concept of gaze augmentation in egocentric videos and summarizes the potential benefits of the technology for physical collaboration.

GAZE-AUGMENTED EGOCENTRIC VIDEO

There are different ways of visualizing gaze information in the video recorded from a head-mounted camera. The simplest way is to show the point of gaze as an abstract visual element, like a small dot (See Figure 1) or ring in each frame of the video. In the following section, we briefly discuss the potential advantages of gaze-augmented video in physical collaboration.

Gaze information for deictic referencing:

Communicating deixis is an important part of collaboration [17]. In present day wearable video communication systems, there are two ways of referencing spatial information: using lengthy verbal descriptions or using hands or other physical pointers visible in the camera frame. Using speech can be ambiguous and time-consuming [26]. Hand pointing can be cumbersome and it is not always

accurate due to parallax. It is also not suitable for tasks



Figure 1. A frame taken from gaze augmented video of a person operating a coffee machine. The red dot on the coffee mug shows the gaze point estimate from the eye tracker.

requiring both hands. Eye gaze naturally carries deictic information, people generally look at objects that they are talking about even when not explicitly pointing at them [13]. If gaze information is available in the video, a collaborator could simply look at an object and say *this* to refer to the object instead of longer verbal description or hand pointing.

Attention cue, situational awareness and grounding

Human gaze is closely related to visual attention. Therefore, gaze information in video can lead to a more precise awareness of the other person’s attention and potentially ease the effort for co-ordination of joint attention. Furthermore, gaze augmentation affords an additional channel for grounding, improving redundancy and reducing ambiguity in communication.

Cooper notes that people tend to look at elements in the visual field when the element or a semantically-related item is referenced in the speech they hear [4]. Gaze behaviour during conversation can hence give an indication of whether the speech was heard and understood. For example, if the collaboration partner says “*the object on your right*”, and the listener’s gaze wanders in the opposite direction, the partner receives an indication that the utterance may have been misunderstood and can then pro-actively repair the conversation.

Gaze-augmented video for collaborative purposes could improve the two key aspects of coordinating a joint activity: situational awareness and conversational grounding.

Intention awareness:

Human gaze, when combined with the contextual information, also carries cues of intention. For example, a nurse might anticipate what equipment is needed next by the surgeon based on where the surgeon is currently looking at [25]. Two functionally distinct types of gaze fixation have been identified in naturalistic tasks: guiding fixation

that is relevant to the current sub-task and look-ahead fixation, which has a role in gathering information and planning for future actions [24]. Look-ahead fixations are task-dependent and occur several seconds before the action. Such fixations are very common in everyday physical tasks. For example, when we are approaching a sink to wash our hands, our eyes may already look at the soap dispenser to locate its position and plan future motor tasks [29]. Look-ahead fixations are a reliable predictor of an upcoming action [24]. To efficiently collaborate, it is often advantageous to not just know what your partner is doing, but also what they are planning to do [37]. Awareness of the partner's intention is a key for successful collaboration. Knowing where the collaborator is looking, in tandem with the contextual information from the egocentric video, could help better predict the conversation partner's intentions. Understanding the partner's intention enables more pro-active assistance from the viewer that could lead to time savings and more confidence in the collaboration.

The following section provides an overview of the related work in egocentric video-based collaboration, gaze awareness in egocentric videos and gaze awareness in desktop-based collaboration scenarios.

RELATED WORK

Egocentric video-based collaboration

Johnson *et al.* [16] studied the effect of mobility on video-based collaboration by comparing a hand-held tablet camera and head-mounted camera based collaboration in both static and dynamic task settings. They found that in tasks requiring higher mobility, a head-mounted camera condition provided a more consistent view of the task space, thereby improving the collaborative behaviour from re-active to pro-active. Zheng *et al.* presented a wearable HMD-based solution for industrial maintenance supporting collaboration between indoor experts and field worker [39]. Procyk *et al.* [31] used egocentric videos to enable shared experiences between remote geocaching players. Fussell *et al.* [10] studied the role of videos recorded from a head-mounted camera in a collaborative bike repair task and noted the need to provide additional cues, in addition to the video, to indicate the user's focus of attention.

There is growing interest on the use of egocentric videos to support remote collaboration and share experiences. The results of previous research suggest that head-mounted cameras are suitable in tasks involving high mobility and could also benefit from additional cues indicating the user's focus of attention.

Gaze awareness in egocentric videos for collaborative physical tasks

Gaze awareness in remote collaborative physical tasks was previously studied by Fussell *et al.* [9,11]. They studied collaboration efficiency in a mentoring style, robot building task, using different communication medium: side by side,

audio only, head-mounted camera with gaze tracking and a stationary scene camera. In the head-mounted camera with gaze tracking condition, the remote expert used an online manual and also saw the video from the head-mounted camera with a crosshair (+) symbol showing the detailed focus of visual attention of the worker in the video [8]. They found that despite the hypothesised potential of the technology, gaze awareness did not improve collaboration. Collaborators were most efficient in the side by side condition, followed by the stationary scene camera condition. They highlighted that the results could be due to technical issues in calibration of gaze tracker and gaze-tracking accuracy in mobile environment and concluded that head-mounted eye-tracking systems "may not yet be robust enough for actual field applications" [11].

Gaze-tracking technology has been improving since the studies by Fussell *et al.* [9,11] in 2003. The growing trend in gaze-tracking research and application development is to move towards more natural mobile settings [3]. We feel that the technical difficulties encountered in the previous work have been resolved to a large extent in the current technology. In our study, we further investigate the usefulness of gaze-augmented egocentric video as a medium for collaboration. We specifically target the potential of gaze augmentation in egocentric videos for improving task-related intention awareness of the collaboration partner.

Visual focus of attention in egocentric videos

Head-mounted cameras provide a view tied with the orientation of the head. Head orientation provides a coarse indication of our visual focus of attention [7,36]. However, gaze control is a co-ordinated activity which may involve movement of eyes, head and trunk. Small shifts in visual focus of attention (< 10 degrees) are typically performed using eye movement only and do not involve any movement of the head [19]. Previous research also shows that there is no simple rule on how people change their visual focus of attention using head and eye movement. There is large variability among people in the usage of head orientation to change gaze direction [36]. Many previous studies have also tried to estimate user's point-of-gaze in egocentric videos using approaches of visual saliency [38], based on hand-eye coordination strategies [21] and head-movements [20]. However, the angular gaze prediction error in these studies varies from 8 – 12 degrees.

We think that a more accurate awareness of the visual focus of attention will be beneficial in remote collaboration. In our study, we hence rely on a head-worn gaze tracker to provide accurate gaze information which is then augmented in the egocentric video.

Gaze sharing in desktop-based collaboration

Qvarfordt *et al.* studied the use of one-directional gaze awareness between user and a remote tourist consultant in a trip planning task [32]. The user's gaze position was

superimposed on a shared map interface on the computer screen as a multi-coloured dot, while the user and the consultant collaborated by voice. Their results suggest that gaze cues not only help spatial referencing but also convey interest, aid topic-switching, reduce communication ambiguity and help attain grounding in communication. Brennan *et al.* used networked gaze trackers to study the use of shared gaze in visual search tasks on a computer screen between two remotely-located participants [1]. The gaze cursor of the remote participant was visualized as a yellow ring on the display. They found that the shared gaze condition outperformed shared voice and also shared gaze plus shared voice conditions. Gaze has a distinct advantage over voice in spatial referencing tasks. Their results suggest that it is possible to achieve grounding in joint activities using shared gaze alone. Neider *et al.* [26] studied the problem of deictic referencing between two stationary collaborators using a shared display under time pressure. The experimental setup used three conditions wherein the collaborators were able to communicate using shared voice, shared gaze or both. They found that shared gaze is more efficient than speech, for rapid communication of spatial information. Stein and Brennan [35] studied the effect of seeing another person's gaze in a software debugging task and found that gaze information, even if produced without the intention to communicate information, could provide useful cues to the viewer to solve similar tasks. Sharma *et al.* [34] studied the role of shared gaze in online learning by augmenting the teacher's gaze in MOOC (Massive Open Online Course) video. They found that gaze augmentation made the video easier to follow for students.

There is a lot of evidence from desktop-based collaboration studies that knowing your partner's gaze information could benefit the interaction. Egocentric video-based collaboration is markedly different from desktop-based situations, due to the inconsistent visual information between partners induced by the limited field of view of cameras and the complexity induced due to mobility. Our study aims to build on this previous knowledge of benefits of gaze awareness in desktop-based collaboration scenarios and verify whether some of the benefits are also present in egocentric videos.

HYPOTHESIS

While the existence of look-ahead fixations is well known in real-world tasks, it is not clear that people are able to detect such patterns from gaze-augmented egocentric videos, and combine it with other contextual information to infer task-related intention of the partner. Verifying this was one of the main motivations for the work reported in this paper.

Informed by previous research on gaze awareness in co-present and desktop-based collaboration, we formed the following hypotheses.

1. Gaze augmentation in video improves the observer's awareness of task-related intention.
2. When gaze information is available in the video, the gaze behaviour of the viewers of the video will be more closely tied with the gaze behaviour of the actor.

The second hypothesis, if true, leads to more similar foci of attention that further increase the likelihood of shared understanding of the scene and the objects of most immediate relevance at a given time. However, the second hypothesis may or may not be true, even if the first part is true. It is only one of the mechanisms that could explain the first hypothesis.

METHOD

To test our hypotheses, we designed a controlled experiment with a simulated car driving scenario as the representative task. The viewers of the driving video were required to predict the direction the driver will take at a four-way intersection.

There were multiple reasons for selecting the driving task in our study. First, driving is a common everyday activity. If gaze-tracking capable head-mounted devices become commonplace they will also be used while driving. Second, driving is a potentially collaborative activity, real-time video from the driver introduces new possibilities for remote collaboration, e.g. remote monitoring of driving, remotely providing navigation instructions [30]. Third, gaze behaviour while driving is extensively studied, confirming the occurrence of look-ahead fixations in turn driving [18,29]. Fourth, the driving task offers a clear quantitative success criterion (number of correct predictions), and is relatively fast so that we could easily measure a number of repetitions without tiring our participants.

The objective of the study was to understand the value of gaze-augmented egocentric video for the collaboration partner's intention prediction. Therefore, we wanted to mask the additional cognitive load associated with real-time collaboration. The experiment was hence conducted in two phases. In the first phase, we recorded the driving videos of actors and the second phase included the lab study, where participants viewed the recorded videos and predicted the direction the driver will take at a four-way intersection.

Phase 1: Video recording

We invited two actors (both 27 years old, one male and one female) from the University community to record the gaze behaviour while driving in a car simulator. Both the actors had valid driver's licences and a number of years of experience in driving, making it safe to assume that they had established patterns of driving behaviour that would be representative of real driving situations. We used the *city*

*car driving*¹ simulator application. The driving simulator set up is shown in Figure 2.

First, the actors familiarised themselves with the simulator, by driving freely for a few minutes. The actors then wore the Ergoneer Dikablis professional 60Hz binocular head-mounted gaze tracker with a 90 degree field of view scene camera, for recording gaze behaviour and the egocentric video. The actors were unaware of the purpose of the study and were simply instructed to drive like they would normally do. This was required because driver's awareness of the purpose of the study may have led to a bias (e.g. actors might have exaggerated their eye movements allowing easier prediction of turn direction). After calibrating the tracker, we recorded 15 videos of each actor driving through the same four-way junction on a road with a single lane to each direction. We set the simulator to present a low volume of vehicular traffic on the road. The crossroad had no traffic signals and the participants were instructed not to use the turn indicator. The actors took all three possible directions (left, right and straight on) five times each. When they had passed the intersection, they took a U-turn and stopped the car. During the stop they were given verbal instruction by the moderator on where to turn next. In addition to the turn videos, we also recorded a few minutes of free driving. The videos were recorded at a resolution of 1920x1080px at 30fps. Soon after the recordings, the purpose of the study was explained to actors. The actors were given an option to withdraw their participation at this point. None of them did.



Figure 2. The driving simulator setup.

Phase 2: Video viewing

Participants

We recruited 12 volunteer participants (6 females and 6 males), aged between 20 and 43 years from the University community. All participants had normal/corrected to normal vision. Nine of the participants were previously familiar with gaze-tracking technology.

Design

We chose a within-subject design with one independent variable (presence of the gaze-point indicator on the video). The two experimental conditions were labelled as follows:

Gaze: The video from the driver's head-mounted camera included the driver's gaze point.

NoGaze: The control condition in which the video from the driver's head-mounted camera was shown without modification.

There were three dependent measures: accuracy of prediction (number of correct predictions of driver's direction), subjective confidence in prediction (median value of self-reported confidence for the predictions) and gaze distance (the average distance between driver's and synchronous viewer's focus of attention).

For each participant we utilized videos from the two drivers so that we got *Gaze* and *NoGaze* data from each participant. Two drivers also helped to reduce the likelihood of driver-dependent behaviours biasing our results. The order of the experimental conditions and the assignment of drivers were counter-balanced between participants.

To test for differences between the *Gaze* and *NoGaze* conditions we used a non-parametric pair-wise (Monte Carlo) randomization test [6]. In randomization test, the null hypothesis is that the pair-wise differences are equally likely to be positive or negative. Repeated resampling ($n=100,000$) with random assignment of sign for the difference between the conditions gives us a sampling distribution of the mean difference. The observed mean difference is compared to the sampling distribution to estimate how likely the observed difference is by chance. A t-test may also have been suitable but, in the absence of any assumption regarding the nature of distribution of the dependent variables, we opted for the non-parametric alternative.

Apparatus

We used a custom C# software based on Microsoft .NET 4.5 framework to present the video stimuli and the on-screen questionnaire. The participants viewed the video on the display of a Tobii T60 gaze tracker. The screen and gaze data of the participants was recorded and analyzed using Tobii Studio (version 3.3) application.

Gaze augmentation

For the gaze-augmented video, the video recorded from the head-mounted camera was superimposed with the gaze data of the driver. The gaze point was visualized as a red semi-transparent circle and was 47px in diameter in the viewer's display (see Figure 3).

Visualizing the raw gaze data was often jittery and difficult to follow for the viewer, so we used a recursive filter to smoothen the gaze data. We borrowed the filter from Quartfordt *et al.* [32].

¹ <http://citycardriving.com/>

$$y(i) = W * x(i) + (1 - W) * y(i - 1), \tag{1}$$

where $y(i)$ was the current displayed position of the eye gaze; $x(i)$ was the current gaze position, W was the percentage weight for the current gaze position and $y(i-1)$ was the last displayed gaze location.



Figure 3. A frame from the gaze augmented video seen by the participants.

A larger weight for the current gaze position would make the gaze cursor more responsive but also more jittery. We used different values of W for horizontal ($W=0.20$) and vertical ($W=0.05$) directions. Based on the pilot tests, the gaze augmentation along the horizontal axis was the more important component that helped in the task and we wanted to make the horizontal component responsive while reducing the jitteriness along the vertical axis.

Procedure

At the beginning of the experiment, the participants signed an informed consent form and filled in a short background questionnaire. The participants were then seated in front of a Tobii T60 gaze tracker, which was then calibrated. The task for the participants was to view the recorded driving videos and predict the direction the driver would take at the four-way intersection. The videos were presented on the display at a resolution of 1280x720px centrally aligned at 30fps.

In the video, the approach of the car to the intersection took 10-15 seconds. The video automatically paused just before the intersection, before any turn-related cues were available through the steering-wheel motion. At this point, the screen turned white and the participants were presented with an on-screen questionnaire. The participants were asked which direction they thought the driver was going to take (options *left*, *right* or *straight* on presented as radio-buttons) and how confident they were with the prediction (7-point Likert scale presented as radio buttons). The participants were instructed to answer the question using the mouse.

Each condition consisted of first watching one minute of free driving video, which included driving through multiple intersections, followed by 15 short videos of the car passing through a four-way intersection and proceeding in one of

the 3 directions (five videos per direction). The videos were presented in random order. The free driving section was included at the beginning, so that the participant could adjust to the driving style and gaze behaviour of the driver. Each participant saw two sets of videos: one from one of the drivers without the superimposed gaze point and another with the gaze point from the other driver.

After completing both the conditions, a final questionnaire was used to collect the subjective opinions of the participant in relation to the two experimental conditions.

RESULTS

Prediction accuracy

Figure 4 shows the number of correct predictions of the driver’s intention in the two experimental conditions. The median value indicates that participants predicted the direction which the driver would take at a four-way intersection 26% more accurately when the video was augmented with the gaze information of the driver. The difference was found to be statistically significant using the pair-wise randomization test ($p=0.01$).

Prediction confidence

Figure 5 shows the boxplot for mean value of the subjective confidence in the prediction between the experimental conditions. The participants felt more confident about their

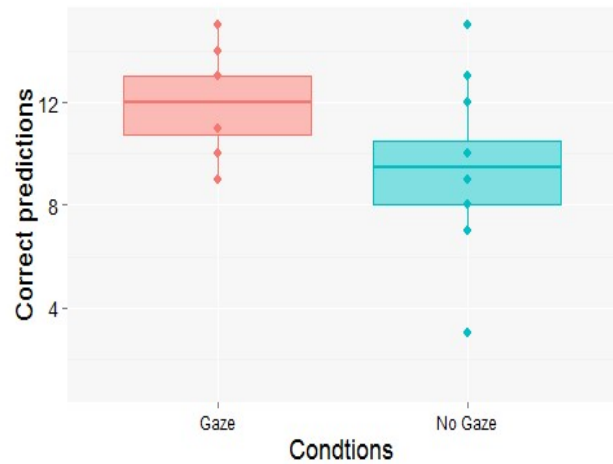


Figure 4. Number of correct predictions for the two conditions. Maximum was 15, random choice would lead to 5 correct predictions on average.

prediction when gaze information was available. The difference was found to be statistically significant using the pair-wise randomization test ($p=0.01$).

Car usage and task performance

It seems reasonable to assume that viewers with more driving experience would be able to utilize the gaze data better, because they understand where a driver needs to

look in order to safely make a given turn at an intersection. Consequently, we analyzed the results based on the self-reported car usage frequency of participants. The car usage frequency reported in 5-point Likert scale data was reduced to 2 levels by combining the top three levels, indicating high car usage, and the lower two levels, indicating low car usage frequency. Each of the resultant levels had 6 participants. There were no noticeable differences in the accuracy of prediction depending on frequency of car usage. Figure 6 shows the prediction confidence for the two levels of car usage. Participants who reported to be

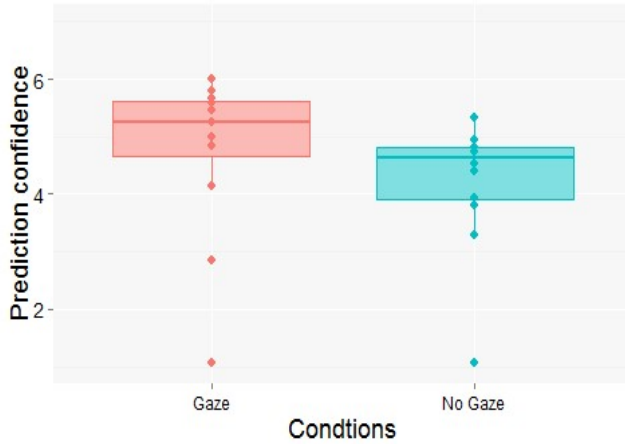


Figure 5. Subjective confidence in prediction for the two conditions.

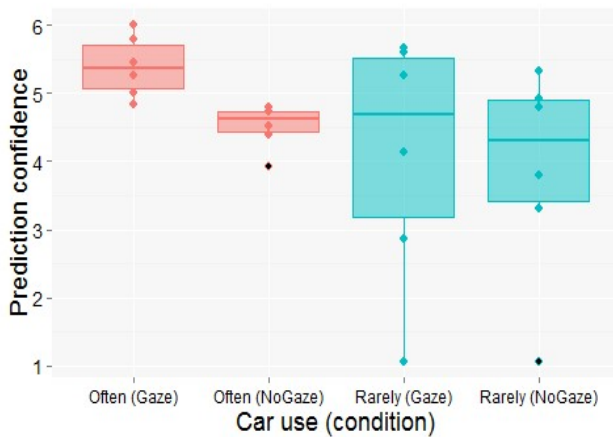


Figure 6. Self-reported car usage and prediction confidence for the two conditions.

frequent drivers, were also relatively more confident with the predictions when gaze information was available. However, these differences were not statistically significant.

Video-viewing behaviour:

The following analysis is based on the data from 10 participants. We had to exclude one participant due to data loss and another due to poor eye-tracking robustness (more than 50% samples missing).

We analyzed the effect of gaze augmentation on how the participants viewed the driving videos. Because gaze data analysis is very labour-intensive, we had to rely on samples of the 15 turns by the two drivers instead of analyzing all of them. We selected 3 videos from each driver. The videos were selected such that there was one video representative for each turn direction and there was a high difference in the number of correct predictions between the two

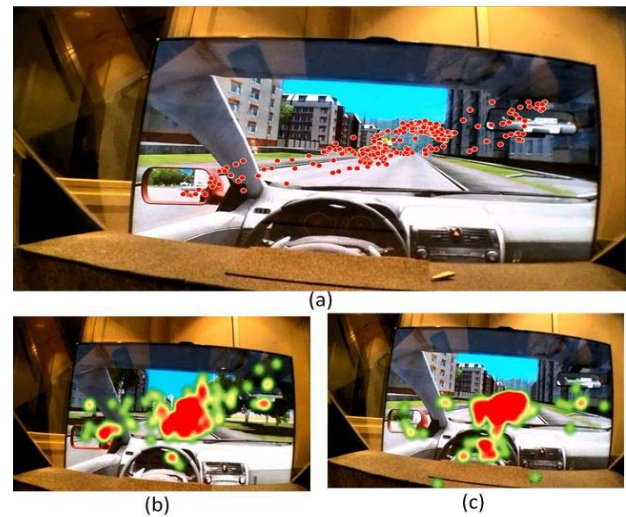


Figure 7. (a) Gaze behaviour of the driver for a single video, (b) Heatmap showing the viewing behaviour of participants when the video was augmented with gaze point and (c) Heatmap showing the viewing behaviour of participants without gaze point.

experimental conditions for the video, suggestive of possible differences in viewing behaviour. Figure 7(a) shows the gaze behaviour of the driver during approach to the intersection in which the target direction was straight. Figure 7(b) and 7(c) show the heatmap of the aggregate gaze behaviour of the participants for the two experimental conditions. It appears that in sub-figure 7(b), where participants saw the driver's point of gaze, their viewing was spread more widely along the axis that the driver's gaze travelled.

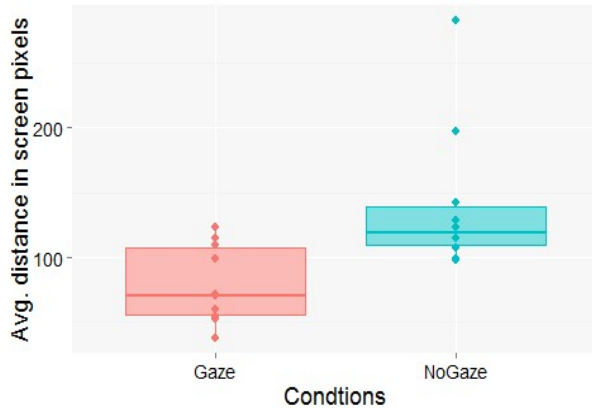


Figure 8. Average distance in screen pixels between driver's focus of attention and viewer's synchronous focus of attention.

We computed the average distance between the driver's gaze point, indicated by the gaze visualization in the video, and the synchronous viewer's gaze point. The metric was computed from the moment each video started until the moment the video paused to show the on-screen questionnaire, upon reaching the intersection. A large value of this metric indicates that viewers were either not focusing on the same areas in the shared visual field as the driver, or not doing so synchronously. A smaller value suggests that both the driver and the viewer of the video looked at the same areas in the scene synchronously, indicating a similar perceptual input of the task environment. Figure 8 shows the average gaze distance in screen pixels of viewer display (resolution: 1280 x 720px) between the driver and viewer for the two conditions. The difference was found to be statistically significant using the pair-wise randomization test ($p=0.004$).

Subjective Evaluations:

10 out of 12 participants felt that the gaze information of the driver helped predict the turn direction confidently. This was also evident in the comments from the participants:

P7: After seeing the example videos, I learned how the driver acts before turning. It (gaze augmentation) gave me more information to base my judgement on.

P9: I was more confident about my prediction with gaze overlay.

One participant felt that gaze augmentation was sometimes misleading. When gaze information was not available, participants said they relied on subtle clues based on speed of the car, position of the car in the lane, head movement of the driver (seen as turn of the egocentric video) and driving dynamics with other vehicles, to predict the turn direction.

DISCUSSION

Our study focused on the usefulness of gaze augmentation in egocentric video to provide better awareness of the intention of the actor to the viewers. Viewers of gaze-augmented video were more aware of the focus of attention

of the driver and could effectively use that information in combination with the contextual information from the video to predict the intention of the driver to proceed in a specific direction at an intersection.

Some of the clues were rather simple to decode. Glances to the left or right often preceded a turn in the same direction. However, there were also subtler clues available in the driver's gaze behaviour that might have helped the participants anticipate the turn direction. For example, the driver sometimes looked to the opposite direction of the turn, possibly to ensure that there were no approaching vehicles. Also, in the presence of oncoming traffic and when intending to take a left turn, the driver often looked at the car approaching from directly ahead to interpret whether it intended to stop or not. This was necessary to avoid a collision at the intersection.

In our simulated driving task there were many different clues that could have potentially helped predict the turn direction (e.g., head orientation of the driver, speed of the car). In both the conditions, the participants could on average predict more than 50% of the turns correctly. Except for one participant who predicted all the turns correctly in both the conditions, all other participants made a few mistakes. The selected task hence presented a scenario of moderate difficulty, with many different contextual channels on which to base the prediction. Our results indicate that the awareness of the drivers gaze behaviour improved the intention awareness over the level that was achieved utilizing the other information sources

The gaze-tracking results of the participants shows that the gaze-augmented video enabled viewers to synchronize their eye movement with the eye movement of the driver. The driver and viewers looked at close by areas in the visual scene at the same moment. This suggests that participants could follow the gaze visualization of the driver. It may not be surprising that the gaze of the driver, visualized as a smoothly moving semi-transparent red circle, attracted the viewer's attention. The close co-ordination of gaze behaviour between the driver and the participants also indicates that gaze augmentation could help the collaborators to enable joint attention and create a feeling of "being on the same page" continually. Such a coordinated gaze behaviour also improves the shared understanding of the scene and could also explain the increased awareness of the driver intention.

Awareness of the partner's visual focus facilitates joint attention and a 'perceptual common ground' in collocated collaborative scenarios. Such a perceptual common ground provides insight into the interaction partner's mind [33]. While this seems almost obvious in theory, it has turned out that measuring the benefit of gaze augmentation in video communication is difficult. For example, Fussell *et al.* [10] did not find a statistically significant benefit of gaze awareness in remote collaboration. In our pilot testing, we went through a number of collaboration tasks where the

possible benefit of gaze-augmented video was too small or the task specific measurement noise too large, for statistical significance with the small number of participants that we can afford to include in our studies. We were satisfied to find this driving simulator task that shows a clear advantage. However, much work remains to be done in exploring the different forms of gaze-augmented video communication to build a fuller picture of the benefits and challenges.

There are two important factors that need to be considered when evaluating the value of gaze augmentation of video. Firstly, the task-related expertise of the collaborators may influence how they use the gaze information of the partner. In our study, the participants who reported to be more frequent car users seemed to be more confident with the prediction than participants who reported to use cars rarely. While this result was not statistically significant, it is indicative of that a task-specific expertise may enable viewers to more confidently rely on the gaze augmentation to predict the partner's intentions. An example scenario would be of an expert support staff helping an experienced field worker troubleshoot complex machinery.

Secondly, there may be a learning effect associated with the use of gaze-augmented video. In our study, the two drivers showed variability in driving style (turn trajectory, approach speed etc.) and gaze behaviour (how often they looked at the rear-view mirrors, how often they looked at other directions to check for approaching vehicles etc.). Our participants had to learn and adapt to the driving style and gaze behaviour of the specific driver. In real-world collaboration scenarios, it could be expected that frequent collaborators would gradually learn the dynamics of the gaze behaviour of the partner in relation to the task and could potentially learn to use the gaze information channel more efficiently with experience.

Before this study, we did not know that gaze augmentation can add value to the egocentric video used for collaboration. The ability to anticipate your collaboration partner's next move can be extremely advantageous in collaboration by giving a sense of "being in control" over the interaction and potentially improving the collaboration efficiency. Now that we know that gaze augmentation can indeed improve awareness of intention, we can investigate further in real video-based collaborative scenarios. Gaze augmentation of video recorded from a head-mounted camera is an exciting and practical technology to explore for real-world video-based collaboration.

Limitations and future work

Our study has a few limitations. First, our study focused specifically on the usefulness of gaze-augmented video as a medium and did not consider an actual collaborative scenario. In real-world collaboration, the viewer of the video could be under additional cognitive load to interpret the partner's speech, respond to the partner and other time pressures imposed by the collaboration. Future work can

examine how the additional cognitive load influences intention prediction in gaze-augmented video-based collaboration.

Second, the study used a simulated driving task. The limited field of view offered by the simulator meant that the participants performed limited head movement to orientate their gaze direction, as opposed to real-world driving. Real-world driving may involve frequent head movement and it is known to be important for driving intent analysis in lane change tasks [5]. This means that both the conditions may have potentially performed better in a real-world task than in the simulator. Further research would be required to validate the benefit of gaze augmentation in driver intention prediction in real-world scenarios.

Third, the information of intention transmitted by gaze in our study would have been trivial to communicate verbally. People sometimes verbally narrate their actions or intentions, if it can be useful for collaboration [23]. Our intention is not to propose that gaze would be used instead of these other practical modalities. Instead, the results should be seen as evidence of the possibility that gaze visualization may be able to complement other communication channels, when they are occupied in other tasks and also increase redundancy in communication.

Fourth, the small sample size (N=12) is a clear limitation of the study. Our conclusions on the effect of task-specific expertise on intention prediction are preliminary and a larger number of participants could have been beneficial. Further work is required to validate this result.

Our study used a simulated car driving task as a representative collaborative scenario of realistic complexity. Gaze behaviour, especially the duration and frequency of look-ahead fixations, varies considerably, depending on the task being performed [14,24]. While performing complex natural tasks, look-ahead gaze fixations towards task-relevant targets are made without conscious intervention and such fixations have been reported in a variety of real-world tasks involving a sequence of action, like model-building tasks [24] or sandwich making [14] where the user fixates on relevant objects 3 - 6 seconds prior to the subsequent reach operation. Hayhoe *et al.* noted that such fixations are a ubiquitous aspect of natural behaviour [14]. We speculate that the benefit of gaze augmentation in egocentric videos for intention prediction may be available also in such natural tasks, enabling the viewers of the real-time gaze-augmented video to use the gaze information along with other contextual information, to be more aware of the task-related intention of the partner. However, further work is required to validate the benefits of gaze augmentation in other collaborative scenarios.

Our study focused on only one of the potential benefits of gaze augmentation in video-based collaborative scenarios, i.e., intention prediction. Further research is required to

understand its benefits in scenarios involving spatial referencing and its role in improving situational awareness and achieving common ground in egocentric video-based collaboration.

CONCLUSION

Our study is aimed as the first step to measuring the usefulness of gaze-augmented video. The study examined the effect of gaze augmentation in a video recorded from a head-mounted camera in a simulated driving task, to improve the awareness of the driver's intention. Our results confirmed that viewers of gaze-augmented videos can efficiently use the actor's gaze information along with the contextual information available through other channels, to predict intention. Our results show the potential utility of gaze-augmented egocentric video for collaboration and encourage further exploration in real-world collaborative tasks.

ACKNOWLEDGMENTS

We thank the members of Tampere Unit for Computer-Human Interaction who provided helpful comments on different versions of this paper. The work was partly funded by Academy of Finland, projects Haptic Gaze Interaction (decisions 260026 and 260179) and Mind Picture Image (decision 266285).

REFERENCES

1. Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, Gregory J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*. Issue 3, 1465-1477. <http://dx.doi.org/10.1016/j.cognition.2007.05.012>
2. Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. 96-105. <http://doi.acm.org/10.1145/2317956.2317973>
3. Andreas Bulling and Hans Gellersen. 2010. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing* 9. <http://doi.org/10.1109/MPRV.2010.86>
4. Roger M Cooper. 1974. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6.1 (1974): 84-107. [http://dx.doi.org/10.1016/0010-0285\(74\)90005-X](http://dx.doi.org/10.1016/0010-0285(74)90005-X)
5. Anup Doshi and Mohan Manubhai Trivedi. 2009. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation System*. 10, 3. 453-462. <http://dx.doi.org/10.1109/TITS.2009.2026675>
6. Pat Dugard. 2014. Randomization tests: A new gold standard? *Journal of Contextual Behavioral Science* 3, 1: 65–68. <http://doi.org/10.1016/j.jcbs.2013.10.001>
7. Tom Foulsham, Esther Walker, Alan Kingstone. 2011. The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*. 51, 17. 1920-1931. <http://dx.doi.org/10.1016/j.visres.2011.07.002>.
8. FOVE virtual reality headset. <http://www.getfove.com/> (accessed: 5 January 2016)
9. Susan R. Fussell and Leslie D. Setlock. 2003. Using Eye-Tracking Techniques to Study Collaboration on Physical Tasks: Implications for Medical Research. *Unpublished manuscript*, Carnegie Mellon University, (2003), 1–25.
10. Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of communication: effects of shared visual context on collaborative work. In *Proceedings of Computer supported cooperative work (CSCW '00)*, 21-30. <http://doi.acm.org/10.1145/358916.358947>.
11. Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. 513-520. <http://doi.acm.org/10.1145/642611.642701>.
12. Darren Gergle, Robert E. Kraut, Susan R. Fussell. 2013. Using Visual Information for Grounding and Awareness in Collaborative Tasks. *Human-Computer Interaction*. 28, 1. 1-39
13. Zenzi M. Griffin and Kathryn Bock. 2000. What the eyes say about speaking. *Psychological science*. 11, 4. 274–279. <http://dx.doi.org/10.1111/1467-9280.00255>.
14. Mary M. Hayhoe, Anurag Shrivastava, Ryan Mruczek, Jeff B. Pelz. 2003. Visual memory and motor planning in a natural task. *Journal of Vision*. 3. 49-63. <http://dx.doi.org/10.1167/3.1.6>.
15. Kori Inkpen, Brett Taylor, Sasa Junuzovic, John Tang, and Gina Venolia. 2013. Experiences2Go: sharing kids' activities outside the home with remote family members. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. 1329-1340. <http://doi.acm.org/10.1145/2441776.2441926>.
16. Steven Johnson, Madeleine Gibson, and Bilge Mutlu. 2015. Handheld or Handsfree?: Remote Collaboration via Lightweight Head-Mounted Displays and Handheld Devices. In *Proceedings of Computer Supported Cooperative Work & Social Computing (CSCW '15)*.

- 1825-1836.
<http://doi.acm.org/10.1145/2675133.2675176>.
17. Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '15). 957-966. <http://doi.acm.org/10.1145/2702123.2702345>.
 18. Tsuneo Kito, Masahiro Haraguchi, Takayuki Funatsu, Motoharu Sato, Michiaki Kondo. 1989. Measurements of gaze movements while driving. *Perceptual and Motor Skills*. 68.1. 19-25. <http://dx.doi.org/10.2466/pms.1989.68.1.19>.
 19. Michael F. Land. 2006. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*. 25, 3, 296-324. <http://dx.doi.org/10.1016/j.preteyeres.2006.01.002>.
 20. Teesid Leelasawassuk, Dima Damen, Walterio W. Mayol-Cuevas. 2015. Estimating visual attention from a head mounted IMU. In *Proceedings of International Symposium on Wearable Computers* (ISWC '15). 147-150. <http://doi.acm.org/10.1145/2802083.2808394>.
 21. Yin Li, Alireza Fathi, and James M. Rehg. 2013. Learning to Predict Gaze in Egocentric Video. In *Proceedings of IEEE International Conference on Computer Vision* (ICCV'13). 3216-3223. <http://dx.doi.org/10.1109/ICCV.2013.399>.
 22. Christian Licoppe and Julien Morel. 2009. The collaborative work of producing meaningful shots in mobile video telephony. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '09). Article 35. <http://doi.acm.org/10.1145/1613858.1613903>.
 23. Paul Luff, Christian Heath, David Greatbatch. 1992. Tasks-in-interaction: paper and screen based documentation in collaborative activity. In *Proceedings of ACM conference on Computer-supported cooperative work* (CSCW '92). 163-170. <http://doi.acm.org/10.1145/143457.143475>.
 24. Neil Mennie, Mary Hayhoe, Brian Sullivan. 2007. Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research*. 179(3). 427-442. <http://dx.doi.org/10.1007/s00221-006-0804-0>.
 25. Bonnie A. Nardi, Heinrich Schwarz, Allan Kuchinsky, Robert Leichner, Steve Whittaker, Robert Scلابassi. 1993. Turning away from talking heads: the use of video-as-data in neurosurgery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '93). 327-334. <http://doi.acm.org/10.1145/169059.169261>.
 26. Mark B. Neider, Xin Chen, Christopher A. Dickinson, Susan E. Brennan, Gregory J. Zelinsky. 2010. Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*. 17(5). 718-724. <http://dx.doi.org/10.3758/PBR.17.5.718>.
 27. Carman Neustaedter and Tejinder K. Judge. 2010. Peek-A-Boo: the design of a mobile family media space. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Adjunct* (UbiComp '10 Adjunct). 449-450. <http://doi.acm.org/10.1145/1864431.1864482>.
 28. Kenton O'Hara, Alison Black, and Matthew Lipson. 2006. Everyday practices with mobile video telephony. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06). 871-880. <http://doi.acm.org/10.1145/1124772.1124900>.
 29. Jeff B. Pelz, Roxanne Canosa. 2001. Oculomotor Behavior and Perceptual Strategies in Complex Tasks. *Vision Research*. 41 (25). 3587-96. [http://dx.doi.org/10.1016/S0042-6989\(01\)00245-0](http://dx.doi.org/10.1016/S0042-6989(01)00245-0)
 30. Bastian Pfleging, Stefan Schneegass, and Albrecht Schmidt. 2013. Exploring user expectations for context and road video sharing while calling and driving. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '13). 132-139. <http://doi.acm.org/10.1145/2516540.2516547>.
 31. Jason Procyk, Carman Neustaedter, Carolyn Pang, Anthony Tang, Tejinder K. Judge. 2014. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). 2163-2172. <http://doi.acm.org/10.1145/2556288.2557198>.
 32. Pernilla Qvarfordt, David Beymer, and Shumin Zhai. 2005. RealTourist: a study of augmenting human-human and human-computer dialogue with eye-gaze overlay. In *Proceedings of the International conference on Human-Computer Interaction* (INTERACT'05). 767-780. http://dx.doi.org/10.1007/11555261_61.
 33. Natalie Sebanza, Harold Bekkering, Günther Knoblich. 2006. Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*. 10(2). 70-76. <http://dx.doi.org/10.1016/j.tics.2005.12.009>.
 34. Kshitij Sharma, Jermann Patrick and Dillenbourg Pierre. 2015. Displaying Teacher's Gaze in a MOOC: Effects on Students' Video Navigation Patterns. In *Proceedings of the 10th European Conference on Technology Enhanced Learning*.
 35. Randy Stein and Susan E. Brennan. 2004. Another person's eye gaze as a cue in solving programming problems. In *Proceedings of the 6th international*

- conference on Multimodal interfaces (ICMI '04)*. 9-15. <http://doi.acm.org/10.1145/1027933.1027936>.
36. Rainer Stiefelhagen and Jie Zhu. 2002. Head orientation and gaze direction in meetings. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '02). 858-859. <http://doi.acm.org/10.1145/506443.506634>.
37. Karl Verfaillie and Anja Daems. 2002. Representing and anticipating human actions in vision. *Visual Cognition*. 9(1). 217-232. <http://dx.doi.org/10.1080/13506280143000403>.
38. Kentaro Yamada, Yusuke Sugano, Takahiro Okabe, Yoichi Sato, Akihiro Sugimoto, Kazuo Hiraki. 2011. Attention Prediction in Egocentric Video Using Motion and Visual Saliency. *Advances in Image and Video technology*. http://dx.doi.org/10.1007/978-3-642-25367-6_25.
39. Xianjun Sam Zheng, Patrik Matos da Silva, Cedric Foucault, Siddharth Dasari, Meng Yuan, Stuart Goose. 2015. Wearable Solution for Industrial Maintenance. In *Proceedings of Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15). 311-314. <http://doi.acm.org/10.1145/2702613.2725442>.



Paper 4

Deepak Akkil and Poika Isokoski. 2016. Accuracy of interpreting pointing gestures in egocentric view. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 262-273. DOI: [10.1145/2971648.2971687](https://doi.org/10.1145/2971648.2971687)

© ACM 2016, Reprinted with permission.

Accuracy of Interpreting Pointing Gestures in Egocentric View

Deepak Akkil

Tampere Unit for Computer-
Human Interaction
University of Tampere, Finland
deepak.akkil@uta.fi

Poika Isokoski

Tampere Unit for Computer-
Human Interaction
University of Tampere, Finland
poika.isokoski@uta.fi

ABSTRACT

Communicating spatial information by pointing is ubiquitous in human interactions. With the growing use of head-mounted cameras for collaborative purposes, it is important to assess how accurately viewers of the resulting egocentric videos can interpret pointing acts. We conducted an experiment to compare the accuracy of interpreting four different pointing techniques: *hand pointing*, *head pointing*, *gaze pointing* and *hand+gaze pointing*. Our results suggest that superimposing the gaze information on the egocentric video can enable viewers to determine pointing targets more accurately and more confidently. Hand pointing performed best when the pointing target was straight ahead and head pointing was the least preferred in terms of ease of interpretation. Our results can inform the design of collaborative applications that make use of the egocentric view.

Author Keywords

Accuracy of spatial referencing, Egocentric video, Gaze augmentation. Collaboration. Pointing.

ACM Classification Keywords

H.5.3 [Group and organization interfaces]: Collaborative Computing; Computer-supported cooperative work

INTRODUCTION

People often use different deictic (pointing) gestures to support their verbal communication. For example a person may point at a distant object with his hand or gaze, while verbally conveying details about it to an onlooker. In addition, pointing gestures are also ubiquitous in Human-Computer Interaction. The most iconic work in this area is “Put that there” [7], which used hand gestures to identify objects on screen and to propose new locations for placing them, while interacting using voice. Pointing is a foundational building block of human–human and human–computer interactions.

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.
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/10.1145/2971648.2971687>

Recent advancements in miniaturization of cameras, improvement in network connectivity, image processing and storage capabilities have led to growing popularity of wearable cameras. Wearable cameras are being increasingly used for a variety of everyday applications, like collaborating with a remote partner, interacting with near-by objects, life logging *etc.*

Unlike other body sites, head-mounted cameras (e.g. cameras integrated in smartglasses) provide an egocentric perspective, providing a consistent view of the current activity and a coarse indication of visual attention of the wearer [2]. There are many commercial head-mounted devices, equipped with a world-facing camera, e.g. Google glass, and Epson Moverio. Such devices offer remarkable potential for use in mobile collaboration [16] and have been studied in remote collaborative scenarios for sharing experiences [18,29,31], ambient telepresence [10] or to collaborate to achieve a specific goal (e.g., to help in repairing a complex machine) [12,16,38].

Communicating spatial information is an integral part of video-based collaboration [17]. Consider a scenario, where a person wearing a head-mounted camera is giving a tour of a flat to a remote viewer, while referring to and explaining interesting details about the flat. There are different ways of visually pointing at the interesting details to the remote viewer in this collaborative scenario. The first option is pointing with the hand. The hand of the wearer, when explicitly used for pointing at a distant object, is often visible in the egocentric video and could be used to convey spatial information. A second option is pointing with the head, by bringing the object of interest to the center of the camera view by turning the head. A third option becomes possible when the point of gaze of the user is superimposed on the video [2]. In such cases, the user can convey the object of interest by just looking at it.

Each of these techniques has advantages and disadvantages. Hand pointing is a natural and familiar pointing technique, however, it is not available when hands are occupied. Also, it is often less desirable in public situations, due to concerns of social acceptability. Head-based pointing is more subtle than hand pointing, however it may be difficult to perform accurately without direct feedback on the center of the egocentric view to the wearer (This is the case when no



Figure 1. An example scenario where high pointing accuracy is desirable. The person wearing the camera is pointing at a specific car (circled in red in the image)

display is available). Gaze-based pointing is natural, however it requires gaze-tracking sensors to be worn and an initial calibration procedure to be performed. Gaze data quality, *i.e.* accuracy and precision of gaze data affects the usefulness of gaze pointing. Another potential difficulty is in distinguishing the use of gaze for pointing and for perception. This is also known as the classic Midas-Touch problem in gaze interaction [24].

Wong and Gutwin [37] divided deictic referencing (*i.e.* conveying spatial information to your collaboration partner by pointing) into four stages. Mutual Orientation, staging of the gesture, production of the gesture and holding. In a collaborative scenario, the goal is to achieve a mutual understanding between the speaker and the listener, across all the four stages. Normally, all the four stages are also accompanied with verbal utterances. Wong and Gutwin note that one of the fundamental questions that must be answered before designing rich support for pointing in collaborative systems, is how accurately viewers can interpret the direction of pointing [37].

It is not known how accurately viewers of egocentric view can interpret the different pointing techniques. For example, Rumelin et al [32] conducted a *Wizard of Oz* study regarding acceptance and applicability of hand-pointing interactions with distant objects in cars. Interestingly, the *Wizard* used the egocentric video with the gaze point overlaid to accurately identify the object being pointed at, rather than interpreting the target of hand pointing. They note that “*informal preliminary tests had shown that with this video stream (egocentric view with gaze overlaid) ... the wizard could identify the objects chosen for pointing accurately*”.

Considering the growing popularity of head-worn cameras, and the ubiquitous nature of pointing in human interactions, there is a need for research to understand and systematically compare the accuracy of the different plausible pointing gestures. One should note that the accuracy required of the pointing technique depends on the task at hand. For example, the accuracy required to convey a distant object from a high-rise building (see figure 1) may be much higher than the accuracy required to point at a lone person standing close by.

We conducted an experiment in a laboratory environment to compare the accuracy of interpreting the different pointing techniques in the egocentric view. Our experiment consisted of four conditions (*hand, head, gaze, hand+gaze*). Our participants viewed egocentric videos of an actor pointing at numbers attached on walls. The participants’ task was to identify the number being pointed at and indicate how confident they were with the interpretation. The rest of this paper is structured as follows. We begin by reviewing some of the related work. Then we describe the user study and its results. Finally we discuss our key findings and their implications.

RELATED WORK

Eye, head, and hand co-ordination during natural pointing

Henriques and Crawford [14] studied hand-eye co-ordination during hand pointing and noted that people normally place their fingertip on the imaginary line joining the dominant/preferred eye and the pointing target, as opposed to pointing with the full arm vector. Similarly, Biguer et al. [6] noted that, while pointing at eccentric targets, users never align their head directly with the target or the pointing arm. Uemura et al. [36] observed that for targets at 50 degrees

lateral displacement, head movement attributes for only 62% of this displacement. However, for target orientation below 10 degrees, head orientation accounts for 93% of the displacement. Stiefelhagen and Zhu [33] noted that is large variability among people in the usage of head orientation in everyday scenarios. Griffin and Bock [13] reported that we look at objects in the environment while speaking about them, even when not explicitly pointing at them. Gaze hence naturally carries deictic information. These eye, head, and hand co-ordinations should be considered, while designing natural and expressive pointing mechanisms for the egocentric view.

Pointing in the egocentric perspective

Colaco et al. [9] developed MIME, a system for hand gestural input with head-mounted display. Tung et al. [35] studied user acceptance of hand gestures as input for smartglasses. Both of these studies relied on the use of hands to point at, and interact with, components on a head-mounted display. Kolle et al. [22] studied hand gestures seen through a head-mounted camera for interacting with ambient objects.

Use of head orientation for pointing was earlier studied by Thomas et al. [34]. They developed ARQuake, a first person augmented reality shooting game that relies on head orientation of the player to aim at targets. Kooper and MacIntyre [23] developed an augmented reality system that uses centering the ambient object in the display screen for interaction.

There are also studies that use gaze information in relation to a head-mounted camera view, to point at and interact with ambient objects, for example Baldauf et al. [5] developed Kibitzer, a head-mounted system that uses gaze information of the wearer to retrieve annotated ambient digital information.

Previous works have shown that hand, head and gaze are plausible pointing techniques in the egocentric perspective to interact with distant objects. The scope of the current study is in the context of collaborative applications that make use of the egocentric view. In our study, we focus on a scenario where the person wearing the camera, do not have an accompanying head-mounted display that could be used to provide feedback on the pointing act. From the perspective of the partner who performs the pointing gesture, hand, head and gaze are natural pointing mechanisms. However, unlike the other pointing methods, gaze information is not automatically a part of the egocentric video stream. A visualization of the gaze point needs to be added into the egocentric view, to be used in the referencing task. In our study, we also included *hand+gaze*, a combination of natural (hand) and technology-augmented (gaze) pointing technique.

Gaze sharing for spatial referencing

Neider et al. [30] studied the role of gaze sharing in referencing a spatial location under time pressure between two remote partners collaborating over a shared display.

They found that sharing gaze allows rapid communication of spatial information. Maurer et al. [25] developed GazeAssist, a shared gaze system for coordinating spatial information between a car driver and co-passenger. Akkil et al. [3] developed GazeTorch, a system that provides gaze awareness in collaborative physical tasks and found that it made collaboration easier by naturally conveying spatial information. Previous studies have shown that gaze sharing can be used for spatial referencing between remote collaborators. In contrast, Muller et al. [28] compared a gaze-sharing system to a conventional mouse-based pointer, while collaborating over a shared computer display. They found that shared gaze induced more ambiguity than a purely intentional mouse-based pointing for collaboration. They argue that gaze sharing systems need to be compared against other pointing modalities, to understand their cost and benefits in collaboration.

Unlike previous works in this area, our study is in the context of egocentric view and, as suggested by Muller et al. [28], we compare the accuracy with which gaze of a collaboration partner can be used to interpret pointing targets to other plausible spatial referencing techniques in the egocentric view.

Accuracy of pointing gestures in collaboration

Avellino et al. [4] studied accuracy of pointing gestures to support collaboration on a large wall-sized display. They compared how accurately a user can interpret the video feed of a remote user showing a shared object. They found that head orientation is more accurate than pointing, using hand or a combination of head and hand. Wong and Gutwin [37] compared the accuracy of hand pointing in real-world and collaborative virtual environments. Their participants observed another person pointing at different objects and were asked to identify the object being pointed at. They found that accuracy of pointing in the real-world is better than in a collaborative virtual environment.

Similar to the above-mentioned studies, our study compares the accuracy of interpreting pointing gestures for collaboration. We also focused on the actual act of pointing and did not study the role of other stages in the deictic referencing process (such as staging). However, unlike the other reported works, our study is in the context of egocentric view.

METHOD

The objective of the study was to understand how accurately viewers of egocentric video can interpret the shared object pointed at by the remote user. The experiment was hence conducted in two phases, similar to Avellino et al. [4]. In the first phase, we recorded videos of actors pointing at numbers attached to the walls of a room using the different pointing techniques. In the second phase, our participants viewed the videos and tried to identify the number being pointed at.

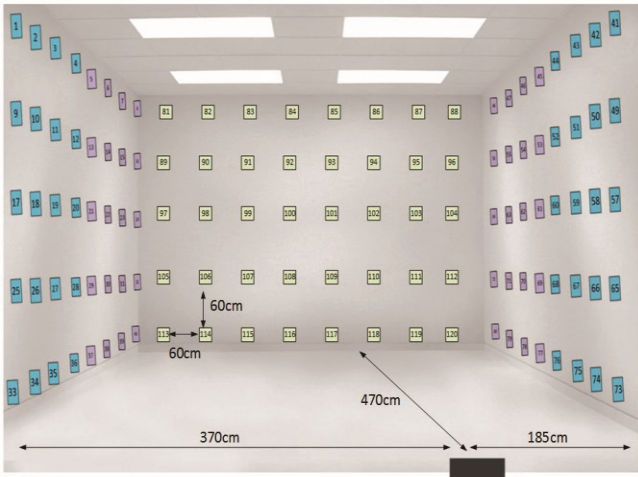


Figure 2. Dimension of the room and the indicative layout of pointing targets. The black rectangle shows the area where the actor was standing while performing the pointing. The colors in the figure indicate the different zones used in the analysis of the results (blue: high eccentricity-low density zone (Zone 1), purple: low-eccentricity-high density zone (Zone 2), yellow: central zone (Zone 3))

Phase 1: Video recording

We prepared a part of a room for recording the pointing tasks by attaching numbered papers on three walls (see Figure 2 and 3). The area used in the study was 4.7m x 5.55m x 2.5m. The numbers were large enough to be easily visible to the actor and also in the recorded egocentric videos. 40 numbers were attached on each of the three walls, 5 rows each with 8

numbers placed 60 cm from each other. Each row of number was also placed 60 cm away from the adjacent row. The choice of placing the pointing targets on all the three walls around the actor was to replicate a real-world scenario where the pointing will not always be confined to the area straight ahead. The distance between adjacent pointing targets was chosen so that it was not too large, based on expectation of pointing accuracy that may be required in some real-world tasks (e.g. pointing at a person standing in a group) and not too small, which would have made the interpretation very difficult for the participants.

We invited two actors (21 and 27 years old, 178 and 180 cm tall, both male) from the University community to perform the pointing task. Both the participants were right-handed, had normal vision and reported that they were not familiar with gaze trackers or head-mounted cameras.

Ocular dominance is known to influence hand pointing [21]. We used the “hole in the hand” test, a variation of Miles test [27] to determine the ocular dominance of the actors. It should be noted that ocular dominance is not a static concept and changes with gaze direction relative to the head position [20] (e.g. when looking at an eccentric target on one of the sides, the eye on that side may be preferred) or even which hand is used for the reach operation [8]. We use a rather simplistic definition of ocular dominance, i.e. preference of one eye over the other, when the target is straight ahead. The actors were instructed to sight an object directly in front of them, with both eyes open and head oriented straight, through the small gap formed between their stretched

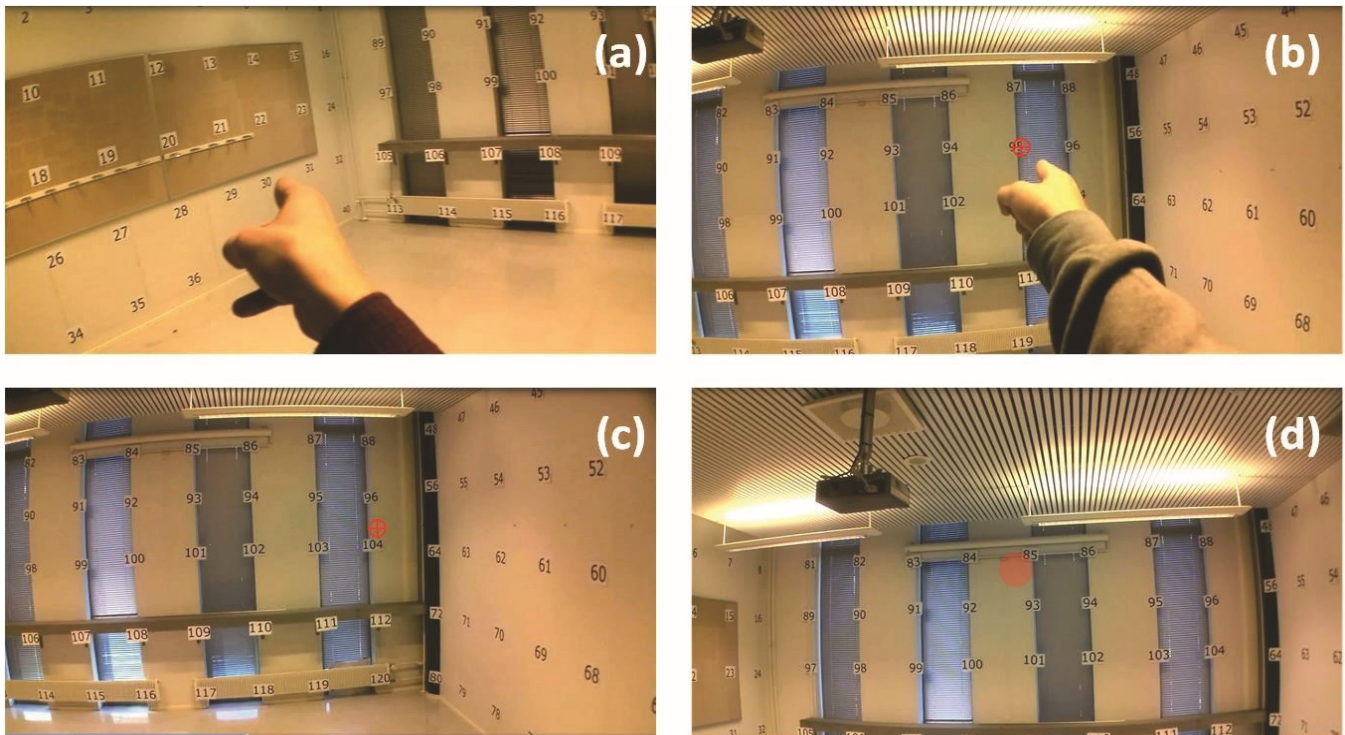


Figure 3. Screenshots from videos: (a) hand (b) hand+gaze (c) gaze (d) head. In gaze and hand+gaze conditions, the red crosshair indicate the gaze point. In head condition, the red circle indicates the center of the egocentric view.

overlapping hands at the eye level. The actors were instructed to alternately close each eye and report the closed eye for which the sighted object is no longer visible through the gap. Based on the test the actors self-reported their dominant eye to be right and left respectively. Miles noted about 34% of the adult population are left eye dominant and 64% are right eye dominant, while a tiny proportion of people do not have any clear eye dominance [27]. Further, about 33% of the population have a cross hand-eye dominance pattern (right-handed with left eye dominant or left-handed with right eye dominant) [27]. Based on these figures, it was reasonable to have actors with differing ocular dominance and hand-eye dominance patterns.

We created four different lists of thirty numbers each to be used as pointing targets for the actors. For each list, two different numbers were randomly chosen from each row on each wall (2 numbers x 5 rows x 3 walls = 30).

The actors first filled in a short background questionnaire and then were instructed to familiarize themselves with the number layout, by visually searching for a few numbers that the moderator read aloud. Later, they wore the Ergoneers Dikablis professional 60Hz binocular head-mounted gaze tracker. The gaze tracker had a scene camera of 90 degree field of view positioned between the eyes. The camera allowed vertical adjustment and the camera was positioned such that it allowed maximum visibility of the numbers along the vertical direction in a normal stance.

The participants were then asked to stand at a marked position, at two-thirds the length of the room (see figure 2). We then video-recorded the actor performing 30 pointing tasks for each of the four test conditions. The moderator read the number to be pointed aloud, the actor then pointed at the number and returned to their normal position. The number to be pointed was randomly selected from the list for that condition.

The four conditions in the video recording phase were:

- *Hand*: Actors were instructed to point like they would normally do extending their full hand.
- *Head*: Actors had to orient the head such that the number is in the middle of the perceived visual field of the camera.
- *Gaze*: Actors were instructed to simply look at the number they wanted to point at.
- *Hand+Gaze*: Actors were instructed to point like they would normally do extending their full hand. We also recorded their gaze direction.

Before each condition, the actors were given a round of practice. During practice, the moderator also gave feedback on the location of the pointing cursor when asked. This was especially useful in the *head* condition, where the moderator iterated the number that is in the middle of the camera field currently. This allowed the actors to fine tune their pointing

behavior. The video recording began after the actors felt confident to start.

After all the four conditions, the actors were instructed to look at one predefined number on each of the four quadrants, as seen through the scene camera. This data was used to ascertain the gaze data quality. Accuracy and precision of the gaze data was measured using a plug-in for TraQuMe [1]. The offset and dispersion was found to be below 3.8 and 0.39 degrees of visual angle respectively for all the quadrants for both the actors. The order of the conditions was reversed for the second actor.

Phase 2: Video viewing

Participants

We recruited 16 participants (14 male and 2 female) aged between 19 and 30 years from the University community. All participants had normal or corrected to normal vision. Six participants said they have used head-mounted cameras occasionally. Two participants were familiar with gaze tracking.

Design

We chose a within-subject design. There were four experimental conditions in the video-viewing phase:

Hand: The video of the actor pointing with the hand as seen through the head-mounted camera was shown (see Figure 3a).

Hand+Gaze: The egocentric video of the actor pointing with the hand, with an additional gaze overlay in the form of a red-colored crosshair was shown. The gaze overlay indicated the raw gaze data averaged for both the eyes (see Figure 3b).

Gaze: Participants were shown the video recorded where the actor was pointing with gaze. The raw gaze data averaged for both eyes was overlaid on the video in the form of a red-colored crosshair (see Figure 3c).

Head: The video of the actor pointing with the head was shown. The egocentric video was overlaid with a semi-transparent red circle of 80px diameter indicating the center of the video (see Figure 3d).

There were two dependent measures: number of errors in interpreting the pointing target and the median subjective confidence in the interpretation. We did not measure the time taken by the participants to interpret the target after watching the individual videos, as we anticipated no large difference based on the pilot tests. To test for differences between the conditions we used a non-parametric pair-wise (Monte Carlo) randomization test [11]. In randomization test, the null hypothesis is that there is no difference between the data under comparison. In other words, under the null hypothesis the pair-wise differences are equally likely to be positive or negative. Repeated resampling ($n=100,000$) with random assignment of sign for the difference between the conditions gives us a sampling distribution of the mean difference. The observed mean difference is then compared to the sampling

distribution to estimate how likely the observed difference is to occur by chance. If it is unlikely ($p < .05$), we reject the null hypothesis and conclude that the variable in question had an effect. If the likelihood of the observed mean is larger than threshold ($p > .05$), we fail to reject the null hypothesis. We further used Holm-modified Benferroni procedure [15] for family-wise type-1 error rate correction setting alpha at .05.

Apparatus

We used a custom C# software based on Microsoft .NET 4.5 framework to present the video stimuli and the on-screen questionnaire. The participants viewed the video in full screen mode on a 24 inch display with a resolution of 1920×1080 at 30fps. The participants answered the on-screen questionnaire using the computer mouse and keyboard.

Procedure

At the beginning of the experiment, the participants signed an informed consent form and filled in a short background questionnaire. The participants were then seated in front of the display. The task of the participants was to view the pointing videos and interpret the number being pointed at and indicate their subjective confidence in the answer.

For each condition, participants viewed 30 videos of the actor performing the pointing task. Each video was approximately 6 seconds long. The order of the videos was randomly selected. The steps in each video included the actor starting for the standing position with head oriented straight, locating the number to be pointed, performing the required pointing gestures and finally returning back to the normal stance.

After each video, an on-screen questionnaire was presented with two questions: “Which number was pointed at?” The participants had to enter the answer in the textbox. The second question was “how confident are you with the interpretation?” The participants had to select the answer on a scale from 1 to 7 (1 being not confident, 7 being very confident). After submitting the answer, the correct number being pointed at was displayed and the same video played again. Replaying of the video was important for the viewers so that they could learn the pointing behavior and tune their interpretations to match the actor’s style. One would expect such learning to take place in a real-world collaborative scenario.

After a pause of 3 seconds, the software showed the next video. The participants were offered a break after finishing two conditions. At the end of the test, participants filled a final questionnaire in which they ranked the pointing techniques based on the ease of interpreting. The order of the conditions and the actor for each condition were counterbalanced.

RESULTS

Aggregate accuracy of Interpretation

Figure 4 shows the aggregate accuracy for the four pointing conditions. *Gaze* condition received median accuracy of 0.98, comparable to *hand+gaze* (0.96), approximately 9% and 20% higher than *hand* (0.90) and *head* conditions (0.81) respectively.

Pairwise randomization test showed that *gaze* and *hand+gaze* conditions were statistically significantly more accurate than *hand* ($p = .001$ and $p = .02$ respectively) and *head* ($p = .001$ and $p = .01$ respectively) conditions. Further, *hand* condition was also found to be statistically significantly more accurate than the *head* condition ($p = .01$). *Gaze* and *hand+gaze* conditions were not statistically significantly different from each other ($p = 0.10$).

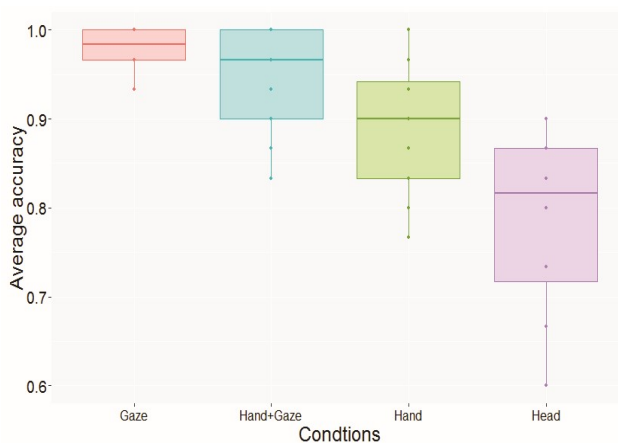


Figure 4. Aggregate accuracy of interpretation for the four conditions. Interpreting all the pointing targets correctly would result in a value of 1.

Figure 5 shows the boxplot for a mean value of the subjective confidence in interpretation of the pointing targets for the four conditions. It follows the same pattern as the results for the accuracy.

Confidence of interpretation

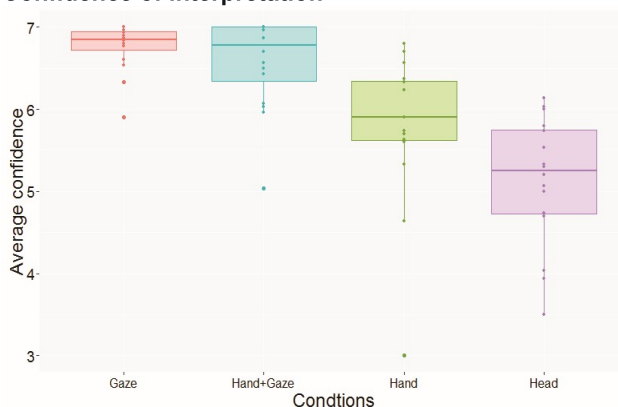


Figure 5. Subjective confidence of interpretation for the four pointing conditions (1-7 Likert scale).

Participants felt statistically significantly more confident in interpreting the targets in the *gaze* and *hand+gaze* conditions than *head* ($p=.001$ and $p=.01$ respectively) or *hand* ($p=.009$ and $p=.01$ respectively) conditions. Participants also felt statistically significantly more confident with *hand* condition than *head* condition ($p=.01$). The differences in subjective confidence for *gaze* and *hand+gaze* conditions were not statistically significant ($p=.09$).

Zone-based analysis

For further analysis, we divided the pointing targets into 3 zones, based on the target density and eccentricity of the target (see figure 2). The three zones were (a) high eccentricity-low density zone (Zone 1), (b) low eccentricity-high density zone (Zone 2) (c) central zone (Zone 3). The high density zone (Zone 2) had a lateral difference of less than 6 degrees between adjacent pointing targets, while zone 2 had a lateral difference of 7-15 degrees between adjacent pointing targets. In zone 3, pointing targets were 6.5-7.5 degrees apart from each other from the perspective of the actor and straight ahead. It should be noted that because our strategy for sampling the pointing target for the actor was to select two random numbers for each row per wall, all the zones did not have the equal number of targets. All three zones had between 8 and 12 pointing targets for each condition.

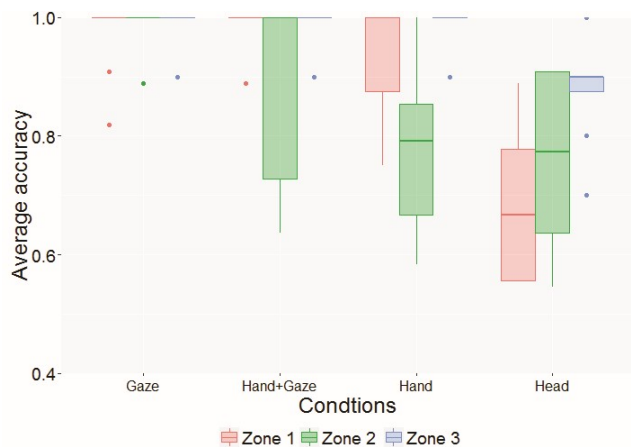


Figure 6. Accuracy of interpreting the pointing target based on the zone.

Figure 6 shows the accuracy of interpreting the pointing target for the different zones. *Gaze* condition performed equally well across all three zones. Randomization test showed that *hand* condition performed statistically significantly better when the target was in zone 3 ($p=.001$) or in zone 1 ($p=.001$) than in the high density zone 2. However, *head* condition performed best when the target was in the central zone (zone 3) than in zone 1 ($p=.001$). *Hand+gaze* condition performed better when the targets were in zone 3 ($p=.025$) or in zone 1 ($p=.025$) than in the high-density zone 2, these difference were however not statistically significant after Holm's correction. Similarly, there was no statistically significant difference between *hand* condition and *gaze* or

hand+gaze conditions when the target was in the zone 3 ($p>.05$).

Error Analysis

We analyzed the errors committed by the participants in interpreting the pointing target to understand if there were any specific underlying pattern. We segregated all instances where there was an error in interpretation, based on the relative direction of the selected number from the correct number being pointed at.

For *hand*, out of the total 54 errors committed, lateral errors (i.e. participants choosing the number to the immediate left or right of the actual number) accounted for 98% of all the errors while the remaining were diagonal errors (i.e. participants interpreted the target to be the number placed diagonally next to the actual number). For the *head* condition, out of the total 104 errors committed, lateral errors accounted for 30%, vertical errors (i.e. either choosing the number up or down from the correct number) accounted for 57% and diagonal errors accounted for 10% of the errors, the remaining 3% being errors resulting from unintentional pointing (e.g. the actor keeping the head still for a short duration, while searching for the pointing target, which the viewers interpret as the number being pointed at).



Figure 7. Distribution of lateral pointing errors for the *hand* condition for the two actors.

For *gaze* condition, out of the 12 errors committed, lateral errors accounted for 33% and the remaining 67% were errors committed due to unintentional pointing. Further, for *hand+gaze* condition, out of the 24 total errors, 87% were lateral errors and the remaining errors were along the diagonal or vertical direction.

We saw an interesting difference in the errors committed by the participants during hand pointing for the two actors. For the *hand* condition, the errors committed for each actor were 22 and 32 respectively. Figure 8 shows the number of right and left lateral errors in interpretation for the two actors. For actor 1, who self-reported to be right-eye dominant, participants often interpreted the number to the right side of the correct number as the pointing target. In contrast, for actor 2, who self-reported to be left-eye dominant, the participants' interpretation of the pointing target was more often skewed to the left (see Figure 7).

Subjective evaluation

All the participants ranked either *gaze* (10/16) or *hand+gaze* condition (6/16) as the easiest to interpret. *Head* condition was also consistently ranked as the most difficult to interpret the target (14/16). Participants also provided comments for their preference:

P12: *Gaze condition seldom left room for error* (P12 ranked *gaze* condition to be easiest)

P8: *When both hand and gaze were slightly off target, I could more easily deduce the correct number because I had more information* (P8 ranked *hand+gaze* condition to be easiest)

P13: *Hand pointing was pretty clear, but only after I noticed a pattern of the actor pointing a bit left of the number.*

DISCUSSION

Accurate spatial referencing is pivotal in video-based collaboration. Techniques to easily point at shared objects could potentially improve both the efficiency and subjective experience of users involved in collaborative tasks. Our study focused on comparing the accuracy of interpreting different pointing techniques in egocentric view. In general, *gaze* and *hand+gaze* conditions outperformed the *head* and *hand* conditions.

From the perspective of the user wearing the head-mounted camera, *gaze* is a natural pointing technique. People normally look at the objects in the environment while talking about them or even while pointing at them with the hand. From the perspective of the viewer, the *gaze* information is artificially overlaid in the egocentric view. In our study, the *gaze* and *hand+gaze* conditions not only led to more accurate interpretation of the pointing target, but also improved the subjective easiness and confidence in the interpretation. Accuracy of interpretation in *gaze* condition was not influenced by the density or eccentricity of pointing targets. Participants who rated *gaze* condition as the easiest to interpret liked it for its clarity. Our results support the use of *gaze* information overlaid in the egocentric view as a precise hands-free pointer to efficiently convey and coordinate spatial information in collaborative scenarios.

Participants who ranked *hand+gaze* to be easiest felt that the additional information provided by the hand conveyed when the act of pointing was taking place and helped confirm the target indicated by gaze. However, we could not see a clear benefit of using hand to point when the gaze information was already available in the video (*i.e.* *hand+gaze* was not significantly different from *gaze* condition). One should note that our study relied on simple pointing gestures (e.g. *look at that*). People may also use more complex hand gestures, along with pointing in collaborative scenarios for showing direction, “*from there you need to go this way*” (while indicating the direction using hand movements), or showing a general area, “*my home is somewhere in that region*” (while showing a closed area enclosing distant buildings using hand movement), or to express affordances of objects, “*you have to turn that knob like this*” (while making a clockwise turn

movement with the hand) etc. In such cases, it may be advantageous to have multiple modalities to effectively convey this information. Hand and gaze would be a natural combination e.g. pointing with the gaze while conveying other expressive detail, like affordance or direction, using simple hand movements.

Hand condition performed comparable to the two gaze conditions, except when the target was in the high density zone 2. There could be multiple reasons for this. Users normally point by placing the fingertip in the imaginary line between target and the dominant eye [14]. However, viewers of the video may use different strategies to interpret hand pointing [4]. For example, some of our participants said they relied on direction of the pointing finger or direction of the whole arm to find the correct pointing target. Such subtle differences in strategies may lead to possible misinterpretation of the target when the targets are closer together.

Further, our head-mounted camera was placed in between the eyes. When looking straight ahead, there is little difference in the camera view and the view seen by the person wearing the camera. Hence, pointing with the hand could have been easier to interpret for a remote viewer of this egocentric view, when the target was straight ahead. However, when pointing to eccentric targets, head movement is required prior to successfully performing the pointing. Normally, people do not move their head all the way to align the head with an eccentric target [6,36], leading to parallax between the tip of the finger as seen through the fixed egocentric view and view of the person wearing the camera. This was most likely not an issue in zone 1, because the distance between the adjacent targets, as seen through the egocentric view were larger and our participants could infer the correct target more easily than zone 2, where the targets would have appeared more close to each other. In a real-world scenario, one should expect that the pointing may not be confined to the area straight ahead. For example, consider a scenario where a user wearing a head-mounted camera walking down the street showing the landmarks around him, to a remote viewer of this egocentric view. Using hand-pointing in such scenarios may not be the most optimal technique due to the additional verbal utterances that may be needed to convey the target efficiently and the costs that would incur to repair the conversation, in case of a wrong interpretation.

Figure 9 shows a schematic of the parallax issue that might arise in hand pointing, when considering the ocular dominance of the actor. Normally, the tip of the pointing hand appears to the right side of the pointing target for right eye dominant person and tip of the pointing hand appears to the left of the pointing target when the actor is left eye dominant. Cross-dominance of hand and eye could also mean that the pointing hand may cover the object being pointed at. Our participants made more errors in the left direction for the left eye dominant actor and more errors

towards the right direction for right eye dominant actor. Hand pointing and its interpretation in the egocentric view could hence be influenced by ocular dominance and handedness.

Head condition had a median aggregate accuracy of 81%. However, participants showed large variability in interpreting head-pointing (mean accuracy of individuals varied from 0.6 to 0.9). Interpreting head-based pointing was most accurate when the target was straight ahead in zone 3. The accuracy was considerably lower in zones 1 and 2. Our results on the accuracy of head pointing suggests that it may be a plausible spatial pointing technique in egocentric view, depending on the accuracy requirements and eccentricity of the targets. One should also note that our participants felt least confident in the *head* condition and also rated it to be the most difficult to interpret. This suggests that there could be an additional cognitive load associated with its use in actual collaborative scenarios.

Limitations and future work

Our current study has limitations. Firstly, it is possible that wearers of head-mounted camera are not always aware of the extent of field of view of the camera [16]. This may have affected the ability of our actors to effectively center the pointing target in *head* condition. This was also evident in some hand-pointing videos, where the pointing hand was visible only at the edge of the camera field of view. One could expect that with prolonged use of the device, the wearer may develop a mental image of the extent of camera view [16]. Our actors were not frequent users of the head-

mounted camera devices and specifically the Ergoneers Dikablis device we used in our study. We provided the actors with a short training session for each condition, but it may be possible that their pointing could have been more accurate with longer training sessions. Alternately, an augmented reality display that shows the borders of the camera field or center of the camera view could have helped our actors in both *hand* and *head* conditions. It would have allowed actors to precisely orient the head to center the pointing target relative to the camera field in *head* condition, or to ensure that a larger portion of the hand is visible in the camera view while pointing with the hand. Further, it could be possible to correct the systematic displacement in hand pointing using computer vision and machine learning approaches [26] and present the hand-pointing target as a cross-hair in the egocentric view. More research is required to understand the effect of these factors on how effectively users can perform hand and head-based pointing and how accurately viewers can interpret them.

Second, our study was conducted in two phases and did not involve an actual collaborative scenario. One would expect that in a real-world collaborative scenario, the partners would repair the conversation if they detect a wrong interpretation of the spatial referencing, enabling both the partners to tune their pointing/interpretation strategy. In our study, after each pointing task, the correct target was shown to the participants and the same video played back. This enabled our participants to learn the pointing pattern and tune the interpretation strategy for the condition and the actor.

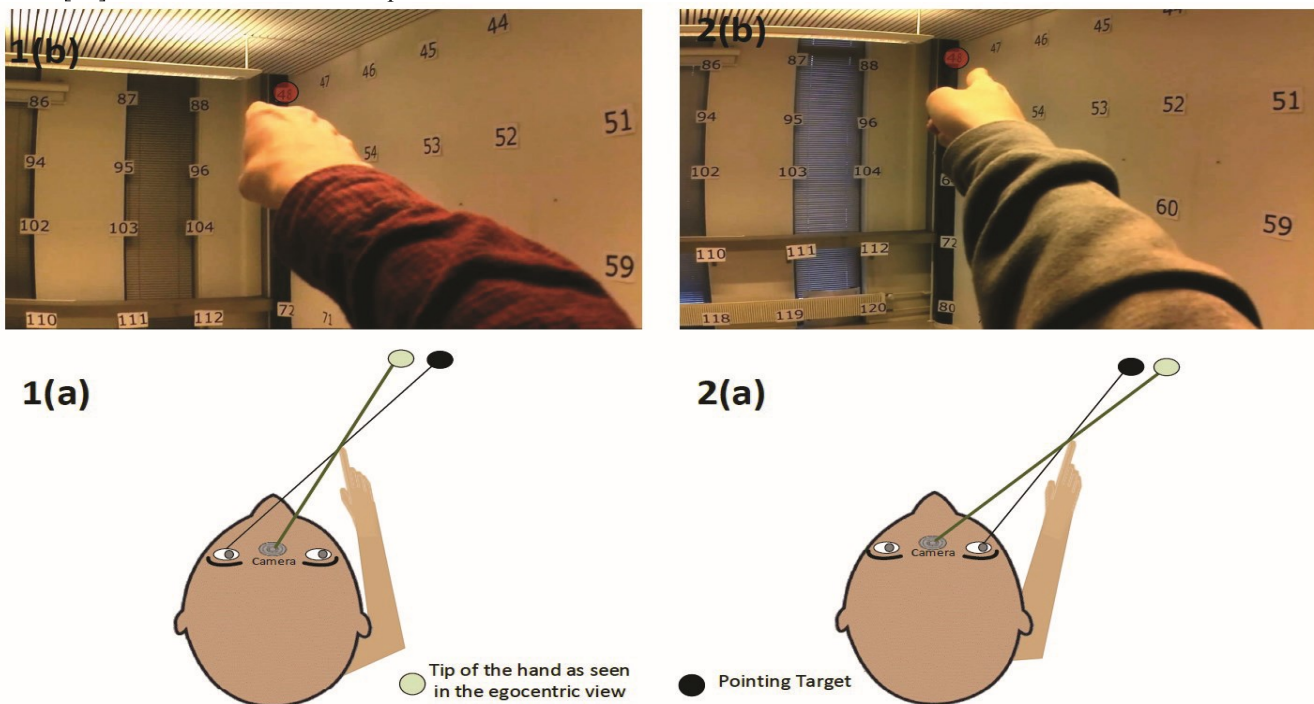


Figure 8. Hand pointing as a function of ocular dominance. For a left eye dominant person (figure 1(a) and 1(b)), the tip of the pointing hand tends to appear to the left of the pointing target. For a right eye dominant person (figure 2(a) and 2(b)), the tip of the pointing hand tends to appear to the right of the pointing target. The red circle in the figure 1(b) and 2(b) indicates the pointing target.

This benefited our participants and was evident in some of their comments (e.g. comment by P13) and could explain the fairly high interpretation accuracy in all the four conditions. However, in real communication situation the partner performing the pointing would also benefit from feedback on how the target was interpreted. This would allow the actors to tune their pointing strategy for easier interpretation. For example, Jones et al. [17] note that people sometimes point relative to the camera frame that the remote partner can view to enable easier interpretation while involved in collaboration using mobile devices. Also, the additional cognitive load and time pressure associated with collaboration may also influence the ability to interpret pointing targets. Additionally, the verbal utterances that normally accompany the pointing gesture might also reduce errors due to wrong interpretation of when the actor is pointing in *gaze* and *head* conditions. Future work should look into how these different collaborative processes influence the interpretation of the pointing direction in egocentric view.

There were only two actors in our study. Also, the actors were instructed how to point in each condition. For example, in the hand condition the actors were instructed to point extending their full arm. There may be subjective differences in how people naturally use hand pointing, such as pointing without extending the arms or simply using the direction of fingers. Further work is required to understand the differences in the use of pointing gestures in egocentric view and how accurately viewers can interpret them.

Gaze-tracking quality in our recorded videos was fairly good. Our video recording session took place in a controlled indoor environment and the tracker itself was mounted on the head of the actor in a stable way. In a real-world scenario, many factors could affect gaze data quality in head-mounted trackers. For example, ambient infrared lighting, vibrations in the environment, or frequent head movements leading to movement of the tracking sensor, may deteriorate the accuracy/precision of gaze tracking. One could anticipate that the *hand+gaze* condition may help in such cases, if viewers can effectively combine hand and gaze information to arrive at a decision (e.g. comment by P8). Further work is required to understand how gaze data quality affects interpretation of spatial pointing, for the gaze and *hand+gaze* conditions in egocentric view.

In our current study, we used the Ergoneers Dikablis gaze tracker, which has the head-mounted scene camera positioned between the eyes. Similar camera set-ups are also used in other smartglasses, like Tobii Glasses 2. The Pupill gaze tracker [19] uses a scene camera set-up just above the right eye. Many other commercial, head-mounted display devices, position the camera on the sides of the head, e.g. Google glass or Epson Moverio 2. Some action cameras (e.g. GoPro, Panasonic A1) provide flexible mounts (e.g. on a helmet or hat). In summary, many different head-mounted camera set-ups are available. Camera position is likely to

influence the accuracy of interpreting the different pointing techniques. Further work is required to understand and quantify this issue in real-world collaborative settings.

CONCLUSION

We investigated the accuracy with which viewers of egocentric video can interpret different pointing gestures. We conclude by presenting the implications of our study for designing collaborative systems that make use of egocentric view:

- Hand, gaze and head pointing are all plausible pointing techniques in collaborative use of egocentric view. Deciding the optimal pointing technique should be based on the task characteristics, accuracy requirements and eccentricity of the targets.
- Hands-free pointing will work in egocentric view without any loss of accuracy. Gaze information when overlaid in the egocentric view can be used as a precise pointer to refer to shared objects.
- For simple pointing tasks, using hand pointing in addition to gaze overlay do not lead to any significant accuracy improvement over gaze-only pointing.
- Interpreting hand pointing is most accurate when targets are straight ahead. For eccentric targets, subtle differences in interpretation strategy, head movements and ocular dominance of the actor can influence accuracy of interpretation.

Our results show the potential utility of overlaying gaze information in egocentric view in tasks that require accurate spatial referencing and encourages further exploration in real-world collaborative tasks.

ACKNOWLEDGMENTS

We thank the members of Tampere Unit for Computer-Human Interaction who provided helpful comments on different versions of this paper. The work was partly funded by Academy of Finland, projects Haptic Gaze Interaction (decisions 260026 and 260179) and Mind Picture Image (decision 266285).

REFERENCES

1. Deepak Akkil, Poika Isokoski, Jari Kangas, Jussi Rantala, and Roope Raisamo. 2014. TraQuMe: a tool for measuring the gaze tracking quality. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14), 327-330. <http://dx.doi.org/10.1145/2578153.2578192>
2. Deepak Akkil and Poika Isokoski. 2016. Gaze Augmentation in Egocentric Video Improves Awareness of Intention. In Proceedings of ACM Conference on Human Factors in Computing Systems (CHI '16). <http://dx.doi.org/10.1145/2858036.2858127>
3. Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. GazeTorch : Enabling Gaze

- Awareness in Collaborative Physical Tasks. In Proceedings of ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16). <http://dx.doi.org/10.1145/2851581.2892459>
4. Ignacio Avellino, Cédric Fleury, and Michel Beaudouin-Lafon. 2015. Accuracy of Deictic Gestures to Support Telepresence on Wall-sized Displays. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), 2393-2396. <http://dx.doi.org/10.1145/2702123.2702448>
 5. Matthias Baldauf, Peter Fröhlich, and Siegfried Hutter. 2010. KIBITZER: a wearable system for eye-gaze-based mobile urban exploration. In Proceedings of the 1st Augmented Human International Conference (AH '10). Article 9, 5 pages. <http://dx.doi.org/10.1145/1785455.1785464>
 6. B Biguer, C Prablanc, and M Jeannerod. 1984. The contribution of coordinated eye and head movements in hand pointing accuracy. *Experimental brain research*. 55, 3: 462–469. <http://doi.org/10.1007/BF00235277>
 7. Richard A. Bolt. 1980. “Put-that-there”: Voice and gesture at the graphics interface. In Proceedings of the 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH '80), 262-270. <http://dx.doi.org/10.1145/800250.807503>
 8. David P. Carey. 2001. Vision research: Losing sight of eye dominance. *Current Biology*. 1.20. 828-830. [http://doi.org/10.1016/S0960-9822\(01\)00496-1](http://doi.org/10.1016/S0960-9822(01)00496-1)
 9. Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, Nan-Wei Gong, Chris Schmandt, and Vivek K. Goyal. 2013. Mime: compact, low power 3D gesture sensing for interaction with head mounted displays. In Proceedings of ACM symposium on User interface software and technology (UIST '13). 227-236. <http://dx.doi.org/10.1145/2501988.2502042>
 10. Mikael Drugge, Marcus Nilsson, Roland Parviainen, and Peter Parnes. 2004. Experiences of using wearable computers for ambient telepresence and remote interaction. In Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence (ETP '04). 2-11. <http://dx.doi.org/10.1145/1026776.1026780>
 11. Pat Dugard. 2014. Randomization tests: A new gold standard? *Journal of Contextual Behavioral Science* 3, 1: 65–68. <http://doi.org/10.1016/j.jcbs.2013.10.001>
 12. Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03). 513-520. <http://dx.doi.org/10.1145/642611.642701>
 13. Zenzi M. Griffin and Kathryn Bock. 2000. What the eyes say about speaking. *Psychological science*. 11, 4. 274–279. <http://dx.doi.org/10.1111/1467-9280.00255>.
 14. Denise Henriques and John D. Crawford. 2002. Role of eye, head, and shoulder geometry in the planning of accurate arm movements. *Journal of neurophysiology* 87, 4: 1677–1685. <http://doi.org/10.1152/jn.00509.2001>
 15. Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 1: 65–70.
 16. Steven Johnson, Madeleine Gibson, and Bilge Mutlu. 2015. Handheld or Handsfree?: Remote Collaboration via Lightweight Head-Mounted Displays and Handheld Devices. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15). 1825-1836. <http://dx.doi.org/10.1145/2675133.2675176>
 17. Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). 957-966. <http://dx.doi.org/10.1145/2702123.2702345>
 18. Kasahara, Shunichi, and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In Proceedings of the 5th Augmented Human International Conference (AH' 14), p. 46. <http://dx.doi.org/10.1145/2582051.2582097>
 19. 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). 1151-1160. <http://dx.doi.org/10.1145/2638728.2641695>
 20. Aarlenne Z. Khan and John D. Crawford. 2001. Ocular dominance reverses as a function of horizontal gaze angle. *Vision Research* 41, 14: 1743–1748. [http://doi.org/10.1016/S0042-6989\(01\)00079-7](http://doi.org/10.1016/S0042-6989(01)00079-7)
 21. Aarlenne Z. Khan and John D. Crawford. 2003. Coordinating one hand with two eyes: Optimizing for field of view in a pointing task. *Vision Research* 43, 4: 409–417. [http://doi.org/10.1016/S0042-6989\(02\)00569-2](http://doi.org/10.1016/S0042-6989(02)00569-2)
 22. Barry Kollee, Sven Kratz, and Anthony Dunnigan. 2014. Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In Proceedings of the 2nd ACM symposium on Spatial user interaction (SUI '14). 40-49. <http://dx.doi.org/10.1145/2659766.2659781>
 23. Rob Kooper and Blair MacIntyre. 2003. Browsing the

- Real-World Wide Web: Maintaining Awareness of Virtual Information in an AR Information Space. *International Journal of Human-Computer Interaction* 16, 3: 425–446. http://doi.org/10.1207/S15327590IJHC1603_3
24. Päivi Majaranta and Kari-Jouko Rähä. 2002. Twenty years of eye typing: systems and design issues. In Proceedings of the 2002 symposium on Eye tracking research & applications (ETRA '02). 15-22. <http://dx.doi.org/10.1145/507072.507076>
 25. Bernhard Maurer, Sandra Trösterer, Magdalena Gärtner, Martin Wuchse, Axel Baumgartner, Alexander Meschtscherjakov, David Wilfinger, and Manfred Tscheligi. 2014. Shared Gaze in the Car: Towards a Better Driver-Passenger Collaboration. In Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14). 1-6. <http://dx.doi.org/10.1145/2667239.2667274>
 26. Sven Mayer, Katrin Wolf, Stefan Schneegass, and Niels Henze. 2015. Modeling Distant Pointing for Compensating Systematic Displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). 4165-4168. <http://dx.doi.org/10.1145/2702123.2702332>
 27. W R Miles. 1930. Ocular dominance in human adults. *Journal of General Psychology* 3: 412–430. <http://dx.doi.org/10.1080/00221309.1930.9918218>
 28. Romy Müller, Jens R. Helmert, Sebastian Pannasch, and Boris M. Velichkovsky. 2013. Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *The Quarterly Journal of Experimental Psychology* 66, 7: 1302–1316. <http://doi.org/10.1080/17470218.2012.737813>
 29. Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15). 113-116. <http://dx.doi.org/10.1145/2677199.2680549>
 30. Mark B Neider, Xin Chen, Christopher A. Dickinson, Susan E. Brennan, and Gregory J. Zelinsky. 2010. Coordinating spatial referencing using shared gaze. *Psychonomic bulletin & review* 17, 5: 718–24. <http://doi.org/10.3758/PBR.17.5.718>
 31. Jason Procyk, Carman Neustaedter, Carolyn Pang, Anthony Tang, and Tejinder K. Judge. 2014. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). 2163-2172. <http://dx.doi.org/10.1145/2556288.2557198>
 32. Sonja Rümelin, Chadly Marouane, and Andreas Butz. 2013. Free-hand pointing for identification and interaction with distant objects. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '13). 40-47. <http://dx.doi.org/10.1145/2516540.2516556>
 33. Rainer Stiefelhagen and Jie Zhu. 2002. Head orientation and gaze direction in meetings. In CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02). 858-859. <http://dx.doi.org/10.1145/506443.506634>
 34. Bruce Thomas, Ben Close, John Donoghue, John Squires, Phillip De Bondi, and Wayne Piekarski. 2002. First person indoor/outdoor augmented reality application: ARQuake. *Personal and Ubiquitous Computing*, 75–86. <http://doi.org/10.1007/s007790200007>
 35. Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). 3327-3336. <http://dx.doi.org/10.1145/2702123.2702214>
 36. Takuya Uemura, Yasuko Arai, and Chiga Shimazaki. 1980. Eye-head coordination during lateral gaze in normal subjects. *Acta Otolaryngologica* 90, 1-6: 191–198. <http://doi.org/10.3109/00016488009131715>
 37. Nelson Wong and Carl Gutwin. 2010. Where are you pointing?: the accuracy of deictic pointing in CVEs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10). 1029-1038. <http://dx.doi.org/10.1145/1753326.1753480>
 38. Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. In *Proceedings of the Human Factors in Computing Systems* (CHI '15). 2125-2134. <http://dx.doi.org/10.1145/2702123.2702305>



Paper 5

Deepak Akkil and Poika Isokoski. 2019. Comparison of Gaze and Mouse Pointers for Video-based Collaborative Physical Task. In *Interacting with Computers*. Article iwc026, 19 pages. DOI: [10.1093/iwc/iwy026](https://doi.org/10.1093/iwc/iwy026)

© Oxford University Press 2019, Reprinted with permission.

Comparison of Gaze and Mouse Pointers for Video-based Collaborative Physical Task

DEEPAK AKKIL* AND POIKA ISOKOSKI

*Tampere Unit for Human-Computer Interaction, University of Tampere, Kalevantie 4,
33100 Tampere, Finland*

**Corresponding author: Deepak.akkil@uta.fi*

Remote collaboration on physical tasks is an emerging use of video telephony. Recent work suggests that conveying gaze information measured using an eye tracker between collaboration partners could be beneficial in this context. However, studies that compare gaze to other pointing mechanisms, such as a mouse-controlled pointer, in video-based collaboration, have not been available. We conducted a controlled user study to compare the two remote gesturing mechanisms (*mouse, gaze*) to video only (*none*) in a situation where a remote expert saw video of the desktop of a worker where his/her mouse or gaze pointer was projected. We also investigated the effect of distraction of the remote expert on the collaborative process and whether the effect depends on the pointing device. Our result suggests that *mouse* and *gaze* pointers lead to faster task performance and improved perception of the collaboration, in comparison to having no pointer at all. The mouse outperformed the gaze when the task required conveying procedural instructions. In addition, using gaze for remote gesturing required increased verbal effort for communicating both referential and procedural messages.

RESEARCH HIGHLIGHTS

We present the first experimental study comparing the value of gaze and mouse cursors for remote gesturing in collaborative physical tasks. Our findings present the following practical implications:

- Task characteristics influence the usability of gaze cursor for collaborative physical tasks.
- Mouse outperforms gaze in tasks involving conveying procedural instructions.
- Shared gaze is better than having no pointer at all for collaborative physical tasks.
- Gaze-tracking accuracy is correlated to quantitative measures of collaboration, such as task completion times and verbal effort.

Keywords: video communication; collaborative physical task; gaze awareness; mixed-reality; distractions; conversation analysis; gaze-tracking accuracy

Editorial Board Member: Dr. Françoise Détienne

Received 17 January 2018; Revised 3 December 2018; Editorial Decision 14 December 2018; Accepted 2 January 2019

1. INTRODUCTION

Physical tasks that may require guidance from another remote person are common in our everyday lives. An important emerging use of video telephony is to facilitate remote collaborative physical tasks. The shared visual context provided by video telephony can enable collaborators to easily understand the current state of the task, efficiently communicate and achieve grounding in communication (Fussell *et al.*, 2000; Gergle *et al.*, 2013).

Imagine a young adult seeking guidance from the parent to cook a dish, an industrial field worker seeking advice from an indoor expert to troubleshoot a complex equipment, an elderly parent asking the help of their child in a distant city to operate the new microwave oven, or a customer seeking the assistance of the manufacturer support staff to assemble a new furniture set. All these scenarios involve a local worker who can directly act on the physical objects, seeking expert

guidance from a remote user situated in a home or office environment and potentially using a desktop or laptop computer for collaboration.

Previous research has shown that, in scenarios such as these, pen-based annotation systems that enable the remote expert to directly draw on the video, or systems that allow overlaying representation of hand of the expert on the video, can be beneficial. However, pen-based annotation systems require a touchscreen interface and representation of the hand of the expert requires additional sensors to track the hands, which are often not available on a standard desktop/laptop computing environment. Usefulness of simple pointers such as mouse, which is the most common and easily available pointing mechanism on a desktop/laptop computer, has shown mixed results (Fussell *et al.*, 2004; Li *et al.*, 2007).

In addition to the mouse, another feasible pointing mechanism in this collaboration context is shared gaze. Gaze-tracking devices are increasingly available in the consumer market. Currently, gaze-tracking sensors are integrated in some consumer products, such as laptops (e.g. Dell Alienware 17) and computer monitors (e.g. Acer Predator Z1), and are also available as an additional accessory to desktop and laptop computers (e.g. Tobii 4C), at prices that are much more affordable than a few years ago. The growing popularity of gaze-tracking devices makes it possible to track the visual attention of the remote expert and present this information to the collaboration partner.

Previous research suggests that sharing gaze information is useful in collaborative physical tasks, compared to having no pointer at all. For example, to help communicate spatial information and for improving the feeling of presence of the collaborators (Akkil *et al.*, 2016; Higuch *et al.*, 2016). However, gaze has never been compared to an alternative remote gesturing mechanism such as the computer mouse in the context of collaborative physical task. Individual studies that show that gaze can be helpful (Akkil *et al.*, 2016; Higuch *et al.*, 2016) and studies showing that a mouse can be helpful for spatial referencing have been made (Alem and Li, 2011; Fussell *et al.*, 2004; Yamazaki *et al.*, 1999). However, in situations where both pointing technologies are available, we need to know which one is better. We also need to know whether a gaze pointer can effectively replace the conventional mouse pointer when a mouse is not available to be used (e.g. when expert is using a device that does not support mouse, when hands of the expert are occupied in other tasks, limiting the usability of mouse, or in situations where a flat surface is unavailable).

Both gaze and mouse have commonalities and unique affordances. In terms of commonalities, gaze and mouse are both continuous pointing mechanisms that can be used for explicit spatial referencing (e.g. the expert could look explicitly at an object to point at it,—or use the mouse to explicitly point) (Akkil *et al.*, 2016). In contrast, gaze also implicitly conveys attention and cognitive processes, while mouse always requires

explicit user action. In remote collaboration, remote experts may have different strategies of gesturing using mouse. Some experts may use it more frequently, some may use it rarely and others may choose to not use mouse pointer at all, despite being available (Müller *et al.*, 2014). The automaticity provided by gaze ensures a level of consistency of what information is transferred between collaborators. Collaborators, however, may still have different strategies of how they use the gaze cursor (Müller *et al.*, 2014).

Another key difference between the shared gaze and mouse cursor emerges when the stationary collaborator is involved in a secondary task outside of the computer display. In such a scenario, the gaze cursor automatically disappears indicating that the partner is no longer attending to the screen, while the mouse cursor remains stationary. Schneider and Pea (2013) note that there is 'a certain degree of uncertainty associated with a [mouse] cursor that stops moving—is your partner thinking, being distracted, or waiting for you?'. On the other hand, the fine level of attention information provided by gaze may be beneficial in overcoming this ambiguity.

From the perspective of understanding the value of shared gaze for collaborative physical tasks, a pertinent question is whether the utility of shared gaze is limited to that of a 2D spatial pointer? Can gaze provide any additional benefit compared to an explicit spatial pointer like the mouse in the context of collaborative physical tasks? Does the presence of a secondary task for the remote expert differentially influence the utility of gaze and mouse cursors? A comparative evaluation of gaze and mouse pointers will provide theoretical understanding of the value of shared gaze and practical implications to the design of future video-based systems for collaborative physical tasks.

We developed a video collaboration system, which uses a stationary overhead camera and projection systems at the worker end. The overhead camera provides a consistent view of the task space to the remote expert, as noted in previous studies (Alem and Li, 2011; Kirk and Stanton Fraser, 2006) and the projection system enables us to show the pointer (gaze or mouse) used by the remote expert directly on the task space of the worker. With this system, we conducted an experiment involving real-time collaboration between a remote expert (from now on referred to as *expert*) and a local worker (from now on referred to as *worker*) in a block assembly task. To tease out the differences between the pointers, we designed an experiment with two independent variables: the pointer type used by the expert (*gaze, mouse, none*) and the level of distraction of the expert (*distraction, no distraction*).

The experiments focused on the following research questions:

RQ1: Is transferring either gaze or mouse pointer beneficial in terms of task completion time and subjective perception of the collaboration, in comparison to having no pointing mechanism in collaborative physical tasks?

RQ2: How does gaze pointer compare against the conventional mouse-based pointer in collaborative physical tasks?

RQ3: In the presence of a secondary task for the expert, is the benefit of the two pointing mechanisms (gaze and mouse) affected differently?

We begin by reviewing the related work on video-based collaborative physical tasks. Then, we describe the user study and its results. Finally, we discuss our key findings and implications.

2. RELATED WORK

We identified five relevant themes of related work. We first discuss the different gesturing mechanisms used in the context of collaborative physical tasks. In the second theme, we discuss the previous studies on gaze awareness in shared display collaboration. In the third and fourth themes, we discuss the previous studies on the value of gaze awareness in collocated collaborative physical tasks and computer-mediated remote collaborative physical tasks respectively. Finally, we describe past work relevant to multitasking and distractions in computer usage, to motivate our experimental design.

2.1. Gesturing in computer-mediated collaborative physical tasks

Collocated individuals collaborating to perform a physical task use two different types of gestures to support their communication: pointing gestures and representational gestures (Fussell *et al.*, 2004). Pointing gestures are used to refer to objects and locations (e.g. 'put that object there', while using hands to point at the object and the location). On the other hand, representational gestures are used to communicate the form of an object and the nature of action to be performed on it (e.g. 'turn the knob like this', while using hands to indicate the direction). One of the reasons why collaboration is more efficient when the involved parties are collocated compared to video-based remote collaboration is because of the effortless use and interpretation of gesturing (Fussell *et al.*, 2004).

Many previous studies have tried to support gesturing mechanisms in remote video-based collaboration, either using simple surrogates such as cursor (e.g. Kurata *et al.*, 2004; Yamazaki *et al.*, 1999), pen-based annotation systems (e.g. Fussell *et al.*, 2004; Kim *et al.* 2013; Kirk *et al.*, 2006), or directly presenting the hand movements of the partner (e.g. Alem and Li, 2011; Higuch *et al.*, 2016; Kirk *et al.*, 2006; Li *et al.*, 2007). Two common approaches have been either to augment the video used for collaboration with the additional gestural information and present this video on a computer display at the worker end (e.g. Fussell *et al.*, 2004; Kim *et al.*, 2013), or physically project the gesture representation directly on the task space of the worker (e.g. Akkil *et al.*, 2016; Alem and Li, 2011; Kirk *et al.*, 2006; Li *et al.*, 2007).

Fussell *et al.* (2004) compared mouse-based pointing with no pointing mechanism and found that simple mouse-based pointing mechanisms do not provide any significant performance benefits. They concluded that while mouse-based pointers are helpful to identify objects and locations, they do not help in giving procedural instructions, which take up the bulk of the time in collaborative physical tasks. Subsequent studies explored the use of pen-based annotation systems (Kirk *et al.*, 2006) and representation of the expert's hand (Alem and Li, 2011; Higuch *et al.*, 2016; Kirk *et al.*, 2006), to support collaborative physical tasks. Such systems were found to be helpful in providing complex procedural instructions and hence beneficial for the collaboration. However, Li *et al.* (2007) compared mouse pointer with hand representation and found no significant difference in task performance. In their study, the users preferred mouse pointing. They note that a mouse pointer is more familiar to the user in a computer-mediated environment and thus more preferred.

While the mouse is seen to be more appropriate to convey pointing gestures than representational gestures in collaborative physical tasks (Fussell *et al.*, 2004), previous work has also shown that mouse-based pointers allow a wide range of communication and representational expressions (e.g. drawing the shape of an object with the mouse pointer), if its movement is also transferred to the collaboration partner (Gutwin and Penner, 2002). Two major differences between the works by Fussell *et al.* (2004) and Li *et al.* 2007 were that, first, the former showed the pointer only upon a button click, while the latter study transferred the mouse location at all times. Second, the former presented the video with cursor overlay on a screen to the worker, while the latter used projection systems to directly project the cursor onto the task space.

The present study revisits the value of pointer-based remote gesturing mechanisms such as mouse and gaze for collaborative physical tasks. Our work used a stationary camera arrangement and projection systems that allowed projecting the pointer directly onto the task space of the worker, a setup similar to previous studies (Akkil *et al.*, 2016; Alem and Li, 2011; Kirk *et al.*, 2006; Li *et al.*, 2007). The objective of our work was to understand the value of mouse and gaze pointers in collaborative physical tasks. Similar to Li *et al.* 2007, we conveyed the mouse movement continuously, instead of only on button press.

2.2. Gaze awareness in shared display collaboration

Sharing gaze information between collaboration partners has been studied in scenarios where two remote partners interact using a shared display. Brennan *et al.* (2008) found that two-way gaze awareness during a collaborative visual search task led to shortest search times when only gaze information was shared and that sharing gaze plus voice and voice only were slower. Qvarfordt *et al.* (2005) studied one-way gaze sharing between a user and a remote tourist guide in a collaborative

trip planning task. They found multiple benefits of sharing gaze information, such as that gaze cues help spatial referencing, aid topic switching, reduce ambiguity in communication and help establish grounding in communication. Stein and Brennan (2004) studied seeing another person's gaze in a software debugging task and found that gaze information from an expert, even if not produced explicitly to communicate information, could provide important cues to viewers to solve similar tasks. D'Angelo and Gergle (2016) found that when a task is linguistically complex, shared gaze allows for more efficient communication, by enabling use of deictic references, than with no pointer at all. All the abovementioned studies compared gaze pointer to a baseline of no pointer for remote collaboration.

Velichkovsky (1995) was the first to compare gaze and mouse in a collaborative puzzle solving task. They found no significant difference between the two pointers. However, they argue that gaze may be particularly better suited than mouse in low redundancy and high complexity tasks. Müller *et al.* (2013) extended the work by Velichkovsky (1995) and reported similar results, in terms of task completion times. However, they found that unlike a purely intentional pointer like the mouse, gaze increased ambiguity in communication. Müller *et al.* (2013) argue that gaze-sharing systems need to be compared against other plausible pointing modalities, to understand their cost and benefits in collaboration. Bard *et al.* (2014) studied the effect of sharing gaze and mouse on language production, during a collaborative tangram construction task. Both the collaborators knew the target tangram to create and could individually manipulate the blocks. However, two objects joined together, only if they were held by different users. They found that only mouse cross-projection, and not gaze cross-projection, increased the use of deictic communication. Table 1 shows a summary of the previous work involving shared gaze in remote collaboration and the low-level characteristics of the experimental task employed.

Unlike typical computing tasks, our study focused on collaborative physical tasks, which is markedly different, in

terms of the task characteristics and collaborator symmetry. The kind of interactions we have with physical objects in the 3D world is more varied and complex than the interactions we have with graphical elements on a 2D computer screen, using a computer mouse. Velichkovsky (1995), Müller *et al.* (2013) and D'Angelo and Gergle (2016) used simple puzzle solving tasks that required participants to point at specific objects to identify them, without the need for any complex manipulations. Bard *et al.* (2014) used a task that also involved 2D manipulation of turning the puzzle blocks. However, both the collaborators could individually orientate the blocks and, thus, the experimental task did not involve communication of complex orientation information. Everyday computing tasks are pointing intensive, while everyday physical tasks involve complex physical manipulations. It is not clear how the cost and benefits of gaze awareness identified in the previous studies will apply in the case of a remote collaborative physical task, which involves conveying extensive amounts of procedural instructions. Are there limits to the utility of shared gaze based on the whether the task requires communicating referential or procedural instructions?

2.3. Gaze awareness in naturalistic collaborative physical tasks

It is well recognised that gaze plays an important role in our face-to-face conversations. However, situations where two or more collocated individuals work together on a physical task that requires referring and manipulating objects in the environment are different from typical conversational situation. Collocated collaborators use a variety of non-verbal cues such as pointing with hand, nodding, shaking the head and also eye gaze to communicate (Clark and Krych, 2004). Macdonald and Tatler (2017) found that in collaborative physical tasks, the specificity of the language used in collaboration and the utility of gaze of the partner interact with each other. Specificity of language influences the utilisation of

TABLE 1. Comparison of previous studies in terms of the task characteristics and experimental conditions.

Publications	Domain of investigation	Low-level task characteristics	Experimental conditions
Brennan <i>et al.</i> (2008) and Neider <i>et al.</i> (2010)	Visual search and consensus	Identify if a target is present among distractors	Shared gaze and none
Qvarfordt <i>et al.</i> (2005)	Trip planning	Seek information about on-screen objects	Shared gaze and none
Müller <i>et al.</i> (2013) and Velichkovsky (1995)	Collaborative puzzle solving	Identify and move 2D blocks to solve on-screen puzzle	Shared gaze, shared mouse and none
D'Angelo and Gergle (2016)	Collaborative puzzle solving	Identify and move 2D blocks to solve on-screen puzzle	Collocated, shared gaze and none
Bard <i>et al.</i> (2014)	Collaborative puzzle solving	Join 2D blocks to solve on-screen puzzle	Shared gaze, shared mouse, none and both

gaze cues and the usefulness of gaze in turn influences the specificity of language used. More recently, *Garcia et al. (2017)* conducted an empirical study involving a collaborative object assembly task. They manipulated the availability of gaze cue of the partner as an independent variable. In one of the experimental conditions, both the collaborators wore goggles that prevented the visibility of eyes to an onlooker and, in the other, the participants collaborated without the goggles. The results suggest that the availability of gaze cues not only enabled higher joint task performance but also led to higher frequency of deictic references and reduced frequency of conversational repair.

Previous studies in collocated collaborative physical tasks suggest that the information communicated by the eye movements can be beneficial. However, it is unclear how the utility of shared gaze transfers in a linguistically complex computer-mediated collaborative physical tasks. Can shared gaze lead to increased task performance and reduce the need for elaborate verbal instruction?

2.4. Gaze awareness in computer-mediated collaborative physical tasks

There are some previous studies that have investigated the value of augmenting gaze information in video from a head-mounted camera for collaborative purposes. *Akkil and Isokoski* found that overlaying gaze data in egocentric video improves a viewer's ability to predict intention of the partner (*Akkil and Isokoski, 2016b*) and help more accurately interpret pointing targets than when pointing with hand or head (*Akkil and Isokoski, 2016a*). *Gupta et al. (2016)* found that conveying gaze information in the egocentric video from the worker to the expert improves collaboration performance in a stationary LEGO building task. In addition, some previous studies have also investigated the value of gaze awareness in collaborative physical tasks using stationary cameras and projection systems at the worker end. *Higuch et al. (2016)* found that, when possible to convey both gaze and hand representation on the task space of the worker, experts tend to use gaze for object identification and hand representation for object manipulation. In a preliminary study, *Akkil et al. (2016)* found that physical visualisation of gaze made collaboration easier, by making spatial referencing effortless and by improving the feeling of presence of the expert. However, *Akkil et al. (2016)* did not compare gaze pointer with other pointing mechanisms and only studied the collaboration from the perspective of the worker.

This paper builds on the previous studies on gaze awareness in collaborative physical tasks. Novelty of our work is that our study compares gaze with a mouse-based pointer. In scenarios where the remote expert is using a desktop computer, both gaze and mouse are feasible gesturing

mechanisms and it is unclear which one would be a superior mechanism for remote gesturing in the context collaborative physical task.

2.5. Multitasking and interruptions in computer usage

Interruptions are common in our everyday computer usage. These interruptions could be another activity on the same computer (e.g. an email notification on the computer screen), or external to the computer (e.g. an incoming call on the mobile device). Previous work has shown that interruptions from an ongoing task carry a cost in terms of time to re-orient the attention back to the current task. However, users employ different strategies to minimise the effect by working faster to compensate for the interrupted time (*Mark et al., 2008*). Such balancing strategies cause larger cognitive load, time pressure and increased stress and frustration, on the user (*Mark et al., 2008*). Also, it is known that users learn to manage interruptions better with repeated exposure, reducing its cost on the primary task (*Hess and Detweiler, 1994*).

Wong et al. (2007) studied sharing of an expert among multiple partners, in video-based collaborative physical tasks. In their study, they did not use any pointing mechanisms. They found the workers will benefit from being more aware of the expert's focus of attention when it is divided. It is unclear how interruptions would affect a collaborative task where mouse movements are also transmitted to the partner. Would a mouse cursor that stops moving confuse the remote worker? Or would a mouse cursor that stays on-screen help the expert orient their attention back to the collaborative task after an interruption? On the other hand, gaze has an advantage in such situations. When the interruptions are external, a gaze pointer implicitly conveys to the worker that the expert is not attending to the screen (i.e. when the expert looks away, the gaze pointer disappears). Would this attention awareness provide any substantial benefit in the collaborative process?

3. USER STUDY

We conducted a controlled user study involving real-time video-based collaboration between a worker and expert supervisor. In addition to two-way audio link, the expert saw the video feed from the task space of the worker and the worker saw the pointer projected on his physical workspace.

3.1. Experimental design

Our experimental setup followed a within-subjects design with two independent variables: pointing condition (the different pointing styles used by the expert) and distraction level of the expert (presence or absence of a secondary task).

The three different pointing conditions were

- *None*: The expert did not have any pointing mechanism but saw the video and used verbal instructions.
- *Mouse*: In addition to verbal instructions, the expert could point at objects using the computer mouse. At the worker's end, the current mouse cursor was shown as a spotlight on the task space.
- *Gaze*: In addition to verbal communication, the gaze of the expert was projected to the task space of the worker as a spotlight. The spotlight disappeared when the expert glanced away from the screen.

The two levels of distraction for the expert were:

- *No distraction*: The expert did not have any additional task and concentrated fully on the collaboration.
- *Distraction*: The expert had a secondary task, designed to simulate the frequent, but momentary distractions such as incoming notification on the smartphone. The secondary task was to solve simple questions that involved the addition of two small numbers (both numbers <20). An arithmetic task was chosen because we could easily arrange and control it. The task appeared periodically, every 15 seconds after the last one was answered, on an adjacent touchscreen display. The display was placed on the same side as the mouse. The expert was explicitly instructed that the arithmetic task is their priority and to be answered as soon as possible, failing which the moderator would remind the expert. As the secondary task was external to the display, the gaze pointer automatically disappeared from the task space when the expert attended the secondary task. In contrast, the mouse pointer was still visible, unless the expert manually moved the mouse cursor to the very edge of the screen area.

Overall, there were six different experimental conditions (three pointing techniques \times two distraction levels).

3.2. Apparatus and experimental setup

The expert and worker were seated in adjacent sound-proof rooms. Our experimental setup consisted of an overhead pocket projector (3M MPRO150) and a Logitech camera at the worker end. The projector and camera were attached close to each other on top of the task space, using a custom-made mount to avoid parallax. Further, both the devices were aligned so that the projected image and the camera field of view covered the same area. At the expert end, we used a Tobii T60, 60 Hz remote gaze tracker, which tracked the point of gaze of the expert.

For the video calling, we developed a custom video calling system, using JavaScript WebRTC API. To reduce network delay, all the data transfer took place through a dedicated high-speed wireless access point. Our system did not provide view of the face of the collaborators i.e. the expert saw only

the task space of the worker and the worker only saw the mouse/gaze pointer overlaid on the task space. View of the face of the collaborators is known to be less useful in instructional collaborative physical tasks (Graver *et al.*, 1993). The experimental software (developed using Microsoft.NET 4.5) collected the gaze or mouse locations from the expert computer and transferred them to the worker system, using a WebSocket connection. The gaze and mouse location were visualised as a white circle floating on a black background. This feed was input to the projector, producing a spotlight representation of the pointer in the task space (see Fig. 1). Visualising the raw gaze data was often jittery and hard to follow for the worker. Therefore, we smoothed the gaze data using a recursive filter, also used in previous studies (Akkil *et al.*, 2016; Qvarfordt *et al.*, 2005).

$$y_i = W * x_i + (1 - W) * y_{i-1} \quad (1)$$

Here x_i is the actual gaze position returned by the tracker, y_i is the smoothed gaze position, W is the weight for the current gaze position and y_{i-1} is the previous smoothed gaze position. The weight (W) for the current gaze position is directly related to the responsiveness of the gaze cursor and inversely related to its jitteriness. This additional smoothing was not applied to the mouse pointer.

The projector also visualised the task space seen by the expert, by projecting a border around the camera view (see Fig. 1a). The dimensions of the task space were 60 cm \times 80 cm, and the diameter of the pointer spotlight was 6 cm.

The secondary task in the distraction conditions was presented on a Microsoft Surface Pro three tablet screen, placed adjacent to the primary display at the expert end (see Fig. 1b). The custom software developed in Microsoft.NET 4.5 presented simple questions involving the addition of two numbers. When a question appeared on-screen, the display of the tablet turned red and an audio beep was played every 2 seconds on the headphone of the expert, until the question was answered. Only correct answers were accepted. In the case of a wrong answer, the expert had to try again.

We further used TraQuMe (Akkil *et al.*, 2014) to evaluate the effect of gaze-tracking accuracy on objective measures of collaboration such as task completion times and verbal effort required to complete the collaboration. Accuracy of a gaze-tracking system is defined as the closeness of the measured gaze point to the point that the tracked eye is looking at and is measured as the average distance between a known stimulus position and the gaze point returned by the tracker.

3.3. Task

The task for the participants was to build predefined structures using pentomino puzzle blocks. Pentominoes are plane geometry puzzle blocks comprising of 12 different blocks. For the

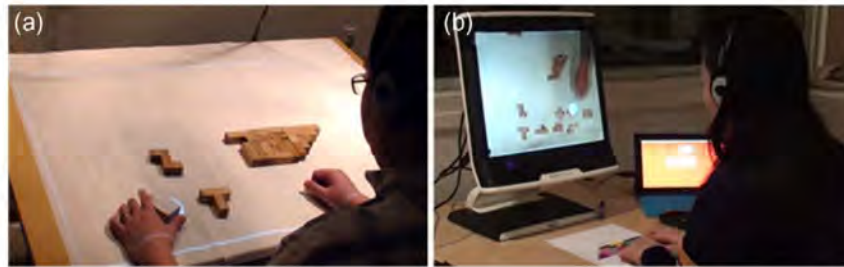


FIGURE 1. (a) The experiment setup at the worker end. Note the pointer and the border indicating the field of view of the overhead camera. (b) Experiment setup at the expert end comprised of Tobii T60 gaze-tracking device and a Microsoft Surface pro tablet positioned adjacent to the display and a pentomino pattern printed on paper. The expert is controlling the pointer using her eyes. The tablet shows the secondary task the expert must attend to.

experiment, we chose six structures of comparable complexity. Figure 2 shows the structures used in the experiment. The expert was given a printout of the structure to build but did not have access to the blocks. The worker had access to the pentomino blocks but did not have the printout of the structure. The schematics provided to the expert used colours to highlight different blocks (see Fig. 2). However, the physical pentomino blocks used at the worker end all had the same wood colour (see Fig. 1). The task for the pairs was to collaborate over the video link, to successfully build the structure. Even though the chosen task involved building 2D structures, it had different sub-tasks, such as identifying linguistically complex objects and performing 3D manipulations such as rotate and flip on an individual block to orient it properly that were representative of a real-world task. To efficiently collaborate on the building task hence required the use of both pointing gestures (to refer to the linguistically complex blocks) and representational gestures (to help correctly orient the blocks).

3.4. Pilot testing

We first conducted a series of pilot tests that involved three pairs of participants. Based on the pilot tests, the weight of the gaze smoothening recursive filter ($W = 0.30$), complexity of the secondary task (addition of two numbers, both numbers < 20) and the frequency of the beep reminder for the expert (every 2 seconds when the secondary task is active) were chosen. The chosen weight of the recursive filter was a good balance between responsiveness of the gaze pointer and the jitteriness. The complexity of the secondary task was chosen so that the task itself is not very complex and requires only 5–8 seconds to complete, on average (comparable to the time required to unlock a smartphone and quickly check the incoming notification).

3.5. Participants

We recruited 24 volunteer participants (12 pairs) from the University community. There were 16 female and eight male

participants, with ages between 23 and 42 years ($M = 26.4$, $SD = 5.3$). The participants were either allowed to sign up in pairs or individually, individuals then being paired by the experimenter. All participants had normal (11 participants) or corrected to normal vision (13 participants). A total of 15 participants were unfamiliar with gaze-tracking, while the remaining had some experience with it from previous experiments/courses. All the experts were frequent users of computer and experienced with using mouse. All participants were proficient English speakers.

3.6. Procedure

At the beginning of the experiment, participants were given an overview of the study, followed by signing of the informed consent form and completing the background questionnaire. Participants were then assigned to the role of either the expert or the worker. As a practice, the pair was asked to build a structure using the pentomino blocks, while standing facing each other over a desk. After the practice round, the expert and worker were seated in different rooms, the expert in front of the Tobii T60 gaze-tracking display and the worker near the task space. After ensuring that the audio/video communication worked as intended, the different experiment conditions were executed. For each condition, the expert was given a printed picture of the structure to build. Further, the worker was verbally instructed about the pointer that will be used (*gaze*, *mouse* and *none*). Once the task was completed successfully, the participants were asked to fill in a short questionnaire with seven different questions using 7-point Likert scale to evaluate the perceived quality of the collaboration (see Table 2 for the original formulation of the questions).

Before each of the gaze conditions, the gaze tracker was calibrated using the standard 5-point calibration procedure and, soon after the completion of the task, the gaze data quality was measured, using 5-point (centre and four corners) quality evaluation process using TraQuMe (Akkil *et al.*, 2014). TraQuMe shows predefined fixation points on-screen, similar to a gaze tracker calibration procedure and, based on

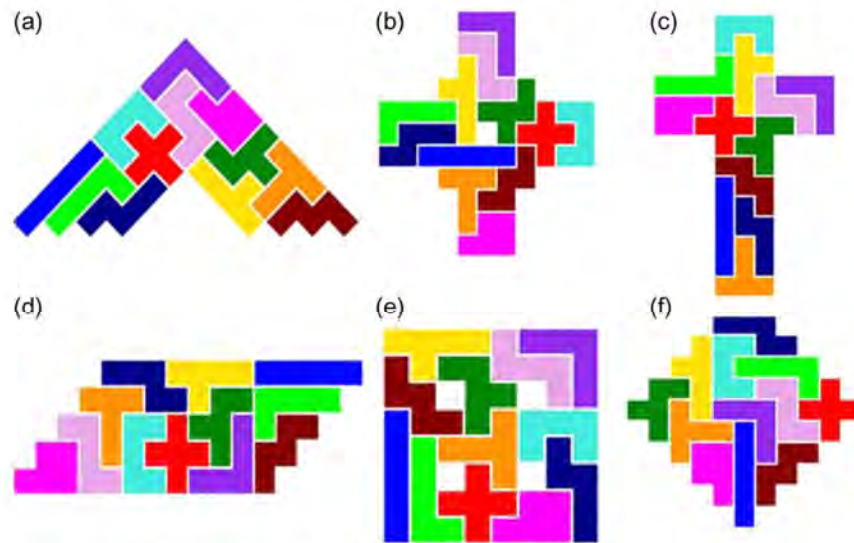


FIGURE 2. The six structures used in the experiment. Coloured representation of the structure was given to the expert on paper. The physical pentomino blocks had only one colour (see Figs 1 and 2).

the gaze data, evaluates the accuracy and precision of tracking. Before each condition, the worker was made aware of the current pointer control mechanism (i.e. gaze, mouse and none). Before the distraction conditions, participants were introduced to the additional task. The expert was asked to handle the distraction without letting the worker explicitly know of the secondary task.

There were altogether six trials (three pointing techniques \times two distraction levels). For each trial slot, the structure to build was fixed and the order of the conditions was counterbalanced between participants. This meant, the pentomino structure to build was not tied to a specific condition. The experiment was video-recorded for later analysis.

3.7. Data analysis

The video was first transcribed while preserving the timestamps of each utterance and task-related events, such as selection and placement of each block. From this transcription, the task completion times, average time to select (TTS) and average time to place (TTP) each block, were derived and independently analysed. Furthermore, we analysed the conversation between the expert and worker by counting the number of phrases required to complete the task, similar to previous works (Fussell *et al.*, 2000; Gupta *et al.*, 2016). A phrase was defined as a distinct verbal utterance. In addition, in order to better understand the effect of gesturing mechanism on the collaboration, we used a conversational coding system, also used in previous studies (Fussell *et al.*, 2000). Each verbal utterance relevant to the collaboration was

TABLE 2. Original formulation of statements used in the questionnaire to understand perceived quality of collaboration.

No.	Statement
1	My partner and I worked well together in this task
2	It was easy to collaborate using this interface
3	I was aware when my partner needed help (expert) My partner was aware when I needed help (worker)
4	It was easy provide instructions using this interface (expert) It was easy to understand instructions using this interface (worker)
5	It was easy to refer to objects
6	I felt I was present with my partner
7	I enjoyed using this interface

classified into one of the five categories, irrespective of whether they were questions, answers or statements:

- **Procedural:** Utterances relating to actions on the task objects (e.g. 'move the block to the right', 'turn it clockwise' and 'fit the block in the gap').
- **Task status:** Utterances describing the state of the task, task completion strategy, or state of objects within the task (e.g. 'we are building a cross shape', 'we will start from the left side to the right' and 'the object does not fit in the space').
- **Referential:** Utterances relevant for directly identifying the task object or its locations (e.g. 'Pick the T block' and 'the block goes here').
- **Internal State:** Utterances communicating the internal state of the collaborators such as their cognitive load,

knowledge and emotions (e.g. 'I think I made a mistake', 'wait a moment, let me find the piece' and 'I am confused now').

- *Acknowledgement*: Utterances providing feedback that a message was heard/understood (e.g. 'okay' and 'mmm').

All the statistical analysis was performed, using two-way repeated measures ANOVA. Post-hoc testing was performed using paired sample *t*-test was used when ANOVA showed significant main or interaction effects. For the 7-point Likert scale questionnaire data and conversational analysis that did not follow a normal distribution, we used a repeated measures Aligned Rank Transform (ART) ANOVA (Wobbrock *et al.*, 2011) and post-hoc testing was done using *t*-test on the ART score, if (ART) ANOVA showed a statistically significant main effect for the pointing conditions. For all analysis, we further used Holm-modified Bonferroni procedure (Holm, 1979) for family-wise type-I error rate correction, setting alpha at 0.05.

4. RESULTS

4.1. Task completion times

Figure 3 shows the boxplot for average task completion time in the different pointing conditions and distraction levels. The median of task completion time was shortest for the *mouse*, followed by *gaze* and longest for *none*. A repeated measure two-way ANOVA revealed a statistically significant main effect of the pointing method ($F(2,22) = 18.9, P < 0.01$). Follow-up *t*-tests showed that the tasks were completed faster with the *mouse* ($M = 221.9, SD = 40.53$) than with the *gaze* ($M = 289.71, SD = 54.85$), $t(11) = 5.3, P < 0.01$ and *none* ($M = 363.83, SD = 101.1$), $t(11) = 5.5, P < 0.01$. Similarly, the *gaze* condition was statistically significantly faster than *none*, $t(11) = 2.7, P = 0.02$. The main effect of the distraction was not statistically significant ($F(1,11) = 2.3, P = 0.16$), and no statistically significant interaction was found ($F(2,22) = 0.3, P = 0.71$).

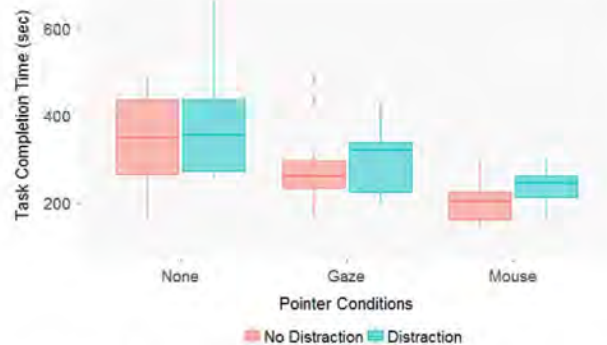


FIGURE 3. Task completion times for the different pointing conditions and distraction level.

4.2. TTS and TTP

We further deconstructed the block assembly task into two phases: *identification* and *placement*. For each of the 12 blocks, the expert had to instruct the worker on which block to select (identification) and on where and how to place the block (placement). From the perspective of the use of supporting gestures, the identification is pointing intensive (e.g. 'take this block'), while placement requires a combination of accurate pointing (e.g. 'place this side of the block here') and representational gestures (e.g. 'turn the block this way').

We computed the TTS and TTP separately for each block. TTS for a block was defined as the time taken to pick the correct block from the moment the previous block was correctly placed. Similarly, TTP was defined as the time taken to correctly place the block from the moment the correct block was selected. We did not include the last block in the TTS computation, as it was obvious that the last piece would be placed next. Also, the TTS and TTP calculations were only done for the *no distraction* condition as, in the presence of distraction, the expert may attend to the secondary task in either of the two phases, adding a confound variable.

Figure 4 shows the boxplot of the average TTS for the different pointing conditions. A paired samples *t*-test showed that *mouse* ($M = 5.1, SD = 1.2$) was faster than *none* ($M = 8.4, SD = 2.1$), $t(11) = 5.1, P < 0.01$. Similarly, *gaze* ($M = 5.75, SD = 1.1$) was faster than *none*, $t(11) = 4.8, P < 0.01$. The difference in TTS between *gaze* and *mouse* was not statistically significant $t(11) = 1.8, P = 0.10$. Figure 5 shows the boxplot of the average TTP for the different pointing conditions. A paired samples *t*-test showed that in the *mouse* ($M = 11.54, SD = 3.6$) condition, participants took statistically significantly less time to orient and place a block than in the *gaze* ($M = 16.8, SD = 6.8$), $t(11) = 3.4, P = 0.01$ or in the *none* condition ($M = 19.2, SD = 7.7$), $t(11) = 4.06, P < 0.01$. The difference in TTP a block between *gaze* and *none* was not statistically significant, $t(11) = 0.9, P = 0.39$.



FIGURE 4. TTS for the different pointing conditions.

4.3. Perceived quality of collaboration

The questionnaire data gathered from the expert and worker was analysed using repeated measures ART ANOVA (Wobbrock *et al.*, 2011). Figure 6 shows the boxplot for the seven 7-point Likert scale questionnaire data from the perspective of the expert and the worker. The statistical analysis relating to analysis of main effect and post-hoc comparisons is summarised in Fig. 7. No statistically significant interaction between pointing condition and distraction level was found for any of the questions.

Furthermore, perceived quality of collaboration varied based on the roles of the collaborators. From the perspective

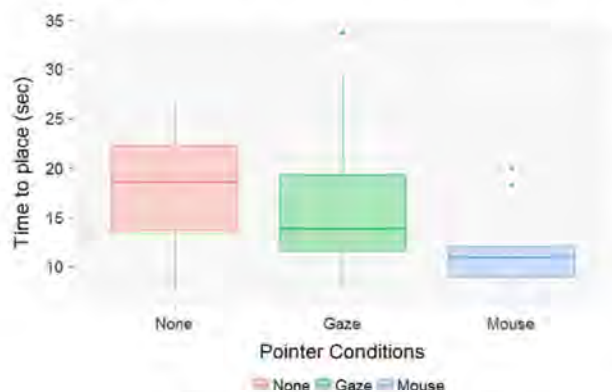


FIGURE 5. TTP for the different pointing conditions.

of the expert, a significant main effect of pointing condition was found for all the questions except Q3 (*aware when partner needed help*) and a significant main effect of distraction level was found for questions Q2 (*ease of collaborating*), Q3 (*aware of partner needing help*) and Q7 (*enjoyment*). On the other hand, for the worker, the main effect of pointing condition was only statistically significant for Q1 (*worked well together*) and Q2 (*ease of referring objects*). Furthermore, the main effect of distraction was statistically significant for Q1 (*worked well together*) and Q6 (*presence*) (see Fig. 7 for detailed *post hoc* analysis).

4.4. Conversation analysis: high level

In general, the expert led the conversation, speaking roughly four times more phrases than the worker. The expert used verbal cues to convey which object to select, how to orient it and where to place it in relation with other blocks, while the worker mainly used verbal cues to acknowledge the expert instructions (e.g. *mmm, okay*) or to clarify the instruction (e.g. *you mean this block?*). However, there were differences in how the expert and the worker communicated, based on the pointing condition and distraction levels. First, we used the total number of utterances spoken by the expert and worker to complete a task as a measure of the overall verbal effort, an approach also followed in previous works, e.g. (Fussell *et al.*, 2000; Gupta *et al.*, 2016; Müller *et al.*, 2013). The verbal effort of both the expert and worker would reduce when

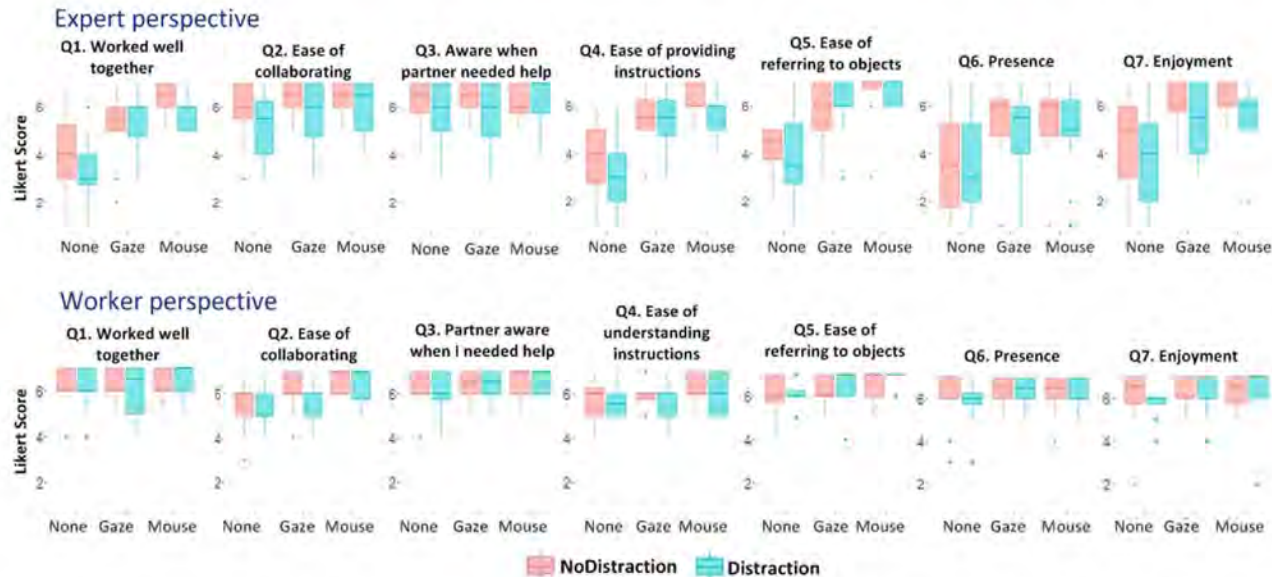


FIGURE 6. Perceived quality of collaboration based on the 7-point Likert scale questionnaire data from the perspective of the expert (top) and worker (bottom).

Question	Role	Main effect of pointing conditions		Main effect of distraction		Post-hoc analysis for pointing condition
		F(2,22)	p-value	F(1,11)	p-value	
1. Worked well together	Expert	20.8	<0.01	3.5	0.06	M > G (p = 0.03) M > N (p < 0.01) G > N (p < 0.01)
	Worker	11.6	<0.01	6.0	0.02	M > G (p = 0.01) M > N (p < 0.01)
2. Ease of Collaborating	Expert	3.6	0.04	5.2	0.03	
	Worker	0.5	0.59	0.4	0.55	
3. Aware of needing help	Expert	0.9	0.42	4.5	0.04	
	Worker	2.5	0.09	0.6	0.44	
4. Ease of providing/ understanding instructions	Expert	25.7	<0.01	3.1	0.09	M > N (p < 0.01) G > N (p < 0.01)
	Worker	2.1	0.14	1.5	0.22	
5. Ease of referring to objects	Expert	30.7	<0.01	0.9	0.34	M > N (p < 0.01) G > N (p < 0.01)
	Worker	5.1	<0.01	1.1	0.29	M > N (p = 0.01)
6. Presence	Expert	11.6	<0.01	0.1	0.88	M > N (p < 0.01) G > N (p < 0.01)
	Worker	0.7	0.50	5.6	0.02	
7. Enjoyment	Expert	12.0	<0.01	9.1	<0.01	M > N (p < 0.01) G > N (p = 0.01)
	Worker	4.0	0.02	2.2	0.14	M > N (p = 0.03)

FIGURE 7. Summary of the ART ANOVA statistical analysis. Colours indicate the role (grey: expert, green: worker). Significant main effects are highlighted in bold. *Post hoc* *t*-test between the conditions mouse (M), gaze (G) and none (N) are performed when a significant main effect of pointing condition is found. Only significant results from *post hoc* analysis are shown ($P < 0.05$).

the expert uses an unambiguous gesturing mechanism (Müller *et al.*, 2013).

Figure 8 shows the count of phrases for the expert for the different distraction levels and pointing conditions. A two-way repeated measures ART ANOVA revealed a statistically significant main effect of the pointing conditions ($F(2,22) = 33.6$, $P < 0.01$). Follow-up *t*-tests on the ART scores showed that, with the *mouse*, the expert used fewer verbal utterances ($M = 50.4$, $SD = 8.3$) than with the *gaze* ($M = 59.4$, $SD = 10.7$), $t(11) = 3.8$, $P = 0.01$ and in the *none* condition ($M = 76.3$, $SD = 10.2$), $t(11) = 6.7$, $P < 0.01$. Similarly, in the *gaze* condition, the experts used fewer utterances than in the *none* condition ($t(11) = 4.2$, $P < 0.01$). The main effect of the distraction level was not statistically significant ($F(1,11) = 0.0$, $P = 0.90$), and no statistically significant interaction was found ($F(2,22) = 0.2$, $P = 0.70$).

Figure 9 shows the count of phrases for the worker for the different distraction levels and pointing conditions. A two-way repeated measures ART ANOVA revealed a statistically significant main effect of the pointing technique ($F(2,22) = 11.2$, $P < 0.01$). Follow-up *t*-tests on the ART scores revealed that, in the *mouse* condition, the worker relied less on the spoken channel ($M = 10.7$, $SD = 7.0$) than in the *none* condition



FIGURE 8. Number of phrases used by the expert for the different pointing conditions and distraction levels.

($M = 21.8$, $SD = 13.6$), $t(11) = 3.8$, $P < 0.01$, and *gaze* condition ($M = 10.7$, $SD = 7.0$), $t(11) = 2.9$, $P = 0.03$. The difference between *none* and *gaze* was not statistically significant, $t(11) = 1.8$, $P = 0.18$. The main effect of the distraction level was approaching statistical significance ($F(1,11)$

= 3.9, $P = 0.05$), and no statistically significant interaction was found ($F(2,22) = 0.3, P = 0.70$).

4.5. Conversation analysis: low level

To gain a more in-depth understanding of how the different gesturing mechanisms influenced the collaboration, we analysed the conversation based on the purpose of each verbal utterance.

Acknowledgements, referential messages and procedural messages constituted the bulk of the conversation. So, we focused our statistical analysis for these message types (see Fig. 10). Three-way repeated measures ART ANOVA (three

pointing conditions \times two distraction level \times three message types) again revealed a main effect of pointing technique ($F(2,22) = 41.54, P < 0.01$) and no main effect of distraction levels ($F(1,11) = 1.0, P = 0.33$). Additionally, ART ANOVA showed a significant interaction effect between the pointing condition and message types ($F(8,88) = 14.2, P < 0.01$). Pairs used significantly less acknowledgements and procedural messages while using *mouse* than in *gaze* ($P = 0.02$ and $P = 0.02$ respectively). The difference between *gaze* and *mouse* was only approaching significance in terms of referential messages ($P = 0.08$). Similarly, pairs used significantly less acknowledgements, referential and procedural messages while using *mouse* than in *none* ($P < 0.01, P = 0.03$ and $P < 0.01$ respectively). The differences in message types between *gaze* and *none* were statistically significant for acknowledgements ($P = 0.01$) and procedural messages ($P = 0.03$) and, approaching significance for referential messages ($P = 0.05$).

The median value of count of different messages showed a common pattern of reduction. The amount of acknowledgements, referential messages and procedural messages were lowest for *mouse*, followed by *gaze* and *none*. However, in terms of the proportional change in the number of messages, the changes in procedural messages and acknowledgements were more than the change in referential messages. Using *mouse* and *gaze* for remote gesturing, compared to *none* resulted in 40 and 17% of reduction in acknowledgements, respectively, 49 and 34% of reduction in procedural instruction, respectively, and 28 and 16% of reduction in referential messages, respectively.

Figure 10 shows the different verbal utterance categories for the different conditions. We do not show the separation

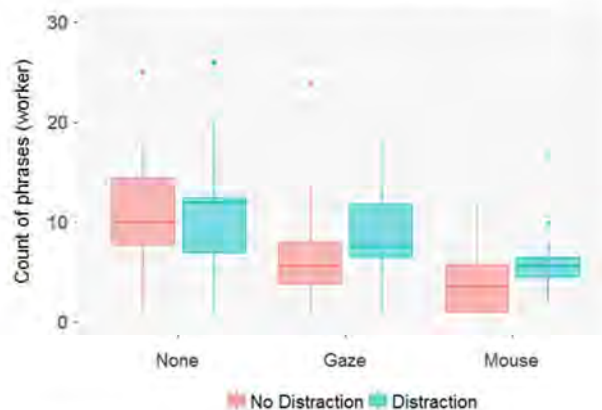


FIGURE 9. Number of phrases used by the worker for the different pointing conditions and distraction levels.

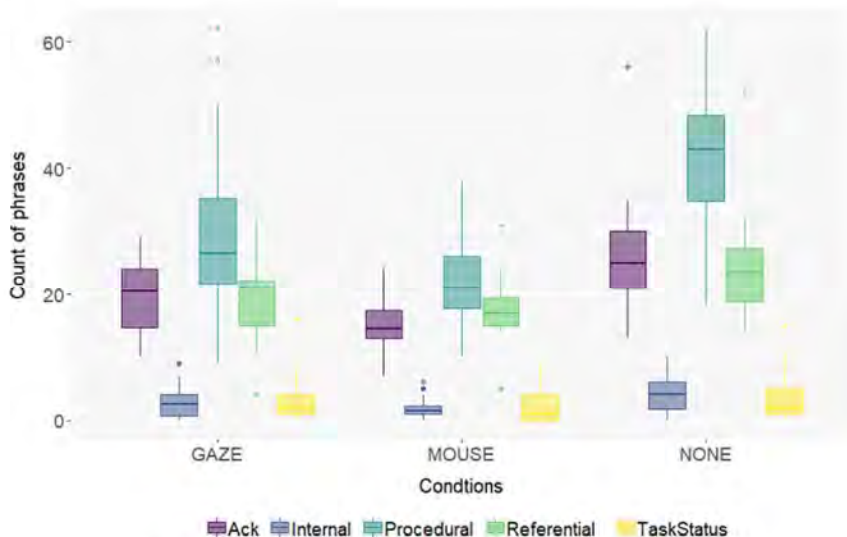


FIGURE 10. Message types for the different pointing conditions.

by distraction level in the figure for simplicity/readability, since the main effect of distraction level was not statistically significant.

4.6. Gaze-tracking quality and task performance

We analysed the relationship between accuracy of gaze tracking, measured using TraQuMe (Akkil *et al.*, 2014), and the overall task completion times. We computed the average accuracy of tracking for each participant, across the five gaze data validation points, for the two gaze conditions (with and without distraction for the expert). The average offset in the gaze data for the participants varied between 0.24 cm or 0.19° and 3.28 cm or 2.62° (mean = 1.07 cm, SD = 0.76 cm) on the expert's display. The offset was proportionally amplified, when gaze was projected onto the workspace that was much larger than the display. Increase in average gaze-tracking offset was weakly correlated with an increase in task completion times ($r = 0.63$, $n = 12$, $P = 0.027$). Similarly, an increase in average gaze-tracking offset was also weakly correlated with increase in the total number of phrases required to complete the task (sum of all phrases spoken by the pair) ($r = 0.70$, $n = 12$, $P = 0.024$).

4.7. Overall preference and subjective evaluations

Overall, eight and 10 (out of 12) experts rated the *mouse* pointing condition as the best in the presence and absence of secondary task, respectively. The others rated the *gaze* as the preferred pointing condition. Similarly, nine (out of 12) workers preferred the *mouse* condition, two preferred the *gaze* and one felt that both *gaze* and *mouse* were equally good. No participants (expert or worker) preferred the *none* condition. Only four out of the 12 workers said they were aware that the expert was involved in a secondary task in some conditions (from the long pauses or movements heard through the microphone). The user preference was also evident in the free form comments:

Pair 1 (worker): *Pointing light (gaze and mouse conditions) helped me a lot. I felt closer to my partner when I was doing the tasks.*

Pair 10 (expert): *Explaining was easier with the mouse. But, with gaze, when I was not gazing at the screen, my partner would notice it and wait for my instruction.* (preferred gaze in the presence of secondary task).

Pair 12 (expert): *I had to think where I should look, which is normally not how we use our eyes. I had to concentrate to not look in the wrong spot* (preferred mouse for both distraction levels).

Pair 4 (expert): *With mouse, I could draw paths and shapes more easily than gaze* (preferred mouse for both distraction levels).

Pair 3 (expert): (gaze condition) *Afterwards, my eyes felt a bit tired. It felt like they had to do more work than normal.*

We also asked the experts if seeing their own gaze visualisation affected them. nine out of 12 participants found it useful, while those remaining felt they initially found it distracting and had to get used to it, especially in situations where the accuracy of tracking was bad, and the gaze visualisation did not exactly indicate the actual point of gaze.

5. DISCUSSION

The focus of our study was to understand the effect of different pointing techniques and the distraction level of the expert on video-based collaborative physical tasks. Our experiment included only one such physical task, and it is likely that different tasks show different benefits for different pointing methods and camera configurations. We will now discuss our findings in relation with our research questions.

5.1. RQ1. Is either gaze or mouse-based pointer more beneficial than having no pointer at all?

Our results suggest that both mouse and gaze pointer can be beneficial in a collaborative physical task, compared to having no pointer at all. The presence of a pointer, irrespective of whether it is controlled with a mouse or with gaze, improved the task completion time, as well as the perception of collaboration from the perspective of the expert on several measures, such as 'ease of providing instruction', 'enjoyment of the task' and 'feeling of presence of the partner'. The presence of a pointer also enabled the expert to provide instructions, with less reliance on the verbal communication channel, with an overall trend towards fewer referential and procedural messages to complete the task.

Previous studies by Fussell *et al.* (2003) compared mouse pointer with no pointer for collaborative physical tasks but did not find a performance improvement due to the mouse pointer. Unlike the task used by Fussell *et al.* (2003), which involved communicating extensive amount of procedural instructions, our task required conveying both pointing and referential instructions. Furthermore, our task was linguistically complex, as all the blocks had the same colour and were of comparable sizes. The shape was the only differentiating factor between the blocks and some shapes were difficult to describe verbally. Hence the benefit of a pointing mechanism that enabled easy referencing of objects may have been more evident in our task. Also, unlike in the work done by Fussell *et al.* (2003), the mouse pointer was available at all times and not only upon a mouse click. This may have enabled our experts to use the pointer also to provide procedural instructions. We observed this in both pointer conditions (e.g. *you should turn it this way*, while making a clock-wise circle with the mouse or looking to the right-hand side with gaze).

Another difference between the studies was the output location of the pointer. *Fussell et al. (2004)* overlaid the pointer information on video and presented this as an external display to the worker, while in our study, this information was projected directly onto the task space. *Kirk et al. (2006)* demonstrated that such differences in output location do not lead to difference in task completion time in a pen-based annotation system. However, the relevance of their finding can be questioned because they did not study continuous pointing methods like mouse or gaze pointers. When the pointer information is displayed on an external monitor, it requires the worker to switch attention from the task space to the monitor. Since the gaze and mouse pointers are always on, cues provided by them are more likely to be missed by the remote worker, when presented externally. Hence, there may be a more pronounced effect of the output location for mouse and gaze pointers.

5.2. RQ2. How does gaze pointer compare against the conventional mouse-based pointer?

Gaze and mouse performed equally well in the identification phase in terms of efficiency. The identification phase did not require accurate pointing, as the blocks were large enough and the worker often placed the blocks in grids to make the task easier. In contrast, the placement phase benefitted from the high accuracy, and versatility offered by the mouse (e.g. 'place the block like this', while drawing a Z shape of the block using mouse or 'turn the block this way' and making a circular movement with the mouse).

All 12 experts used the mouse pointer to convey orientation and placement information. In contrast, only 6 out of 12 used the gaze pointer to convey orientation and placement information (e.g. this 'block goes like this', while making a horizontal eye movement). A few participants also used gaze to draw shapes to instruct how a block needs to be arranged or oriented. However, this often led to limited success as the expert could not perform the accurate eye movements required to aid the verbal instruction. In addition to physiological limitations associated with eye movements (e.g. it may be impossible to perform a quick smooth circular eye movement to indicate a turn direction), another issue with using eye movements to indicate directions was that the experts often performed exaggerated eye movements. For example, to instruct horizontal placement, the expert performed precise pointer movement with the mouse but could not do the required movement precisely with eyes and the movements performed often moved beyond the screen area, causing the gaze pointer to disappear from the task space.

The low-level conversation analysis showed that using the gaze required more verbal effort, than using a mouse-based pointer. Overall, the larger reliance on the verbal channel while using gaze, compared to using the mouse, can be

attributed to a larger number of verbal acknowledgements, and additional verbal effort required to communicate referential and procedural instructions.

There were multiple scenarios where using the gaze required further verbal support to overcome ambiguity. First, when two blocks were close to each other and, due to tracking issues, the gaze pointer fell between the two blocks. In such situations, participants had to use additional verbal instruction to clarify (e.g. *not that one. Take the block next to the one you picked*), as opposed to crisp verbal instruction which accompanied accurate mouse pointing (e.g. *take this*). Second, when the expert was visually searching for a block, the gaze cursor frequently moved from one block to another. This was often accompanied by explicit verbal instruction from the expert, to prevent the worker from selecting the wrong block, or wrongly interpreting the gaze fixation as part of the search as an explicit pointing act by the expert (e.g. *wait a moment, let me find the block we need*). This is inherently similar to the Midas-Touch problem associated with the use of gaze as an input modality in human-computer interaction (*Jacob, 1991*). Third, in the gaze condition, the worker used additional verbal interaction in the grounding process, by confirming the block that was being pointed to by the expert (e.g. by asking *do you mean this one?*) and providing more acknowledgement to the expert instructions. Fourth, in the placement phase, experts relied more on verbal instruction using gaze, than while using mouse to communicate the procedural instructions.

Gaze of a person provides several implicit cues about different cognitive processes, such as attention, interests and intentions. However, in our study, the most useful gaze cues were produced when the gaze was used purposefully (pointing with eyes, looking in a direction to convey direction, etc.). We also observed rare scenarios when the worker could 'predict' the location a block needs to be placed in, based on the eye movement of the expert, without the need for any verbal commands (*Higuch et al., 2016*). Experts would sometimes look at the block placement location, while providing instructions on which block to select. A few participants could effectively use that information, to predict the location to place the block. In such situations, the experts sometimes seemed pleasantly surprised that the worker did not need any instructions. However, the ambiguity associated with eye movements meant that most participants waited for verbal confirmation before 'trusting' the gaze signal.

Common ground is critical to the success of remote collaboration (*Olson and Olson, 2000*). Acknowledgements of understanding an instruction is central to establishing and developing common ground. Verbal acknowledgements such as 'okay' and 'hmmm' provide the conversation partner a feedback loop to ascertain understanding and monitor comprehension. Another way of providing similar acknowledgement of understanding is by letting the actions 'speak' (i.e. by directly following the instruction to pick/place an object

without additional verbal acknowledgement). Gergle *et al.* (2004) note that when collaborators have a shared visual context, the workers are less likely to provide verbal acknowledgement of understanding an instruction when compared to collaborations without a shared visual context. Similarly, the remote experts are more likely to follow the worker's action with the next instruction, without providing an acknowledgement for the correctness of the action. Such a behaviour is in line with Clark's principle of least collaboration effort to reach the conversational goals (Clark *et al.*, 1986). Our results from the low-level conversation analysis suggest a similar collaborative model in remote gesturing for collaborative physical tasks. In the *gaze* condition, collaborators relied on extensive verbal acknowledgements to establish and develop common ground. On the other hand, in the *Mouse* condition, collaborators relied on actions as an alternative for verbal acknowledgement, there by communicating more efficiently. Table 3 shows excerpts of pairs selecting and placing a block in *gaze* and *mouse* conditions highlighting the difference in terms of providing acknowledgement. This highlights the ambiguity introduced in the collaboration, due to the shared gaze and the additional verbal effort by both the expert and worker to develop common ground.

Müller *et al.* (2013) compared gaze and mouse for shared display collaboration and found that, while gaze and mouse perform equally well in terms of task completion times, gaze induced more ambiguity and complicated grounding in conversation. Their task involved only pointing, to help identify the on-screen puzzle blocks and their new locations. In contrast, our study focused on physical tasks and involved 3D physical manipulation of the blocks. Our results extend the findings by Müller *et al.* (2013) to physical tasks. In the identification phase (which is pointing intensive), both gaze and mouse performed equally well. Furthermore, mouse outperforms gaze in terms of task completion time, when task requires conveying procedural instructions or use of representational gestures. Further, using gaze for remote gesturing requires increased reliance on the verbal channel compared to using the mouse.

The purposeful use of eyes for remote gesturing required the experts to make both unnatural eye movements (e.g. stare

at a block for a longer duration than normal) as well as intentionally suppress the natural eye movement (e.g. not to look at specific blocks or screen areas, to avoid chances of misinterpretation by the worker). Unnatural eye movement can lead to eye fatigue and some of the experts also explicitly commented that their eyes were strained after the gaze condition (see comment by pair 3 expert). In contrast, using the mouse for remote gesturing did not have any such effects.

5.3. RQ3. How do distractions affect the use and benefit of gaze and mouse pointing?

Generally, distraction did not significantly affect the task completion times for the different pointing conditions. However, distractions influenced the subjective perception of the collaboration. We noticed that the expert used different strategies while managing the secondary task and the mouse pointer in the event of the secondary task. Often, the expert attended to the secondary task after the end of the identification phase, once the worker had picked up an object, or at the end of the placement phase, after a specific block was placed. Depending on the phase, using the mouse to instruct created situations where the worker wrongly interpreted the current mouse location as the location to place the blocks. Also, if the cursor accidentally moved onto another block when the expert attended to the secondary task, the worker inadvertently believed that block had been indicated as the next block to select. Such situations required additional verbal effort from the expert to repair the collaboration, but such events were less frequent to affect any other objective measures we used. However, two of the experts noticed this issue and deliberately moved the mouse away to the sides, when attending to the secondary task. In terms of the objective measures we used, our study does not provide any evidence for differential degradation of the value of gaze and mouse for remote gesturing mechanisms in the presence of a secondary task.

In the gaze condition, the workers were aware of the attention of the expert. Based on the subjective feedback, many of the experts and workers appreciated this awareness. There was also a difference in how the workers used this awareness. P10

TABLE 3. Excerpts of pairs selecting and placing a block in gaze and mouse conditions, highlighting the difference in verbal communication in terms of providing acknowledgement.

Shared gaze	Shared mouse
Expert: So, we can begin with L. <points with gaze>	Expert: And then this one <points block with mouse>
Worker: okay.	<Worker picks the block>
Expert: Yeah, and put it to the top right corner.	Expert: It will come around this corner, like this.
Worker: mmm. Ok Top right corner.	<Worker places the block in the correct spot>
<worker places the block in the correct spot>	Expert: and then take this block
Expert: Just like that.	
Worker: okay	
Expert: then, the Z.	

(expert) noted that the collaboration partner would notice the attention shifts and wait for instructions when the expert was involved in a secondary task in the gaze condition, while we also observed that most other participants used this awareness to regain the attention of the expert by asking more questions.

Based on our observations, a combination of mouse and gaze may be useful in scenarios where interruption of the expert is frequent. The mouse cursor that disappears when the user is not attending to the screen may mitigate the problems we encountered with ambiguity associated with the mouse cursor, when the expert is involved in an additional task. In some scenarios we observed that the experts had to be reminded by the worker to use the mouse (e.g. 'you can use the mouse, you know?'). Similar observations are also made in previous studies, that sometimes collaborators do not use the shared mouse cursor even when available (Müller *et al.*, 2013). A combination of gaze and mouse may also be advantageous, to achieve a level of consistency in remote gesturing for collaboration. For example, a gaze cursor that appears when mouse is not being actively used for collaboration may help support the individual variability in the use of mouse for remote gesturing.

The task completion times and the total number of utterances required to complete the task were correlated with the accuracy of gaze tracking. In a situation when the accuracy of tracking was low, both expert and worker used more verbal communication to overcome the inaccuracy. Experts relied more on the verbal channel to compensate for the reduced accuracy and recover from possible errors, while the worker resorted to confirming the instruction before acting on it (e.g. 'did you mean this block'). D'Angelo and Gergle (2016) provided qualitative insights into how collaborators compensate the lack of gaze data accuracy with detailed verbal description. Our results extend the findings by D'Angelo and Gergle (2016), by showing a quantitative link between critical collaboration measures, such as task completion times, number of verbal utterances and the accuracy of gaze data. Accurate gaze tracking is critical in shared gaze collaborative applications and issues with accuracy can negatively impact on task performance. Sharing gaze pointer when the accuracy is below a critical limit for the task may be counter productive to the collaboration.

From a methodological perspective, this is the first study to objectively report and quantitatively analyse the role of gaze-tracking accuracy on collaboration quality. Despite the understanding that accuracy of gaze data can influence collaboration quality in shared gaze interfaces, previous studies have neither analysed nor reported the gaze-tracking accuracy during the experiments and its influence on the results. Müller *et al.* (2013) refers to the spatial accuracy reported by the gaze tracker manufacturer ($<0.5^\circ$), which is often measured in ideal situations on 'ideal' users, and can be very misleading. Another aspect which could influence the research findings is the criterion for recalibration of tracker or exclusion of data from

analysis on the grounds of 'bad data'. Most of the previous studies do not employ or report objective grounds for recalibrating the gaze tracker or excluding a participant's data. For example, D'Angelo and Gergle (2016) mentions that they recalibrated the participants when gaze data became inaccurate, for seven out of 18 (roughly 40%) of the pairs. Guo and Feng (2013) note that the participants were '*monitored throughout the study and recalibrated as necessary*'. D'Angelo and Begel (2017) relied on the participants to tell when accuracy had degraded enough to require recalibration. Relying on the subjective judgements of the participant or the moderator, without objectively stating the accuracy achieved, could introduce large variability in the gaze data quality across studies, and thus affect the validity, comparability and repeatability of research findings. We believe measuring, analysing and reporting gaze-tracking accuracy to be a small, yet significant steps towards methodological consistency in the field of shared gaze research.

Our setup used physical projection of the gaze point in the task space of the worker and the camera at the worker's side captured this physical projection of gaze point. This enabled the expert user to see their own gaze point on the video feed, get direct feedback on the accuracy of gaze tracking and be more aware of the situations when the tracker could not see the eyes of the expert. Previous work has noted that showing the user his or her own gaze pointer as feedback is not effective in overcoming tracking issues, as it leads to feedback loop that causes people to follow their own gaze cursor (D'Angelo and Gergle, 2016). However, our observation and subjective comments from the experts suggest that seeing their own gaze point was useful in multiple ways. It is possible that this is true more generally in collaboration systems and that users can learn to use the feedback channel. This can happen in different ways. First, users can learn to overcome the accuracy issues with the gaze tracker, by either looking a bit away from the actual target, so that the gaze cursor would fall on the intended block, or by using supplementary verbal instructions to aid the worker (e.g. *to the left of where the pointer is*). Second, the visualisation can help the experts to be aware of the gaze-tracking status and to reposition themselves, when the tracker cannot see the eyes and the gaze pointer disappears. Third, the visualisation can help the expert to be more aware of their own eye movements and proactively use verbal instructions when the eye movements are misleading (e.g. explicit verbal instruction 'wait, I am searching for the block' when the expert was searching for the block).

5.4. Limitations and future work

Our study has a few limitations. First, gaze was novel and hence also unfamiliar input modality for many of our participants. All our participants had extensive experience using the mouse and limited to no experience using gaze. It is possible that the novelty/unfamiliarity of gaze influenced the user

preferences. Few of the participants who acted as the expert explicitly commented that they preferred the mouse when they were required to perform the secondary task, because using the mouse was familiar to them and they did not want to use a new technique under the additional cognitive load. It could be that the collaboration partners learn to use their gaze more effectively with extensive practice. Future longitudinal studies are required to understand if expert users learn to use the gaze modality more effectively and if the worker learns to interpret the subtle and implicit cues available from the eye movement in collaborative physical tasks.

Second, our setup involved stationary cameras and projection systems. This was the simplest setup we found for clearly understanding the differences in the pointing conditions. In addition to stationary cameras, a growing trend is to use mobile phones and head-mounted cameras/displays for collaborative physical tasks. Further research is required to understand how to extend our results to other camera-display configurations and how the inherent mobility involved in such setups would affect the different pointing conditions.

Third, our study focused on a specific kind of collaborative physical task, involving remote guidance in an object assembly task. It is possible that different task may show different costs and benefits of shared gaze awareness. For example, a collaborative learning activity or visual search in the physical world may benefit more from precise gaze awareness. Further research is required to understand how our results may apply to other collaborative tasks.

Fourth, we only compared shared gaze with a shared mouse pointer. There exist other rich ways of communicating referential and procedural instructions (e.g. presenting hand of the collaborator or touchscreen annotation systems). Future research should study how gaze compares against other techniques and beneficial ways to combine the different channels to enable an effective and efficient remote collaboration for physical tasks.

Fourth, the secondary task we designed required only momentary attention shift and occurred every 15 seconds. It could be expected that if interruptions due to the secondary task are more frequent, or require more time to complete, it may have a stronger impact on the task completion times and subjective perception of the collaboration of both the expert and worker. Also, our workers were not explicitly made aware of the secondary task for the expert user. Further work is required to understand if knowing that the collaboration partner may be involved in another task affects how the worker reacts to precise attention awareness available in the gaze condition.

6. CONCLUSION

An emerging use of video telephony is to enable remote guidance in physical tasks such as, operating, troubleshooting and repairing unfamiliar equipment, assembling a new purchased

device or even everyday scenarios such as cooking a dish. We investigated the value of gaze and mouse-based remote gesturing mechanism in such collaborative physical tasks.

We conclude by presenting the practical implications of our study:

- The presence of any pointer (*gaze or mouse*) improves efficiency and the perceived quality of collaboration, compared to having no pointer at all.
- Mouse pointer, being more accurate and versatile, can be useful in conveying both pointing and representational gestures. Mouse outperforms gaze when the task requires conveying procedural instructions.
- Gaze can be a feasible pointing modality, when the task characteristics require more pointing gestures than representational gestures. In some cases, using mouse may not be feasible. For example, when the expert is using a device that does not have a mouse (e.g. a virtual reality headset), or an expert is involved in another task that requires use of hands. It is important to evaluate the task characteristics and context of use, before choosing the optimal pointing modality.
- When using shared gaze, collaborators rely more on the verbal channel to establish and develop common ground and communicate complex instructions.
- Gaze-tracking accuracy affects the task performance, when gaze is used as a pointer. An increase in the gaze-tracking offset is associated with an increase in the task completion times, as well as with an increase in the total number of phrases required to complete the task.
- Use of mouse to point while the expert is also performing a secondary task can cause confusion and sometimes lead to wrong interpretation of the pointer by the worker. Similarly, there may be inconsistencies in how individuals choose to use the mouse for remote gesturing. A combination of mouse and gaze (e.g. mouse pointer visible only when the expert attending to the task, or a gaze pointer that appears when collaborator is not using the mouse actively) may be a useful combination in such scenarios.

SUPPLEMENTARY MATERIAL

Supplementary data is available at *Interacting with Computers* online.

ACKNOWLEDGEMENTS

We would like to thank members of Tampere Unit for Human-Computer Interaction research unit for their help and feedback in designing the experiment. This work has been partly funded by the Graduate school funding, School of Communication Sciences, University of Tampere.

REFERENCES

- Akkil, D. and Isokoski, P. (2016a) Accuracy of Interpreting Pointing Gestures in Egocentric View. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM. pp. 262–273.
- Akkil, D. and Isokoski, P. (2016b) Gaze Augmentation in Egocentric Video Improves Awareness of Intention. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM. pp. 1573–1584.
- Akkil, D., Isokoski, P., Kangas, J., Rantala, J. and Raisamo, R. (2014) TraQuMe: A Tool for Measuring the Gaze Tracking Quality. In Proceedings of the Symposium on Eye Tracking Research and Applications. ACM. pp. 327–330.
- Akkil, D., James, J.M., Isokoski, P. and Kangas, J. (2016) GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM. pp. 1151–1158.
- Alem, L. and Li, J. (2011) A study of gestures in a video-mediated collaborative assembly task. *Adv. Hum. Comput. Interact.*, 2011, 1. doi:10.1155/2011/987830.
- Bard, E.G., Hill, R.L., Foster, M.E. and Arai, M. (2014) Tuning accessibility of referring expressions in situated dialogue. *Lang. Cogn. Nurs.*, 29, 928–949.
- Brennan, S.E., Chen, X., Dickinson, C.A., Neider, M.B. and Zelinsky, G.J. (2008) Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition*, 106, 1465–1477.
- Clark, H.H. and Krych, M.A. (2004) Speaking while monitoring addressees for understanding. *J. Mem. Lang.*, 50, 62–81.
- Clark, H.H. and Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition*, 22, 1–39.
- D'Angelo, S. and Begel, A. (2017) Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM. pp. 6245–6290.
- D'Angelo, S. and Gergle, D. (2016) Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM. pp. 2492–2496.
- Fussell, S.R., Kraut, R.E. and Siegel, J. (2000) Coordination of communication: Effects of shared visual context on collaborative work. In Proceedings of the 2000 ACM conference on Computer supported cooperative work. ACM. pp. 21–30.
- Fussell, S.R., Setlock, L.D., Parker, E.M. and Yang, J. (2003) Assessing the Value of a Cursor Pointing Device for Remote Collaboration on Physical Tasks. In CHI'03 Extended Abstracts on Human Factors in Computing Systems. ACM. pp. 788–789.
- Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E. and Kramer, A.D. (2004) Gestures over video streams to support remote collaboration on physical tasks. *Hum. Comput. Interact.*, 19, 273–309.
- García, P., Olvido, J., Ehlers, K.R. and Tylén, K. (2017) Bodily constraints contributing to multimodal referentiality in humans: the contribution of a de-pigmented sclera to proto-declaratives. *Lang. Sci.*, 54, 73–81.
- Gergle, D., Kraut, R.E. and Fussell, S.R. (2004) Action as Language in a Shared Visual Space. In Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04), Vol. 6(3). p. 487.
- Gergle, D., Kraut, R. and Fussell, S. (2013) Using visual information for grounding and awareness in collaborative tasks. *Hum. Comput. Interact.*, 28, 1–39.
- Graver, W.W., Sellen, A., Heath, C. and Luff, P. (1993) One Is Not Enough: Multiple Views in a Media Space. In Proceedings of INTERCHI. pp. 335–341.
- Guo, J. and Feng, G. (2013) How Eye Gaze Feedback Changes Parent-Child Joint Attention in Shared Storybook Reading? In Nakano, Y., Conati, C. and Bader, T. (eds) *Eye Gaze in Intelligent User Interfaces*. pp. 9–21. Springer, London.
- Gupta, K., Lee, G.A. and Billingham, M. (2016) Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Trans. Vis. Comput. Graph.*, 22, 2413–2422.
- Gutwin, C. and Penner, R. (2002) Improving Interpretation of Remote Gestures with Telepointer Traces. In Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work. ACM. pp. 49–57.
- Hess, S.M. and Detweiler, M.C. (1994) Training to Reduce the Disruptive Effects of Interruptions. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 38, pp. 1173–1177. SAGE Publications, Sage, CA/Los Angeles, CA.
- Higuch, K., Yonetani, R. and Sato, Y. (2016) Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM. pp. 5180–5190.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, 1, 65–70.
- Jacob, R.J. (1991) The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.*, 9, 152–169.
- Kim, S., Lee, G.A., Sakata, N., Dunser, A., Vartiainen, E. and Billingham, M. (2013) Study of Augmented Gesture Communication Cues and View Sharing in Remote Collaboration. In 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE. pp. 261–262.
- Kirk, D. and Stanton Fraser, D. (2006) Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. pp. 1191–1200.
- Kurata, T., Sakata, N., Kourogi, M., Kuzuoka, H. and Billingham, M. (2004) Remote Collaboration using a Shoulder-worn Active Camera/Laser. In Eighth International Symposium on Wearable Computers, 2004 (ISWC 2004), Vol. 1. IEEE. pp. 62–69.
- Li, J., Wessels, A., Alem, L. and Stitzlein, C. (2007) Exploring Interface with Representation of Gesture for Remote Collaboration. In Proceedings of the 2007 Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of

- Australia on Computer-Human Interaction: Design: Activities, Artifacts and Environments (OZCHI '07), p. 179. <https://doi.org/10.1145/1324892.1324926>
- Macdonald, R.G. and Tatler, B.W. (2017) Do as eye say: gaze cueing and language in a real-world social interaction. *J. Vis.*, 13, 1–12.
- Mark, G., Gudith, D. and Klocke, U. (2008) The Cost of Interrupted Work: More Speed and Stress. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. pp. 107–110.
- Müller, R., Helmert, J.R. and Pannasch, S. (2014) Limitations of gaze transfer: without visual context, eye movements do not to help to coordinate joint action, whereas mouse movements do. *Acta Psychol. (Amst)*, 152, 19–28.
- Müller, R., Helmert, J.R., Pannasch, S. and Velichkovsky, B.M. (2013) Gaze transfer in remote cooperation: is it always helpful to see what your partner is attending to? *Q. J. Exp. Psychol.*, 66, 1302–1316.
- Neider, M. B., Chen, X., Dickinson, C. A., *et al.* (2010) Coordinating spatial referencing using shared gaze. *Psychon Bull Rev.* 17, 718–724.
- Olson, G.M. and Olson, J.S. (2000) Distance matters. *Human-computer interaction. HuM. Comput. Interact.*, 15, 139–178.
- Qvarfordt, P., Beymer, D. and Zhai, S. (2005) Realtourist—a study of augmenting human-human and human-computer dialogue with eye-gaze overlay. *Hum. Comput. Interact.*, 2005, 767–780.
- Schneider, B. and Pea, R. (2013) Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *Int. J. Comput. Support. Collab. Learn.*, 8, 375–397.
- Stein, R. and Brennan, S.E. (2004) Another Person's Eye Gaze as a Cue in Solving Programming Problems. In Proceedings of the 6th International Conference on Multimodal Interfaces. ACM. pp. 9–15.
- Velichkovsky, B.M. (1995) Communicating attention: gaze position transfer in cooperative problem solving. *Pragmatics Cogn.*, 3, 199–223.
- Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. (2011) The Aligned Rank Transform for Nonparametric Factorial Analyses using only ANOVA Procedures. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. pp. 143–146.
- Wong, J., Oh, L.M., Ou, J., Rosé, C.P., Yang, J. and Fussell, S.R. (2007) Sharing a Single Expert among Multiple Partners. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. pp. 261–270.
- Yamazaki, K., Yamazaki, A., Kuzuoka, H., *et al.* (1999) GestureLaser and GestureLaser Car Development of an Embodied Space to Support. In ECSCW 1999: Proceedings of the Sixth European Conference on Computer Supported Cooperative Work, September. pp. 12–16.



Paper 6

Deepak Akkil, Biju Thankachan, and Poika Isokoski. 2018. I see what you see: gaze awareness in mobile video collaboration. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 32, 9 pages. DOI: [10.1145/3204493.3204542](https://doi.org/10.1145/3204493.3204542)

© ACM 2018, Reprinted with permission.

I See What You See: Gaze Awareness in Mobile Video Collaboration

Deepak Akkil
Tampere Unit for Computer-Human
Interaction (TAUCHI)
University of Tampere, Finland
deepak.akkil@uta.fi

Biju Thankachan
Tampere Unit for Computer-Human
Interaction (TAUCHI)
University of Tampere, Finland
biju.thankachan@uta.fi

Poika Isokoski
Tampere Unit for Computer-Human
Interaction (TAUCHI)
University of Tampere, Finland
poika.isokoski@uta.fi

ABSTRACT

An emerging use of mobile video telephony is to enable joint activities and collaboration on physical tasks. We conducted a controlled user study to understand if seeing the gaze of a remote instructor is beneficial for mobile video collaboration and if it is valuable that the instructor is aware of sharing of the gaze. We compared three gaze sharing configurations, (a) *Gaze_Visible* where the instructor is aware and can view own gaze point that is being shared, (b) *Gaze_Invisible* where the instructor is aware of the shared gaze but cannot view her own gaze point and (c) *Gaze_Unaware* where the instructor is unaware about the gaze sharing, with a baseline of shared-mouse pointer. Our results suggests that naturally occurring gaze may not be as useful as explicitly produced eye movements. Further, instructors prefer using mouse rather than gaze for remote gesturing, while the workers also find value in transferring the gaze information.

CCS CONCEPTS

•Human-centered computing → Collaborative and social computing; Computer supported cooperative work;

KEYWORDS

Mobile phone, video, communication, collaboration, gaze awareness, implicit, explicit, video conferencing, physical task

ACM Reference format:

Deepak Akkil, Biju Thankachan, and Poika Isokoski. 2018. I See What You See: Gaze Awareness in Mobile Video Collaboration. In *Proceedings of 2018 Symposium on Eye Tracking Research and Applications, Warsaw, Poland, June 14–17, 2018 (ETRA '18)*, 9 pages. DOI: 10.1145/3204493.3204542

1 INTRODUCTION

Mobile devices such as smart phones and tablet computers have revolutionized how we communicate. In addition to communication using text and audio, recent advancements in video technologies and network connectivity, have enabled seamless anytime-anywhere video communication using mobile devices. There are numerous

video telephony services and applications that support communication between devices of multiple form factors (e.g. Skype). These services allow a mobile user to video call a remote partner who could be using another mobile device, desktop computer or laptop.

There is currently a growing trend to move beyond a "talking head" video communication, towards using video to support joint activities and share experiences between remote users. Imagine a traveller exploring a new city with the help of a remote guide, a novice driver troubleshooting an issue with the car with the guidance of a remote mechanic, an industrial field worker repairing a complex machinery with the help of an indoor expert, or a shopper video calling a friend to seek suggestions on what to buy from a store. The mobility and flexibility offered by mobile video telephony makes smartphones an ideal choice of device to collaborate in such situations. All the scenarios above involve a mobile user seeking guidance from a stationary remote partner who may be using a desktop/laptop computer for the collaboration.

Effective collaboration in these novel scenarios requires tools and features to improve the mutual awareness of collaborators and to efficiently communicate complex spatial and procedural information. Previous studies using stationary camera set-ups have shown that remote gesturing mechanisms, such as mouse [Fussell et al., 2004], gaze [Akkil et al., 2016], pen-based annotation systems [Fussell et al., 2004] and hand representations [Kirk and Stanton Fraser, 2006], could be beneficial in satisfying these needs.

Gaze-tracking technology is now increasingly available, at lower prices than ever before (e.g. Tobii 4C). In a video-based remote collaborative scenario involving a mobile user and a stationary computer user, it is now possible to accurately track the gaze of the stationary user and present this information, in real-time, on the mobile phone display of the collaboration partner.

Previous studies have explored the value of gaze awareness in remote collaborations involving stationary tasks performed on a computer screen [Brennan et al., 2008, Qvarfordt et al., 2005], or physical tasks involving limited mobility [Akkil et al., 2016, Gupta et al., 2016]. The results indicate that gaze awareness could enable easier collaboration by allowing effortless reference to spatial information and contribute to an improved feeling of presence. However, similar studies using mobile video communication do not exist.

There are two inherent differences between the stationary set-ups studied in the previous work and mobile phone video collaboration. In mobile video collaboration, the visual information communicated to the stationary user is limited by the field of view of the camera and fully controlled by the mobile user. The frequent movement and the subsequent view changes may affect the gaze

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.

ETRA '18, Warsaw, Poland

© 2018 ACM. 978-1-4503-5706-7/18/06...\$15.00

DOI: 10.1145/3204493.3204542

behaviour of the remote user and thus the usefulness of shared gaze. Further, the mobile user needs to shift the attention between the hand-held mobile display to acquire the gaze information of the remote user and the physical world to perform the task. Thus, specific cues, even if accurately transferred, may not always be perceived by the mobile user, or used in the collaboration.

Further, previous studies that investigated the usefulness of gaze awareness in remote collaboration have used different design for the shared gaze interface. For example, Qvarfordt [2005] studied the value of gaze produced implicitly, as opposed to intentionally, in a collaborative trip-planning task. The participant whose gaze was shared to the collaboration partner was not aware of the gaze sharing and thus did not use it as an explicit mechanism to communicate. D'Angelo et al. [2016] studied two-way shared gaze in a collaborative puzzle-solving task. In their study, the participants were aware that gaze was being shared and used it explicitly. However, they did not get any feedback on the actual gaze point by the tracker i.e. they did not see the gaze point themselves. In contrast, Akkil et al. [2016] studied shared gaze in a collaborative construction task, where gaze of the remote user was physically projected onto the task space. The remote user was thus not only aware of the shared gaze, but could also see the physical projection of it in the camera view, attaining direct feedback on the gaze data returned by the tracker. All other previous studies involving shared gaze communication have used one of these three configurations. However, the previous studies have not compared these three configurations in terms of efficiency and user experience.

We conducted a controlled user study on mobile video communication in an object arrangement task. A remote stationary instructor knew the arrangement of the objects but could not act on the objects. The mobile worker could manipulate the objects but did not know the target arrangement. The target arrangement required the instructor to identify the right block, specify the target location of the block, and communicate the 3D orientation of the object. Thus, the task required communicating both pointing and procedural instructions. The focus of the study was to compare the following four configurations.

- **Gaze_Unaware:** Instructor is unaware of the gaze sharing. The worker is aware that the gaze is implicitly produced.
- **Gaze_Invisible:** Instructor is aware of the gaze sharing. However, the instructor cannot see her own gaze data, only the worker can.
- **Gaze_Visible:** Instructor is aware of the gaze sharing, and both instructor and worker can see the gaze information of the instructor.
- **Mouse:** Mouse position of the instructor is continuously shared to the worker.

We begin by reviewing the relevant related work. Then, we describe our study. Next, we report the results of our study, followed by discussion of the results and their practical implications.

2 RELATED WORK

2.1 Mobile Video Collaboration

The flexibility offered by the mobile devices to easily switch camera feeds and change device orientation enable their use for collaborative physical tasks. O'Hara [2006] conducted a diary study. In their

sample 28% of video calls involved showing things in the environment to talk about and 22% were for performing functional tasks (e.g. planning events or seeking guidance). Similarly, Brubaker et al. [2012] note a growing trend in using video to support joint activities (e.g. seeking guidance to accomplish physical tasks) and experiences (e.g. giving a tour of a flat). Jones et al. [2015] studied how people collaborate using mobile video and found that a serious shortcoming of commercial mobile video conferencing services is the lack of support for remote gesturing, which is known to be important to efficiently use video as a collaborative activity space. Previous work has shown the growing trend in mobile video telephony to use "video-as-data", instead of the conventional "talking heads" [Nardi et al., 1993]. These new applications of mobile video require gesturing mechanisms e.g. to point out interesting details in the environment, or to effectively communicate procedural instructions in a physical task.

2.2 Gaze sharing in collaboration: Does the level of awareness matter?

There have been numerous studies on gaze awareness in collaboration, in tasks involving visual search [Brennan et al., 2008], programming [Stein and Brennan, 2004], trip-planning [Qvarfordt et al., 2005] and puzzle-solving [Velichkovsky, 1995]. The most common approach to provide gaze awareness in collaboration is to present the gaze of the partner as an abstract visual element, such as a dot, ring or icon of the eye, overlaid on the shared visual space (notable exceptions are Trosterer et al. [2015] and D'Angelo et al. [2017]).

The previous studies on shared gaze can be classified, based on the level of awareness the producer of the gaze has on the gaze sharing. Qvarfordt et al. [2005] studied value of naturally occurring eye movement in a collaborative trip-planning task and found that gaze, even if not explicitly produced with the intention to communicate can aid deictic referencing, aid topic switching and help reduce ambiguity in communication. Similarly, Stein et al. [2004] found that eye gaze produced instrumentally (as opposed to intentionally), can help problem solving in a programming task. Liu et al. [2011] noted that naturally occurring gaze can help to efficiently achieve referential grounding. In all these studies, the producer of gaze was not aware that the partner would see their gaze point and thus did not use gaze as an explicit communication channel. Thus, the gaze point reflects their natural gaze behaviour.

In contrast, other studies used a setup where the collaborator is aware that their gaze is being shared, and thus they use their gaze more explicitly to communicate. However, some studies showed the collaborators their own gaze point, providing accurate awareness of the point that is transferred and others did not show own gaze point to the collaborator. Akkil et al. [2016] and Higuch et al. [2016] studied a set-up where gaze of the remote instructor was physically projected to the task space of the partner. Thus, the instructor saw the physical projection of their own gaze point on the video captured by the situated camera, giving the instructor direct feedback of their own eye movements and accuracy of gaze tracking. Others have studied shared gaze in a collocated scenario, where both the collaborators are in front of the same display, enabling the collaborator to see their own gaze. Zhang et al. [2017] studied

collocated visual search on a large screen, Maurer et al. studied gaze sharing between a collocated game spectator and gamer [2015], and between passenger and driver in a driving simulator [2014]. Similar set-up was also used by Duchowski et al. [2004] in collaborative virtual environments.

Similarly there are a number of studies involving shared gaze used in a set-up where the collaborators are aware of shared gaze and saw their partner's gaze, but did not see their own gaze point. Examples include, Brennan et al. [2008] in a collaborative visual search task, D'Angelo et al. [2016] and Muller et al. [2013] in puzzle-solving tasks, Lankes et al. [2017] during online game viewing and Maurer et al. [2016] during online cooperative gaming. Interestingly, Maurer et al. [2016] note that their participants commented they would have liked to see their own gaze point, along with the partner's gaze. In contrast, D'Angelo [2016] note that showing the own gaze pointer may not be a good idea, since it "can produce a feedback loop that causes people to follow their own cursor", when gaze tracking is not accurate.

In summary, previous studies on shared gaze have used three different configurations of gaze sharing, and found value for all three in the collaboration. This brings us to the question, are they all equally effective? A comparative evaluation between the three setups would give us novel insights into the utility of each of the configurations. This was the focus of our work.

2.3 Gaze Awareness in Collaborative Physical Tasks

Akkil et al. [2016] noted that gaze overlaying in egocentric videos improves accuracy of interpreting hand-pointing gestures. Similarly, Gupta et al. [2016] found that in collaborations using head-mounted cameras, gaze sharing improves collaboration performance in a stationary LEGO building task. Other studies explored physically projecting gaze to the task space in a circuit assembly [Akkil et al., 2016] and block arrangement task [Higuch et al., 2016]. They found that gaze sharing made referring objects easier and also improved the feeling of presence between collaborators. We recently conducted a study involving an object arrangement task using a similar experimental setup, involving stationary cameras and physical projection of gaze, but we compared shared gaze with a shared mouse for remote gesturing [Akkil and Isokoski, 2018]. We found that shared gaze improved collaboration compared to having no gesturing mechanism at all. However mouse outperformed gaze in both objective and subjective measures. There was no difference between shared gaze and shared mouse cursor in tasks that required only pointing. However, when the task required conveying procedural instructions (e.g. "turn the object like this or orient it like this"), mouse was the better of the two remote gesturing mechanisms. In summary, previous studies on shared gaze in collaborative physical tasks involving stationary tasks have shown that while gaze is useful, it may not be as useful as the mouse. The focus of this study is on mobile video collaboration. Mobility of the task and additional complexity due to the hand-held device may influence how the remote gesturing is perceived and used by the instructor and the worker, and perhaps even on the effectiveness of gaze and mouse-based remote gesturing.

3 USER STUDY

We conducted a controlled user study, using a within-subject experimental design, with 4 experimental conditions: *Gaze_Unaware*, *Gaze_Invisible*, *Gaze_Visible*, and *Mouse*. Our study focused on answering the following research questions:

- **RQ1: Does sharing gaze of the instructor that is produced implicitly, provide benefits comparable to when the gaze is produced with the intention to communicate?**

When the instructors are aware of the shared gaze, they may use it explicitly to communicate (e.g. "pick the object I am looking at"). When the instructor is not aware of gaze sharing, the eye movement of the instructor reflects their natural gaze behaviour. Thus, by experimentally manipulating the awareness of the instructor regarding gaze sharing, we get insights into the usefulness of sharing natural gaze versus intentional gaze.

- **RQ2: Does the visibility of the gaze point on the instructor's side influence the usability of shared gaze interface and the collaboration dynamics?**

Seeing one's own gaze data can be helpful in multiple ways. When the instructors can view the gaze pointer, they may be more likely to explicitly use it in the collaboration. Further, the instructors can be more aware of their own gaze behaviour and when it can be potentially misleading to the worker, and verbally correct it (e.g. *Please wait, I am searching for the block*). It also allows the instructor to be more aware of the accuracy of gaze tracking and to correct the gaze pointer, when it can be potentially misleading to the worker (e.g. by looking slightly away from the target, so that gaze pointer is on the target). On the other hand, visualizing the gaze data can be potentially distracting [D'Angelo and Gergle, 2016].

- **RQ3: How does shared gaze compare against a shared mouse pointer in a mobile collaborative physical task?**

Gaze and mouse have several commonalities but also unique affordances. Gaze automatically conveys attention and intention, while mouse needs explicit user action to be meaningful. When the mobile camera moves, the mouse pointer needs to be manually adjusted to keep it on the target, while the gaze of the instructor will automatically track the target even during camera movements. Further, the automaticity provided by the gaze ensures a level consistency between the information that is transferred between pairs, while there may be large variability between instructors on the extent of use of mouse for remote gesturing [Muller et al., 2013]. On the other hand, mouse allows pointing accurately and gesturing flexibly (e.g. by drawing shapes or conveying rotations).

We did not include a no pointer condition as the baseline in our study, as is generally the practice followed in previous studies involving shared gaze collaboration [Akkil et al., 2016, Brennan et al., 2008, D'Angelo and Begel, 2017, Qvarfordt et al., 2005], for three main reasons. First, mouse is also a plausible remote gesturing mechanism in our context of investigation and therefore also

serves as a "stricter" baseline comparison. If any of the gaze configurations performs as good/better than mouse, then most likely they also perform better than having no pointer at all. Second, the *Gaze_Unaware* condition was disguised as a no pointer condition to the instructor, and thus adding another no pointer condition might have influenced the credibility of *Gaze_Unaware* condition (e.g. instructors could have been suspicious about two No pointer conditions). Third, because earlier studies have already included the no pointer condition multiple times, the likelihood of new findings regarding it was lower than with the mouse and gaze comparison, despite the new mobile context.

The *Gaze_Unaware* condition was included in the study only to understand the usefulness of naturally occurring gaze in collaboration. The gaze of a user may contain private information regarding the person's preferences, emotions, and personality. Sharing gaze of a user without their knowledge is a privacy violation and we do not recommend developing services and applications to share gaze of a user without their knowledge. However, our study was conducted in a controlled lab environment and consequently the deception involved was benign and also revealed to the participants at the end of the study with the option of withdrawing their data from the study without penalty. In addition to theoretical interest, the utility of including the *Gaze_Unaware* condition was that it resembles a situation where the instructor is aware of, and accepts, the gaze being transmitted, but no longer pays attention to it.

3.1 Apparatus and Experimental Set-up

The instructor and mobile worker collaborated from two adjacent sound-proof rooms. At the worker's end, there were two tables, the blocks to arrange were placed on one table and the task was performed on another table, placed at 2.5 meter distance. The worker used a Samsung Galaxy S4 smartphone with the camera, microphone, and speaker for video and verbal communication. The rear camera feed of the phone was displayed on the phone and also streamed to the remote instructor. In addition to the video, the worker saw either gaze or mouse pointer of the instructor, visualised as a semi-transparent blue dot with 1cm diameter.



Figure 1: The setup at the worker's end. One table had the blocks to arrange and the other table was the taskspace where the blocks had to be arranged.

At the instructor's end, we used a Tobii T60 gaze tracker. The mobile worker's camera feed was shown on the T60 display. In addition, the instructor communicated to the worker, via a headphone. For video communication between the mobile phone and



Figure 2: The setup at the instructor's end. The blue dot on screen indicated the gaze pointer. The same pointer was visible at the worker's end.

instructor's computer, we developed a custom LAN-based video conferencing system using the Javascript WebRTC API and Microsoft .NET 4.5. The experimental software collected the gaze and mouse cursor location at the instructor's end and transferred it to the browser-based video-calling clients. The pointer was displayed on the instructor's computer only during *Gaze_Visible* and *Mouse* condition. The worker saw the pointer in all four conditions. We used a recursive filter (with weight for current gaze position $W=0.3$), to smoothen the gaze data following previous studies [Akkil et al., 2016, Qvarfordt et al., 2005]. This additional smoothing was not applied on the mouse data.

3.2 The Task

The experimental task was to arrange 10 unique pentomino puzzle blocks on specific locations and orientation on an A3 sized paper. The paper was marked with 60 randomly generated non-overlapping dots and each pentomino block had to be placed on one of the dots. For the experiment, we chose 4 arrangement tasks of comparable complexity. For each task, different background dot arrangement was used. Figure 3 shows two representative arrangements used in the experiment. The expert was given a physical model of the structure to build. The task for the pairs was to collaborate over the mobile video link, to successfully arrange the blocks as quickly and as accurately as possible. Even though the chosen task was artificial in nature, it had different sub-tasks such as identifying linguistically complex objects and performing 3D manipulations such that also appear in real-world tasks. Using pentomino puzzle blocks allowed us better opportunities for repeating and isolating these tasks than the real-world tasks that we found. Also, it enabled us to create multiple tasks of comparable complexity, enabling us to leverage the strength of a repeated measures experimental design.

3.3 Participants

We recruited 24 participants (12 pairs, 10F, 14M) from the University community, with ages between 20-34 years ($M=25.4$, $SD=4.0$). The participants were either allowed to sign up in pairs or individually. The individuals were paired by the experimenter. All participants had normal (10 instructors, 6 workers) or corrected to normal vision (2 instructors, 6 workers). 11 participants were unfamiliar with gaze

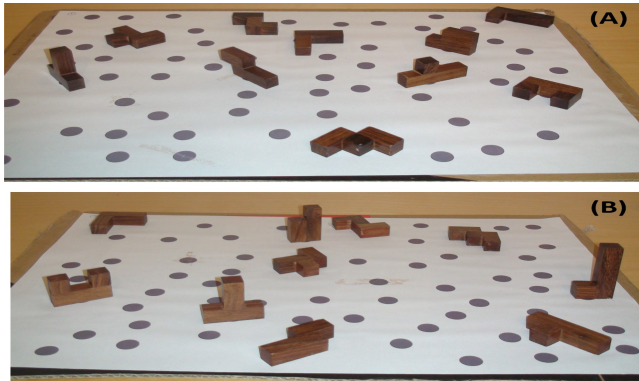


Figure 3: Two representative block arrangements from the study. The task was to arrange 10 pentomino blocks in specific location and 3D orientation.

tracking, while the remaining had some experience with it from previous experiments/courses. All the participants were frequent users of smartphones. All instructors were experienced in using a mouse. All participants were also proficient English speakers.

3.4 Procedure

At the beginning of the experiment the participants were given an overview of the study and introduced to the study set-up. They signed an informed consent form and completed a background questionnaire. The participants were then assigned to the role of instructor or worker. Since we expected a learning in the use of the mobile device and in arranging the puzzle blocks, the participants completed two practice tasks. First, a pentomino block arrangement task involving 12 blocks with both participants standing next to the same table. Instructors were allowed to use hand gesturing in this practice task. Then, the instructor was seated in front of the Tobii T60 tracker, in the adjacent room, followed by completing a 9-point gaze tracker calibration. The instructor was shown her own gaze point, and the calibration was repeated if the instructor/moderator felt gaze tracking was not accurate enough. Later, after ensuring that the audio/video communication worked as intended, the pairs completed another round of practice using the mobile video communication. A paper with 3X3 grid of dots was given to the worker. The instructor was asked to randomly pick 8 pairs of dots and ask the worker to connect those dots with pen. This task was repeated two times (with mouse and *Gaze_Visible* conditions). The *Gaze_Visible* condition was chosen for the practice because it allowed the instructor to understand how the eyes move naturally while giving instructions (e.g. while searching for the block), which could have helped the collaborators in other gaze conditions.

Later, the different experimental conditions were executed. Once the task was completed successfully, both the instructor and worker were asked to fill in a short questionnaire with 5 different questions using the 7-point Likert scale to evaluate the perceived quality of the collaboration. Soon after the completion of the gaze conditions, the gaze data quality was measured using a 9-point quality evaluation process using TraQuMe [Akkil et al., 2014]. TraQuMe shows

predefined fixation points on screen, and measures the accuracy and precision of tracking.

Before each condition, the worker was made aware of the current pointer control mechanism (i.e. *Gaze_Unaware*, *Gaze_Visible*, *Gaze_invisible*, *Mouse*). The *Gaze_Unaware* condition was disguised as the no pointer condition to the instructor, and the worker was explicitly instructed that the blue dot represents the naturally occurring gaze of the instructor. The worker was allowed to take advantage of the gaze pointer, but prohibited from telling the instructor that it was visible. After the post-test questionnaire was completed, the instructor was made aware of the deception in gaze sharing and its rationale. For each trial slot, the block arrangement was fixed. The order of the conditions was counterbalanced between participants. The experiment was video-recorded for later analysis.

3.5 Data Collected and Related Analysis

Human-human collaboration is often complex and plastic. Simple measures such as task completion time may not always reflect the effect of experimental manipulations, such as viewing or not viewing own gaze point, even if an effect exists. Thus, along with measures such as task completion times, we also recorded and analysed the conversation of the collaborators, and subjective opinions and preferences using questionnaires. The post-trial questionnaire had five 7-point Likert scale questions: (1) Ease of collaborating (2) Ease of providing/Understanding instructions, (3) Ease of referring objects, (4) Presence, and (5) Enjoyment.

The video was first transcribed. Then, we analysed the conversation between the expert and worker by counting the number of phrases required to complete the task, similar to previous works [Fussell et al., 2000, Gupta et al., 2016]. A phrase was defined as a distinct verbal utterance. All the statistical analysis was performed using one-way ANOVA and post-hoc testing using the t-test. For the 7-point Likert scale questionnaire data that did not follow a normal distribution, we used the non-parametric Friedman's rank sum test and post-hoc testing using Wilcoxon signed-ranks test. When interpreting the tests, we used the Benferroni-Holm procedure [Holm, 1979] for family-wise type-1 error rate correction with alpha at .05.

4 RESULTS

4.1 Overall task completion times

First, we analysed the effect of learning on the task completion times. A one-way repeated measures ANOVA showed no statistically significant differences in task completion times for the four trial slots ($F(3,33)=1.2$, $p=.33$) i.e. participants were not statistically significantly faster as the experiment proceeded. Next, we analysed the effect of the experimental conditions on the overall task completion times. Figure 4 shows the boxplot of task completion times for the different conditions. A one-way repeated measures ANOVA showed a statistically significant effect of conditions on task completion times, $F(3,33)=5.6$, $p=.003$. Post-hoc t-test showed that *Mouse* ($M=360.0$, $SD=52.6$) was significantly faster than *Gaze_Unaware* ($M=447.8$, $SD=68.7$), $t(11)=3.86$, $p=.016$. The difference between *Gaze_Invisible* ($M=391.1$, $SD=46.5$) and *Mouse* $t(11)=2.68$, $p=.085$, *Gaze_Visible* ($M=394.1$, $SD=78.3$) and *Gaze_Unaware* $t(11)=2.50$, $p=$

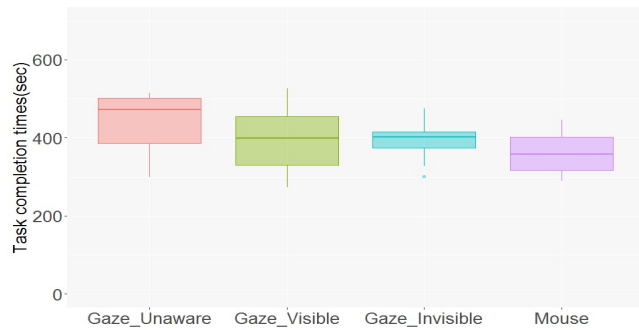


Figure 4: The task completion times for the different experimental conditions

.09, and *Gaze_Invisible* and *Gaze_Unaware* $t(11)=2.95$, $p=.065$ were only approaching statistical significance after the Benferroni-Holm correction. Other differences were not statistically significant ($p>.10$).

4.2 Conversation Analysis

Figure 5 shows the total number of utterances (i.e. sum of utterances by the instructor and worker). In the presence of an accurate and unambiguous remote gesturing mechanism, the verbal effort required to complete the task was expected to reduce. A repeated measures ANOVA showed a statistically significant difference in the number of utterances required to complete the task for the different experimental conditions, $F(3,33)=6.4$, $p=0.001$. Follow up t-tests showed that in the *Gaze_Unaware* condition ($M=126.6$, $SD=24.8$), participants required statistically significantly more verbal effort than all other conditions: *Gaze_Visible* ($M=101.6$, $SD=18.8$), $t(11)=4.6$, $p=.004$, *Gaze_Invisible* ($M=99.5$, $SD=18.4$) $t(11)=3.17$, $p=0.045$, and *Mouse* ($M=98.1$, $SD=26$) $t(11)=3.04$, $p=.045$. Other differences were not statistically significant ($p>.05$).

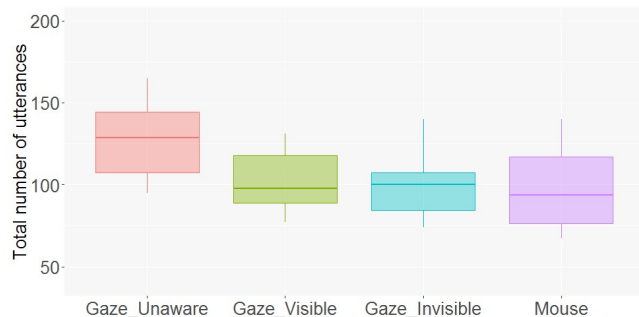


Figure 5: Boxplot showing total number of utterances spoken to finish the task for the different conditions.

4.3 Questionnaire data

Friedman's rank sum test showed significant differences in responses to the post-trial questionnaire for the instructor in 4 questions (ease of collaborating, ease of providing instructions, presence and enjoyment of using the interface) and in 1 question for the

worker (ease of collaborating). Follow up comparison, using the Wilcoxon rank-sum test, showed significant pairwise differences for 3 questions for the instructor, after Benferroni-Holm correction. See figure 6 for the boxplots and Figure 7 for summary of the analysis of significant results.

4.4 Gaze Tracking Accuracy

The average accuracy of gaze tracking varied from 0.34cm (0.27 deg) to 2.67cm (2.19 deg) on the desktop display ($M= 1.14$ cm, $SD=0.52$ cm). This offset in screen distance was proportionally reduced when gaze was presented on the mobile display. Friedman's rank sum test showed no statistically significant difference in the average accuracy of gaze tracking (for 9 validation points), between the three gaze conditions, ($\chi^2(3)=0.844$, $p=.65$). An increase in gaze-tracking offset was weakly correlated with increase in task completion times in the case of *Gaze_Invisible* ($r=0.58$, $n=12$). Similar correlation did not exist in the case of *Gaze_Visible* ($r=0.22$, $n=12$) and *Gaze_Unaware* ($r=-0.01$, $n=12$).

4.5 User Preferences

Overall, the user preference was mixed and there was a difference in preferences based on the roles of the participants. Eight instructors preferred the mouse sharing condition, while two felt *Gaze_Invisible* and mouse were equally good and the rest preferred *Gaze_Invisible*. On the other hand, only four workers preferred mouse. Interestingly, three workers preferred the *Gaze_Unaware* condition, while the remaining five workers felt both *Gaze_Visible* and *Gaze_Invisible* to be comparable and preferred. The following participant comments illustrate the different factors that the users considered when making their preference decisions:

- **P5 Instructor:** "With mouse, I could express myself better and even describe the actions. When [my own] gaze point was visible, it was a bit annoying. When gaze point was invisible, I was not confident at the first, that my partner knew where I was looking. After a while, I could trust it more." (Preferred Mouse).
- **P3 Worker:** [In *Gaze_Visible* and *Gaze_Invisible* conditions], My partner pointed dots well and did not try to overuse it like [in the] Mouse [condition]."
- **P11 Worker:** "Gaze felt more natural. Eyes would correct the pointer even if the mobile device moved in my hand" (Preferred both *Gaze_Visible* and *Gaze_Invisible* conditions).
- **P6 Worker:** "When other person was unaware,She give good verbal instructions. I could easily focus on the verbal instruction and use gaze as a support". (preferred *Gaze_Unaware*)

5 DISCUSSION

RQ1: Does sharing gaze of the instructor that is produced implicitly, provide benefits comparable to when the gaze is produced with the intention to communicate?

It can be argued that sharing gaze information that is produced implicitly can be useful compared to having no pointer at all. There were numerous instances, where the worker relied on the implicit gaze of the instructor to identify the correct block, and the target

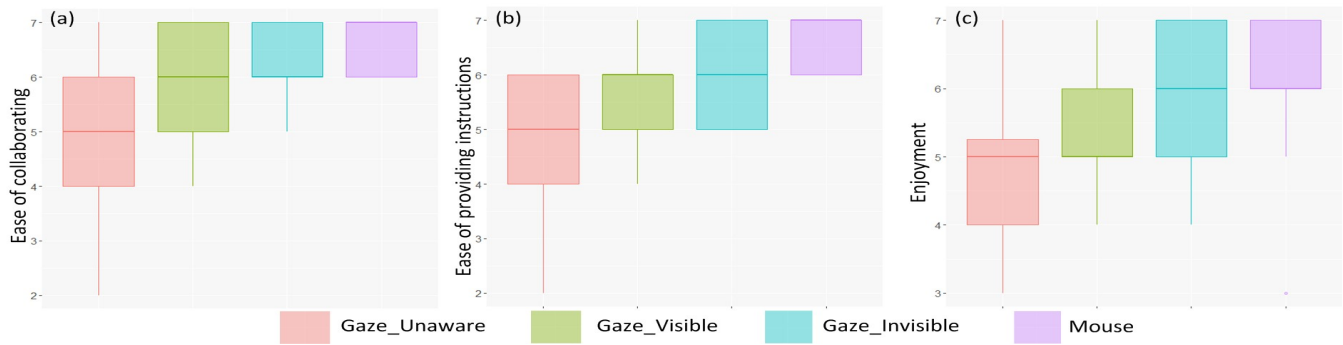


Figure 6: Boxplot showing responses to the post-test questionnaire for 3 questions for the instructor (a) ease of collaborating, (b) ease of providing instructions, and (c) Enjoyment.

Question	Role	Friedman's omnibus test		Post-hoc analysis
		χ^2	p-value	
Ease of collaborating	Instructor	19.7	<.01	G_unaware - Mouse : p=.02 G_unaware - G_visible: p=.06 G_visible - Mouse: p=.07
Ease of providing/understanding	Instructor	19.3	<.01	G_unaware - Mouse : p=.03 G_unaware - G_visible: p=.04 G_visible - Mouse: p=.05
Enjoyment	Instructor	17.4	<.01	G_unaware - Mouse : p=.02 G_unaware - G_visible: p=.06

Figure 7: Summary of analysis of questionnaire responses for the instructor. Differences were not significant for the worker.

location, even when the verbal instructions were ambiguous or incomplete. The implicit gaze of the instructor was useful as a supporting modality to understand the verbal instructions. In the absence of awareness of the shared gaze, the instructors often used extensive verbal communication to establish a shared understanding regarding the target locations, that were otherwise not directly required for the task. (e.g. *"the dot to the left of the block we just put, two dots to the top of that dot is where you should place this block"*). Implicit gaze helped the worker understand these indirect instructions easily. However often times implicit gaze did not directly lead them to the exact target block or location. Thus, in the *Gaze_Unaware* condition, pairs spent considerably more time and verbal effort to complete the task. Also, we observed that when not aware of the shared gaze, some instructors spent more time looking at their target arrangement model while formulating the verbal instructions or used hand gestures while communicating possibly obstructing the gaze tracker.

In terms of subjective preference, none of the instructors preferred the *Gaze_Unaware* condition, while interestingly three workers did stating that this was because the instructor gave extensive verbal instructions in this condition. Thus, in the *Gaze_Unaware* condition, the worker could focus their attention on the task space, relying on the verbal instructions of the partner to complete the task. Switching attention to the mobile device was necessary only when the verbal instructions were ambiguous. In contrast, while

the gaze was used explicitly to communicate, the worker needed to attend to the mobile display every time the instructor used gaze to explicitly communicate (e.g. *"take this block"*).

Our results suggests that sharing implicit gaze, while useful in the collaboration to support verbal instructions and subjectively preferred by some workers, may not be as useful as explicit use of gaze or mouse in terms of task completion times and verbal effort required to complete the task.

RQ2: Does the visibility of the gaze point on the instructor's side influence the usability of shared gaze interface and the collaboration dynamics?

There were no significant differences between the *Gaze_Invisible* and *Gaze_Visible* conditions in terms of the objective measures such as task completion time and number of utterances required to complete the task. However, there was differences in preference between the two conditions based on the roles of the collaborators. The workers generally felt both the conditions were comparable. However, majority of our instructors preferred the *Gaze_Invisible* condition to *Gaze_Visible*, as seeing their own gaze point was distracting and few participants also specifically mentioned that their eyes were more strained after the *Gaze_Visible* condition.

In the *Gaze_Invisible* condition, the instructor did not receive any feedback regarding status of gaze tracking (i.e. are the eyes being tracked without any technical problems?), or accuracy of gaze tracking. From our observations, it was evident that the instructors could have benefited from such a feedback. Our instructors often tried to get these information by other means, by asking the worker (e.g. *"can you see the dot now?"*, *"can you point where I am looking at now?"*). Our results suggests that in real-time shared gaze applications, it would be best to provide a feature to toggle the visualisation of own gaze on and off for the instructor, in order to allow the instructor to easily ascertain quality of tracking when needed, while avoiding the distraction of seeing own gaze point. Zhang et al. [2017] have also proposed the option to toggle gaze pointer ON and OFF, in the context of co-present collaborative visual search on a large display.

The gaze tracking accuracy showed a weak correlation to the task completion times in the case of *Gaze_Invisible*, while such a correlation did not exist in the case of *Gaze_Visible*. Our results

give preliminary indication that the *Gaze_Invisible* maybe more prone to issues with accuracy of tracking than *Gaze_Visible*. When the accuracy of tracking is low and the instructor is aware of it, they may pro-actively try to overcome this inaccuracy, e.g. by explicitly adjusting the gaze point by looking slightly away from the target or complementing the gaze information with additional verbal instructions (e.g. *"a little bit to the left of where the point is"*). However, when the instructor cannot see their own gaze point, such situations may lead to wrong interpretation of the gaze pointer by the worker and would incur additional time and verbal effort to repair.

RQ3: How does shared gaze compare against a shared mouse pointer?

Previous results in stationary contexts suggest that shared mouse pointer may be more effective than shared gaze in collaborative physical tasks, since mouse enables providing complex procedural instructions e.g. by drawing shapes using the mouse cursor [Akkil and Isokoski, 2018]. However, this study focused on a mobile context and the differences between mouse and explicit use of gaze pointers were not that straight forward in terms of objective measures such as task completion times and verbal effort required to complete the task. Completing the task was faster with the mouse, on an average, than the *Gaze_Invisible* and *Gaze_Visible* conditions.

There was a difference in preferences between gaze and mouse pointers depending on the role of the participants. The majority of instructors preferred using mouse to gaze-based conditions and majority of the workers preferred one of the three gaze-based conditions. While the mouse does enable the instructor to draw shapes and accurately point, the mobile device introduces additional challenges on using mouse for remote gesturing.

First, the orientation of the mobile device is controlled by the worker and there are often minor movements of the device which changes the visual information presented to the instructor. For mouse to keep pointing at a location, the instructor needs to explicitly move the mouse to negate the device movement. Sometimes this led to situations where the worker misinterpreted the location pointed by the cursor or the instructor asking the worker to keep still (e.g. *"do not move, I am pointing"*). Such situations were rare when gaze was shared.

Second, we also noticed that there was some variability in how the instructors used mouse to communicate. Some instructors had to be reminded by the worker that they should use mouse (e.g. *"may be you can point you know"*), and not all of our instructor used the mouse to give complex procedural instructions, possibly because they felt it would be complex due to the mobile characteristics of the task. Further, when instructors did use mouse to communicate accurately point at parts of the block and actions (e.g. *"this small part of the block needs to face this direction"*), it was not always possible to accurately perceive the small movements of the cursor on the mobile display for the worker. In addition, many of our workers noted that gaze sharing allowed them to roughly ascertain the target location even before the verbal point of disambiguation.

Our results suggests that even though mouse is faster than gaze, mobile workers find value in gaze sharing. Gaze sharing could be an alternative or complementary to shared mouse pointer for mobile

video collaboration, especially in scenarios where using mouse may not be possible e.g. when hands are occupied or device form factor not supporting mouse or as an additional channel when instructor is not actively using mouse. Further research should also look at novel ways of combining gaze and mouse pointers to effectively support mobile video collaboration.

An important aspect that influences the usefulness of gaze-sharing systems is the accuracy of gaze tracking. D'Angelo et al. [2016] showed that users in desktop-based collaboration overcome the accuracy issues with verbal communication (e.g. *"you mean this one?"*). Mobile video collaboration enables new ways to overcome to inaccuracy in gaze tracking. We observed two such ways. First, workers used physical actions (e.g. touching one of the block) as a feedback mechanism to gather more instruction from the remote partner, similar to previous studies suggesting collaborators with shared visual space use actions as communication cues [Gergle et al., 2004]. Second, workers would move the phone closer to the target area indicated by the gaze cursor. This increased the distances between the potential targets and enabled easier target disambiguation.

Our study has a few limitations. First, an important aspect to consider when generalising our results is that our sample was small (n=12 pairs). A larger sample size might have resulted in more clear statistical difference between the gaze conditions in measures such as task completion times. Second, our participants were new users of gaze-augmented video communications. More experienced users might be able to utilize the gaze data better. However, exploring this requires a new longitudinal study. Third, the *Gaze_Unaware* condition was disguised as the No pointer condition to the instructor. However, it is possible that some of the actions of the workers may have implicitly communicated the visibility of the gaze (e.g. picking the right block communicated by gaze even before verbal instructions). It is also possible that the workers deliberately did not utilised gaze (e.g. by waiting for the verbal instructions when the correct location of blocks are already available through gaze) in order to avoid implicitly communicating the awareness of gaze. Fourth, our study only focused on remote guidance during collaborative physical task. It is possible that in other remote collaborative scenarios, such as collaborative learning (e.g. [Schneider and Pea, 2013]), the effect of the three gaze configurations may be different.

6 CONCLUSION

Based on our results we can see that the instructors preferred the mouse because of its better support for giving procedural instructions. However, workers also find value in knowing the gaze of the instructor. If gaze is used, it is best to make sure that the instructor is aware of the gaze being tracked and transferred because this improves their performance and reduces the need for verbal utterances. The worker may be able to utilize the gaze data even if the instructor is not aware of its existence. However, this is futile as the instructor will take the time to explain everything verbally when he/she is not aware of the gaze being visible to the worker.

REFERENCES

Deepak Akkil and Poika Isokoski. 2016. Accuracy of Interpreting Pointing Gestures in Egocentric View. In *Proceedings of the 2016 ACM International Joint Conference*

- on *Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 262–273. DOI: <http://dx.doi.org/10.1145/2971648.2971687>
- Deepak Akkil and Poika Isokoski. 2018. Comparison of Gaze and Mouse Pointers for Video-based Collaborative Physical Task. *Interacting with Computers (Under Review)* (2018).
- Deepak Akkil, Poika Isokoski, Jari Kangas, Jussi Rantala, and Roope Raisamo. 2014. TraQuMe: A Tool for Measuring the Gaze Tracking Quality. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 327–330. DOI: <http://dx.doi.org/10.1145/2578153.2578192>
- Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1151–1158. DOI: <http://dx.doi.org/10.1145/2851581.2892459>
- Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (2008), 1465–1477. DOI: <http://dx.doi.org/10.1016/j.cognition.2007.05.012>
- Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on Shared Experiences: Moving Beyond the Camera in Video Communication. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 96–105. DOI: <http://dx.doi.org/10.1145/2317956.2317973>
- Sarah D'Angelo and Andrew Begel. 2017. Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6245–6290. DOI: <http://dx.doi.org/10.1145/3025453.3025573>
- Sarah D'Angelo and Darren Gergle. 2016. Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2492–2496. DOI: <http://dx.doi.org/10.1145/2858036.2858499>
- Andrew T. Duchowski, Nathan Cournia, Brian Cumming, Daniel McCallum, Anand Gramopadhye, Joel Greenstein, Sajay Sadasivan, and Richard A. Tyrrell. 2004. Visual Deictic Reference in a Collaborative Virtual Environment. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA '04)*. ACM, New York, NY, USA, 35–40. DOI: <http://dx.doi.org/10.1145/968363.968369>
- Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, New York, NY, USA, 21–30. DOI: <http://dx.doi.org/10.1145/358916.358947>
- Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Hum.-Comput. Interact.* 19, 3 (Sept. 2004), 273–309. DOI: http://dx.doi.org/10.1207/s15327051hci1903_3
- Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2004. Action As Language in a Shared Visual Space. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 487–496. DOI: <http://dx.doi.org/10.1145/1031607.1031687>
- K. Gupta, G. A. Lee, and M. Billinghurst. 2016. Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 22, 11 (Nov 2016), 2413–2422. DOI: <http://dx.doi.org/10.1109/TVCG.2016.2593778>
- Keita Higuchi, Ryo Yonetani, and Yoichi Sato. 2016. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5180–5190. DOI: <http://dx.doi.org/10.1145/2858036.2858438>
- Sture Holm. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 2 (1979), 65–70. <http://www.jstor.org/stable/4615733>
- Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 957–966. DOI: <http://dx.doi.org/10.1145/2702123.2702345>
- David Kirk and Danae Stanton Fraser. 2006. Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1191–1200. DOI: <http://dx.doi.org/10.1145/1124772.1124951>
- Michael Lankes, Daniel Rammer, and Bernhard Maurer. 2017. Eye Contact: Gaze as a Connector Between Spectators and Players in Online Games. In *Entertainment Computing – ICEC 2017*, Nagisa Munekata, Itsuki Kunita, and Junichi Hoshino (Eds.). Springer International Publishing, Cham, 310–321.
- Changsong Liu, Dianna L. Kay, and Joyce Y. Chai. 2011. Awareness of Partner's Eye Gaze in Situated Referential Grounding: An Empirical Study. In *Proceedings of Eye Gaze on Intelligent Human-Machine Interaction*. 44–43.
- Bernhard Maurer, Ilhan Aslan, Martin Wuchse, Katja Neureiter, and Manfred Tscheligi. 2015. Gaze-Based Onlooker Integration: Exploring the In-Between of Active Player and Passive Spectator in Co-Located Gaming. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. ACM, New York, NY, USA, 163–173. DOI: <http://dx.doi.org/10.1145/2793107.2793126>
- Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. 2016. EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game. In *Entertainment Computing – ICEC 2016*, Günter Wallner, Simone Kriglstein, Helmut Hlavacs, Rainer Malaka, Artur Lugmayr, and Hyun-Seung Yang (Eds.). Springer International Publishing, Cham, 102–114.
- Bernhard Maurer, Sandra Trösterer, Magdalena Gärtner, Martin Wuchse, Axel Baumgartner, Alexander Meschtscherjakov, David Wilfinger, and Manfred Tscheligi. 2014. Shared Gaze in the Car: Towards a Better Driver-Passenger Collaboration. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14)*. ACM, New York, NY, USA, 1–6. DOI: <http://dx.doi.org/10.1145/2667239.2667274>
- Romy Muller, Jens R. Helmert, Sebastian Pannasch, and Boris M. Velichkovsky. 2013. Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *Quarterly Journal of Experimental Psychology* 66, 7 (2013), 1302–1316. DOI: <http://dx.doi.org/10.1080/17470218.2012.737813> PMID: 23140500
- Bonnie A. Nardi, Heinrich Schwarz, Allan Kuchinsky, Robert Leichner, Steve Whittaker, and Robert Schlabassi. 1993. Turning Away from Talking Heads: The Use of Video-assata in Neurosurgery. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 327–334. DOI: <http://dx.doi.org/10.1145/169059.169261>
- Kenton O'Hara, Alison Black, and Matthew Lipson. 2006. Everyday Practices with Mobile Video Telephony. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 871–880. DOI: <http://dx.doi.org/10.1145/1124772.1124900>
- Pernilla Quarfordt, David Beymer, and Shumin Zhai. 2005. RealTourist – A Study of Augmenting Human-Human and Human-Computer Dialogue with Eye-Gaze Overlay. In *Human-Computer Interaction – INTERACT 2005*, Maria Francesca Costabile and Fabio Paternò (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 767–780.
- Bertrand Schneider and Roy Pea. 2013. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning* 8, 4 (01 Dec 2013), 375–397. DOI: <http://dx.doi.org/10.1007/s11412-013-9181-4>
- Randy Stein and Susan E. Brennan. 2004. Another Person's Eye Gaze As a Cue in Solving Programming Problems. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)*. ACM, New York, NY, USA, 9–15. DOI: <http://dx.doi.org/10.1145/1027933.1027936>
- Sandra Trösterer, Martin Wuchse, Christine Döttlinger, Alexander Meschtscherjakov, and Manfred Tscheligi. 2015. Light My Way: Visualizing Shared Gaze in the Car. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 196–203. DOI: <http://dx.doi.org/10.1145/2799250.2799258>
- Boris Velichkovsky. 1995. Communicating Attention: Gaze Position Transfer in Cooperative Problem Solving. *Pragmatics and Cognition* 3, 2 (1995), 199–223.
- Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21, 1 (01 Feb 2017), 173–186. DOI: <http://dx.doi.org/10.1007/s00779-016-0969-x>

Details of the dissertations are available at
<http://www.uta.fi/sis/tauchi/dissertations.html>

1. **Timo Partala:** Affective Information in Human-Computer Interaction
2. **Mika Käki:** Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula:** Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari:** Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi:** Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen:** Software Tutoring in Speech User Interfaces
7. **Harri Siirtola:** Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen:** Face Analysis Techniques for Human-Computer Interaction
9. **Oleg Špakov:** iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications
10. **Yulia Gizatdinova:** Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta:** Text Entry by Eye Gaze
12. **Ying Liu:** Chinese Text Entry with Mobile Phones
13. **Toni Vanhala:** Towards Computer-Assisted Regulation of Emotions
14. **Tomi Heimonen:** Design and Evaluation of User Interfaces for Mobile Web Search
15. **Mirja Ilves:** Human Responses to Machine-Generated Speech with Emotional Content
16. **Outi Tuisku:** Face Interface
17. **Juha Leino:** User Factors in Recommender Systems: Case Studies in e-Commerce, News Recommending, and e-Learning
18. **Joel S. Mtebe:** Acceptance and Use of eLearning Solutions in Higher Education in East Africa
19. **Jussi Rantala:** Spatial Touch in Presenting Information with Mobile Devices
20. **Katri Salminen:** Emotional Responses to Friction-based, Vibrotactile, and Thermal Stimuli
21. **Selina Sharmin:** Eye Movements in Reading of Dynamic On-screen Text in Various Presentation Formats and Contexts
22. **Tuuli Keskinen:** Evaluating the User Experience of Interactive Systems in Challenging Circumstances
23. **Adewunmi Obafemi Ogunbase:** Pedagogical Design and Pedagogical Usability of Web-Based Learning Environments: Comparative Cultural Implications from Africa and Europe
24. **Hannu Korhonen:** Evaluating Playability of Mobile Games with the Expert Review Method
25. **Jani Lylykangas:** Regulating Human Behavior with Vibrotactile Stimulation
26. **Ahmed Farooq:** Developing Technologies to Provide Haptic Feedback for Surface Based Interaction in Mobile Devices
27. **Ville Mäkelä:** Design, Deployment, and Evaluation of Gesture-Controlled Displays in Ubiquitous Environments
28. **Pekka Kallioniemi:** Collaborative Wayfinding in Virtual Environments

29. **Sumita Sharma:** Collaborative Educational Applications for Underserved Children: Experiences from India
30. **Tomi Nukarinen:** Assisting Navigation and Object Selection with Vibrotactile Cues
31. **Deepak Akkil:** Gaze Awareness in Computer Mediated Collaborative Physical Tasks