Barış Serim Adapting Interaction Based on Users' Visual Attention Aalto University 2020

Department of Design

Adapting Interaction Based on Users' Visual Attention

Barış Serim

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 72/2020

Adapting Interaction Based on Users' Visual Attention

Barış Serim

Aalto University
School of Arts, Design and Architecture
Department of Design

Supervising professor
Professor Turkka Keinonen, Aalto University

Thesis advisor
Professor Giulio Jacucci, University of Helsinki

Preliminary examiners
Professor Carl Gutwin, University of Saskatchewan, Canada
Professor Raimund Dachselt, Technische Universität Dresden, Germany

Opponent
Professor Stephan Wensveen, Eindhoven University of Technology, Netherlands

Cover image courtesy of the Getty's Open Content Program.

Aalto University publication series
DOCTORAL DISSERTATIONS 72/2020

ISBN 978-952-60-3870-4 (printed)
ISBN 978-952-60-3873-5 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-3873-5

Unigrafia Oy
Helsinki 2020

Finland

NORDIC SWAN ECOLABEL
Printed matter
4041-0619

Aalto University
Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Barış Serim

Name of the doctoral dissertation
Adapting Interaction Based on Users' Visual Attention

Publisher School of Arts, Design and Architecture

Unit Department of Design

Series Aalto University publication series DOCTORAL DISSERTATIONS 72/2020

Field of research Human–Computer Interaction

Language English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

Successful interaction with many information systems depends on our ability to visually attend to the system feedback as well as to our own actions. However, at a given time, we are able to attend to only a portion of the available information. Among other constraints, what can we attend is limited by the spatial acuity of our eyes. Aware of this limitation, researchers have long pursued interfaces that decrease our dependence on visual attention during interaction. The newly proliferating sensing technologies such as eye and head tracking as well as methods for user modeling provide a novel venue for addressing this limitation: An information system can utilize users' visual attention information to change how it responds to user actions.

This thesis contributes design knowledge about adapting the interaction based on users' level of visual monitoring during input through a series of prototypes that have been developed for different use cases. I first distinguish between different implications of visual attention information for interface design, and identify visual attention as a measure of user awareness as the main focus of the work presented in this thesis. Lack of visual attention during input decreases users' awareness of the environment. In these cases, the system can adapt the interaction through a number of methods such as handling input more flexibly or remediating the lack of visual attention through novel visual feedback techniques. These interaction methods have been formulated as part of a constructive research program and applied to single-user applications that require users to split their visual attention between multiple interface regions during pointing and also to collocated and synchronous multi-user applications. User studies provide evidence for the increased uncertainty during input with low visual attention and also show in which situations these interaction techniques can improve performance. The dissertation discusses these empirical findings in terms of the previously identified trade-offs between time and spatial multiplexing, and between predictability and adaptiveness in interface design. The thesis also makes a theoretical contribution to the general design challenge of building adaptive or context aware systems through an analysis of the concept implicit interaction.

Overall, the thesis contributes to the existing line of work on attentive interfaces by developing interaction methods that specifically target handling user input with low visual attention, and contributes to the ongoing discussions about the integration of eye tracking into human–computer interaction.

Keywords

ISBN (printed) 978-952-60-3870-4 **ISBN (pdf)** 978-952-60-3873-5

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2020

Pages 144 **urn** http://urn.fi/URN:ISBN:978-952-60-3873-5

Preface

The journey that has led to this dissertation has been one of personal and intellectual maturing. This would not have been possible without the presence and support of many others. Here, I would like to express my gratitude to all of them.

An immense amount of work goes into creating a fertile and resourceful research environment. During these years, I have been privileged to be part of my advisor Giulio Jacucci's research group at the University of Helsinki and take part in many interesting research projects that expanded my horizon. Thanks for your continuous encouragement, trust and feedback, as well as for giving me the flexibility to pursue many side research projects. Those side projects have culminated in this dissertation.

My supervisor Turkka Keinonen has been a great help during my time at the Department of Design doctoral school and guided me through the final dissertation process. Hans Gellersen has been part of my journey, first indirectly, through his group's cutting edge research, and later more directly, by hosting me as a visiting researcher at Lancaster University. My pre-examiners, Carl Gutwin and Raimund Dachselt deserve special regard for their thorough reviews and feedback that improved the final text. I am already excited to have Stephan Wensween as an opponent.

I had the chance to collaborate with many brilliant colleagues. I am indebted to Thanh Tung Vuong, Chen He, Michiel Spapé, Luana Micallef, Tuukka Ruotsalo, Ken Pfeuffer, Hassan Abbas, Nishad Aluthge, Patrik Pluchino, Pedram Daee, Iiris Sundin, Luca Chech, Salvatore Andolina, Andrea Bellucci, Valeria Orso, Andrea Vianello, Luciano Gamberini, Benjamin Blankertz and Samuel Kaski. Among my collaborators special thanks go to Oswald Barral and Khalil Klouche, not only for being great colleagues but also for their friendship and company over countless lunch breaks and after-work beers.

My gratitude extends to many who made Helsinki an intellectually stimulating place. Antti Salovaara, Antti Oulasvirta and Sampsa Hyysalo have all been truly inspiring with their dedication to research and good scholarship; I am grateful for their efforts in organizing the local research communities and for their thoughtful feedback. Yi-Ta Hsieh, Kumaripaba Athukorala, Imtiaj Ahmed,

Suzan Bayhan, Ashwin Rao, Matti Nelimarkka, Mikko Kytö, Joanna Bergström, Diogo Cabral and many others at HIIT and HCI communities greatly enriched the research environment with their presence.

This dissertation in its current form would not materialize without the public research grants from the EU and the Academy of Finland, particularly through MindSee and Re:know projects. My smooth functioning during all these years is in debt to the excellent work of the CS Department and Kumpula campus support staff; thanks for making Kumpula a great place to work in.

Finally, I owe a lot to my friends and family for keeping me sane throughout my studies. Thank you Vincent, Sveta and Martina for your loving friendship and many memorable moments over the years. Thank you 'Kitap Kurtları' for being a constant source of joy. Seviliyorsunuz! Not every parent would encourage their son to leave his secure day job and sail into the unknown. I thank my parents Fevziye and Mehmet for this, and for their unconditional love and support.

Helsinki, April 24, 2020,

Barış Serim

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Barış Serim, Giulio Jacucci. Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5789-5800, May 2016.

II Barış Serim, Khalil Klouche, Giulio Jacucci. Gaze-Adaptive Above and On-Surface Interaction. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, 115-127, June 2017.

III Barış Serim, Ken Pfeuffer, Hans Gellersen, Giulio Jacucci. Visual Attention-Based Access: Granting Access Based on Users' Joint Attention on Shared Workspaces. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2, 3, Article No. 133, September 2018.

IV Barış Serim, Giulio Jacucci. Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Paper No. 417, May 2019.

Author's Contribution

Publication I: "Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input"

The research has been conceptualized, prototyped and evaluated by Barış Serim and was then conducted under the supervision of Giulio Jacucci.

Publication II: "Gaze-Adaptive Above and On-Surface Interaction"

The idea of combining above and on-surface sensing with eye tracking has been provided by Barış Serim, and Khalil Klouche participated in the early conception phase. The prototypes have been jointly developed by Barış Serim and Khalil Klouch, who also participated in writing. The evaluation and analysis has been conducted by Barış Serim. Giulio Jacucci provided feedback and supervision during the process.

Publication III: "Visual Attention-Based Access: Granting Access Based on Users' Joint Attention on Shared Workspaces"

The idea of visual-attention based access has been provided by Barış Serim who also wrote the majority of the text. Ken Pfuffer and Hans Gellersen participated in the early conception phase. The implementation and evaluation of the prototypes have been conducted by Barış Serim. All the authors provided feedback during the writing.

Publication IV: "Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research"

The majority of the literature review and writing has been done by Barış Serim, under the supervision of Giulio Jacucci, who provided feedback to the text and referred me to additional literature.

1. Introduction

Designing information systems is partly a problem of configuring which tasks should be executed by computers or humans. An obvious starting point for determining this division of labor between the two parties is the relative differences in capabilities. It is commonly known that computers can perform predefined instructions tirelessly and with precision, but lack the interpretative and ad-hoc decision making capabilities possessed by humans. A parallel problem emerges when designing the interface between the humans and the computer system they will use; the two parties need to communicate, but allocating the communicative work of displaying, sensing and interpreting information is a complex design problem. Among other considerations, the problem is informed by the respective sensing and interpretation capabilities of computers and humans. Computers are limited in terms of what they can sense as input through various sensors, while users are limited by their sensory organs as well as their finite attentional resources. The division of labor is in constant flux due to changing demands and capabilities. The sensing capability of early computers was limited to dedicated input devices such as the keyboard and the mouse, and the system's interpretation of user input was straightforward in the sense that user actions mapped directly to predefined commands. The success of communication in these systems consequently depended on users' ability to accurately model the system's behavior and provide the right inputs. In the last few decades, the sensing capabilities of interactive systems have diversified from these dedicated input devices to a plethora of sensors that gather additional information about the user and environment while the modeling capabilities of computer systems increased.

This thesis investigates how a particular type of information, namely users' visual attention, should be utilized by the system during interaction. Visual perception is a crucial sensory capability for humans. It is thus no surprise that user interfaces are designed to take advantage of this capability by displaying graphical elements. Furthermore, the design of many interfaces assumes the presence of visual perception during input; in order to accurately select an item on a touchscreen, users must move their fingers to the target position in a coordinated fashion, all the while taking on board continuous visual feedback from the environment. When we look at a graphical interface or the environment, our visual field can be populated by many objects and their numerous features. Yet we mentally process only a portion of them at a given time. The selectivity in what we process corresponds to the concept of visual attention. Visual attention is continuously facilitated by various actions such as eye and head movements. In most previous human–computer interaction, the significance of these actions lied in their value of monitoring the system's state and the environment. However, the developing sensing capability of computers makes it increasingly possible to use these actions for purposes other than just visual monitoring. Said actions instead become inputs that provide information about what the users are visually attending to, which in turn leads to changes in how the system behaves. This development brings forth the problem of utilizing visual attention information for interaction. Two lines of research in HCI are particularly relevant to this problem.

First is the research on novel input modalities that provide a more direct measure of visual attention. A particularly notable technology is eye tracking, which makes it possible to sense the user's gaze direction with far greater accuracy than that allowed for by other input devices. Eye tracking has been a means of communication for disabled users who cannot operate manual input devices, but its use as an everyday input modality—long envisioned by researchers [9, 55, 56]—is yet to materialize. However, relatively accurate eye tracking is now available through affordable equipment, which means that there is an opportunity for the general population to benefit from it. The emergence of eye tracking and other sensing technologies raises the question of how various tasks should be divided between different human capabilities (e.g. moving hands, moving eyes) and, by extension, different input methods.

Second is the long line of research on attentive interfaces [49, 96, 119] which aims to adapt the interaction based on users' attention information. The departure point of attentive interfaces is that human attention is a limited resource and interfaces should keep track of this limited resource to determine users' workload and environmental awareness [49]. The attention information, in turn, can be used to decide when to interrupt users [49] or how to compensate users' temporary attentional disconnect through visual changes on the interface [21, 41]. A separate but related line of research addresses the option of delegating control to the system when the users are considered to be less capable of control and decision making [32, 88, 122]. Common among this work is the treatment of users' capabilities as a contextual phenomenon that is partly determined by their ability to monitor the environment. Attentive interfaces thus raise the question of configuring the division of labor between the user and the system based on users' contextually changing capabilities.

The two lines of research also highlight different and partly conflicting considerations related to visual attention. On the one hand, eye tracking and other novel input modalities promise performance improvements over existing interaction methods [e.g., 55, 104, 129]. On the other hand, human visual attention, itself constrained by the spatial acuity of the eyes, poses its own limitation for human performance. How these different considerations can be reconciled for interface design remains an open question.

1.1 The Scope and Contributions of the Thesis

This thesis aims to address the challenge of designing for two considerations, namely utilizing users' gaze information and designing for limited visual attention. I frame this in terms of a constructive research program of adapting interaction to users' level of visual monitoring during input. Said program has been realized through a series of prototypes that were designed for different use cases. The common focus of the research program and its main contributions can be summarized in the following questions:

1.1.1 What are the implications of limited visual attention for interface design?

The thesis formulates the research program of designing for limited visual attention as a combination of different considerations related to visual attention. I make distinctions between visual attention information as a 1) measure of what users prioritize in terms of the elements they are monitoring in the environment, 2) what they have already monitored and 3) what they aim to signal to the system. The first two considerations of visual attention information relate to the cognitive role of visual attention in monitoring the environment, while the third interpretation is communicative and intimately related to the way a system utilizes visual attention information. I show how different design approaches that utilize visual attention information can be expressed in terms of these different considerations. The emergence of visual attention information as an input makes the third consideration ever more relevant, although a design approach in HCI has been to ground this communicative aspect of users' visual attention actions on their monitoring function, leading to expectations of implicitness. The publication IV provides an operational definition of implicit interaction, which I use to express the problem of utilizing visual attention information for interaction.

1.1.2 How can an interactive system adapt interaction based on users' level of visual attention?

The main focus and contribution of the thesis is constructive and includes various input handling and visual feedback methods that compensate for users' lack of visual attention. These techniques have been prototyped for various applications that operationalize visual attention through different instruments (e.g. eye and head orientation tracking). The application cases were:

Pointing on touchscreens: Pointing, that is the selection of an interface item through positional input, is one of the most common actions in HCI. For graphical user interfaces, pointing is often performed through mouse cursors or finger contact on touchscreens. At the same time, unlike keyboards or other input devices that provide tactile cues, accurate input using a mouse or touchscreen requires users to monitor their hand movements. This requirement of visual monitoring is a bottleneck in situations where users need to divide their attention between multiple interface regions. The publications I and II aimed to support interaction in these cases by adjusting the system's handling of user inputs based on their degree of visual monitoring. If a user had performed an input while looking elsewhere on the interface, the system handled the input as positionally inaccurate and relied on other inputs to resolve uncertainty. Another interaction technique involved adjusting the position and the size of the visual feedback based on the user's gaze direction. Both publications used eye trackers that provide fine-grained data pertaining to visual attention.

Collaborative work on shared screens: Many interaction tasks benefit from concurrent input of multiple users in shared workspaces. Yet concurrent input also introduces the challenge of maintaining coordination between multiple users, in particular ensuring that the work done by individual users is relevant to the joint activity and that individual users' actions do not interrupt those of others. Because human attention is limited, the public availability of information does not guarantee individual users' awareness of others' actions. A potential opportunity for design is to track users' locus of visual attention and adapt the interaction accordingly. The publication III investigated adapting the access rights based on how multiple users visually attend to the interface and each other's actions on a shared display.

1.1.3 How do low visual attention and interaction techniques that adapt to users' visual attention affect performance and user experience?

The empirical contribution of the thesis is the data gathered through different studies that evaluate the performance and user experience of the interaction methods developed for various applications. The empirical questions that guided the individual publications were:

- RQ1.1: How is touch accuracy affected by decreased visual monitoring?
- RQ1.2: What are the particular considerations for touch input without visual monitoring?
- RQ2.1: How does the performance of a gaze-aware interaction technique compare with traditional input for acquisition and manipulation tasks?
- RQ3.1: What are the visual attention-based access preferences for different

actions?

- RQ3.2: What are the motivations for different visual attention-based access preferences?

These questions show the effect of low visual attention and the particular interaction methods that compensate for low visual attention regarding user performance and experience. I discuss these observations in terms of the trade-offs between time and spatial multiplexing and between adaptiveness and predictability. The qualitative analysis of the data gathered through video analysis and interviews points to additional considerations for future system design.

1.2 The Structure of the Thesis

This thesis is based on a number of publications that document my work on using visual attention data to adapt interaction based on users' level of visual monitoring during input. The following chapters provide an overview of the central concepts relevant to the research, while also describing the overall research strategy and situating the interaction techniques and empirical observations within the larger domain of HCI research. The chapters are organized as follows:

Chapter 2 introduces the concept of visual attention through an overview of related work on human cognition and interpersonal communication. A main distinction is made between visual attention as a cognitive measure of monitoring the environment and visual attention as a communicative signal that is available to other entities in the environment. Importantly, various conceptions of visual attention point to different considerations for the use of visual attention information for interaction.

Chapter 3 frames the thesis in terms of a constructive research program of *adapting interaction to users' level of visual monitoring during input* and describes how individual publications instantiated this program in different ways. The differences concern the unit of visual attention (single or multiple users), how the visual monitoring is operationalized, and the particular input handling and visual feedback methods that aim to compensate for users' lack of visual monitoring.

The prototypes that have been developed as part of the constructive research program have been evaluated through different user studies to answer various empirical research questions. Chapter 4 summarizes the studies' designs and their main results.

Chapter 5 discusses the observations in the context of more general concepts in HCI research, namely the trade-off between time and spatial multiplexing and the uncertainty introduced by adaptive interfaces. The chapter summarizes the main contributions of the thesis and provides an outlook for future work.

2. Background: Visual Attention as Cognition and Communication

This chapter introduces the constructs of attention and visual attention through an overview of previous work on human cognition and interpersonal communication. Importantly, different conceptions of attention—and by extension visual attention—bring forth different insights, which ultimately point to different considerations for interactive system design. I make one major distinction between two different interpretations of actions that facilitate visual attention, first as a cognitive measure of what is being monitored in the environment, and second as communicative signals that are available to and utilized by other agents. This distinction is important as the emergence of sensing devices such as eye tracking as input methods entails a partial shift from cognitive to communicative considerations.

2.1 Visual Attention and Cognition

2.1.1 Attention as a Psychological Construct

The psychological treatment of the concept of attention can be traced to as early as the nineteenth century when James described it as "taking possession by the mind...withdrawal from some things in order to deal effectively with others" [57, p. 403]. The emergence of attention as a mature area of study, however, coincides with the development of information-processing models in cognitive psychology [80, p.179] that reframed diverse phenomena such as memories, thoughts or sensory experience under the unifying framework of information. Within this framework, attention emerged as a necessary construct to explain the discrepancy between information available to the cognitive system and what can actually be processed [1, 58, 73]: At a given time, the mind is tasked with processing information of internal (e.g., memories, thought) or external origin (e.g. visual, auditory stimuli) [17], but is limited by its processing capacity. A direct consequence of this limitation is the selectivity regarding the processed information [1, 73, 79]. Attention has been conceptualized either as the cause of this selectiveness or as the by-product of different priming events or inputs that compete for representation and processing in the brain [19, 25, 58]. Yet central to all conceptions is the ability to prioritize certain stimuli over others. The functional outcome of this prioritization is roughly summarized in Lindsay and Norman's characterization of attention as a "two-edged sword" [73, p.356]. On one hand, attention involves focusing on information that is of immediate interest and filtering out competition and interferences. On the other hand, selectivity of attention results in limitations to what can be tracked at a given time, leading to potential omissions of useful information.

It should be noted that the precise application of information processing models to human beings has been non-trivial[1] and diverse interpretations led to different conceptions of attention. While the processing capacity or the communication bandwidth of machines can be specified through design, experimental psychologists' knowledge on the limits of human processing capacity relies on the performance data observed through behavioral measures such as reaction time, response accuracy or memory reports. Experiments generally operationalize attention by observing the extent information can be accurately memorized or reacted upon by participants after they have been exposed to multiple audio streams [e.g. 10] or crowded visual fields [e.g. 19, 58]. The particular information that can be accurately reported during these attentional "overload" situations provides evidence for a limitation posed by the information processing capacity. Early work such as Broadbent's filter theory [10] characterized attention as a single channel bottleneck that filters different inputs early on during the cognitive process based on their task relevance. Distributing information to different streams (such as spatially distributing audio through multiple speakers or utilizing multiple sensory modalities) is deemed helpful for filtering out irrelevant information, but the final processing is conceived as a many-to-one convergence [81]. As such, information processing is a time resource that can not be shared between different tasks: *"When no material is to be discarded there is comparatively little advantage in using two or more sensory channels for presenting information"* [10, p.34].

Later work put the non-shareable single channel bottleneck model into question by observing the concurrent accomplishment of multiple tasks with different attentional demands [81, 17, 121]. The limitations of human processing and attentional selectivity predicts an upper limit to performance, which can be observed in the trade-offs between the attentional demand of the primary task and the performance of the secondary task [121]. Importantly, this performance limit is not static but sensitive to a person's skill and the type of stimuli. For example,

[1] In fact, the definitions of human information processing capability do not necessarily mirror information theory as formulated by Shannon. Luce observes: *"Of course, the word information has been almost seamlessly transformed into the concept of 'information-processing models' in which information theory per se plays no role. The idea of the mind being an information-processing network with capacity limitations has stayed with us, but in far more complex ways than pure information theory."* [75, p.185].

multi-tasking performance during simultaneous writing, comprehension and reading can improve with practice [105]. Another observation is the improved multitask performance if two separate tasks are conducted over different sensory channels, such as when providing vehicle drivers audio instead of visual instructions [74], showing a cross-modal advantage for time-sharing between different tasks. Such findings led to a conception of information processing as a collection of multiple and situated resources that can be deployed in parallel and with relative independence [121] instead of the unitary model of attention that is conceived as a single channel bottleneck. The division between resources can be done on the basis of information processing stages, perceptual modalities (e.g., auditory or visual), channels (e.g., foveal or peripheral vision) and codes (e.g., spatial or symbolic information) [121].

2.1.2 Visual Attention

Visual attention accordingly refers to selectivity in processing visual information [85]. As with general attention research, visual attention can be conceptualized as a cause or an effect [25]. One example of the cause conception is to treat visual attention as a spotlight that covers only a portion of a vast visual field [58, 90]. The spotlight metaphor also suggests selectivity based on spatial location within this visual field. Yet the precise entity of visual attention, and thus its limitation, has been the subject of ongoing research. In addition to spatial selection, previous work identified discrimination-based (such as color and shape) and object-based (where attention is limited by the number of separate objects) selection criteria [19, 24].

In addition to different selection criteria, visual selection has been conceptualized as the product of both top-down and bottom-up processes (sometimes referred to as endogenous and exogenous) [19, 85, 112]. The distinction is based on the source of bias that directs attention. Bottom-up control is defined as stimuli-driven, determined by the feature properties in the environment [112]. Empirical support for bottom-up control comes from various saliency models that predict visual attention based on various visual variables such as contrast, movement or color. Top-down control, on the other hand, is defined as goal-driven [112], or more broadly as cognitively biased [19]. Previous research has observed better response accuracy and reaction times if experimental subjects are provided visual cues, providing evidence for top-down control of attention informed by the prior knowledge about visual field [90].

Visual attention and gaze direction

There has been a long line of research that correlates attention with motor behavior, particularly with that of eye movements [93]. The spatial distribution of acuity in the human visual field poses limitations to what can be sensed. The acuity is highest on the foveal region (the central 2° of vision) and gradually decreases further into parafovea (which extends 5° from the center) and peripheral regions [93]. The decreasing acuity means that perception is continuously facilitated by foveal alignment (i.e., eye movements that spatially align the gaze direction with the locus of attention). This is typically done using high velocity movements called "saccades", which are followed by relatively still "fixations".[2] Most experimental work uses screen based stimuli in the study of visual attention (e.g. [48, 90]), where eye movements can solely facilitate foveal alignment. Yet foveal alignment with regions further in the periphery can require head movements. Thus, an alternative categorization is to divide the visual field based on the different types of action required for perception. Sander's distinction between "stationary", "eye" and "head" fields is in this direction [99]. Furthermore, the types of actions needed for gaze alignment can be expanded to include re-orienting one's body posture or moving in the space when objects of interest are distributed in the space.

Visual attention and gaze direction are not intrinsically tied, but it has been suggested that separation is often the result of tightly controlled experimental conditions [90], such as when screen-based stimuli is flashed for a limited amount of time [48]. Observations of more complex scenarios such as reading text concluded that *"there is no appreciable lag between what is being fixated [by eyes] and what is being processed"* [59], leading to the general "eye-mind assumption". Similar observations have been made for motor tasks of manually reaching to targets in experimental studies [6, 92] or in the naturalistic observations of making tea [72] or preparing a snack [45]. An observation from the latter set of studies is the high degree of synchrony between gaze direction and hands, with eye fixations often preceding the handling of an object, providing evidence for the top-down control of visual attention in task-based scenarios [111]. Part of the synchrony is due to the need for visually guiding hand movements when reaching to an object. The need for visual guidance has also been shown in controlled studies that reported decreased accuracy during manual target acquisition for arm movements without visual guidance [8, 106, 11].

Visual attention, manipulation and coordination

The actions described so far, such as eye and head movements or moving in space for gaze alignment, correspond to what has been called "sensor actions" [66]; they involve adapting one's own body, but do not cause any other changes in the environment. However, for many realistic use cases, the repertoire of actions involved in perception can be expanded to include manipulating the environment. For example, perception can require removing obstacles in the visual field or positioning objects to locations that are easier to gaze at. The role of manipulation in visual selection has been observed early in the development process of visual skills [127] when infants manipulate objects to bring them to the centre of their visual field and closer to their eyes. In doing so, they increase

[2] In addition to saccades, previous studies identified pursuit, vergence, and vestibular eye movements, but saccadic eye movements are generally considered more relevant due to their high correlation with stimulus in experimental settings [93].

the available stimuli but also filter out other objects of lesser importance [127]. The environment can be manipulated to ease perception by arranging the objects to ease comparison [65]. Part of the physical arrangements is done to decrease the load on memory such as using hands as placeholders [64]. Such manipulation actions put an agent in a better position for perception by decreasing the visual complexity of the environment, increasing the salience of certain items and bringing related items together.

Various actions for facilitating perception and cognition are partly interchangeable. For example, a person can memorize information in a visual field before looking elsewhere, or alternatively use the visual field as an external memory through continuous visual attention. Similarly, one can visually attend to an object by reorienting gaze through eye and head movements or by physically moving the object to the center of the foveal region using hands. Visual comparison of two items can involve continuous reorienting of one's gaze between them or physically bringing them closer to decrease the amount of eye movements. Which strategy is more economic, that is, whether the savings from eye movements make up for the effort spent in physical manipulation is an open question that depends on the particulars of a given task.

2.2 From Cognition to Communication: Signalling Visual Attention

The work described so far conceptualized various actions (such as eye and head movements) as a means for monitoring the environment. This focus has a practical justification: as we perform our daily activities, many objects that we monitor are not affected by how we monitor them. This means that the significance of various actions that facilitate visual attention can be researched solely through their value for monitoring the environment. As such, the main research interest is cognitive.

At the same time, there are cases in which our actions that facilitate visual monitoring can lead to changes in the behavior of other entities. A paradigmatic example is face-to-face communication between humans, where participants are able to see where the other party looks at. The use of gaze in interpersonal interaction has been studied in social psychology [67] and conversation analysis [97]. Previous studies have shown that humans are able to detect others' gaze direction with remarkable accuracy (within a few degrees of deviation if looking straight to the other person [35]). The availability of gaze information means that interlocutors in a conversation are not passively monitored as is the case with inanimate objects, but have the capacity to sense and adjust their behavior in response to others' gaze.

Thus, the study of human communication early on distinguished between gaze *"as an act of perception by which one interactant can monitor the behaviour of the other, and as an expressive sign and regulatory signal by which he may influence the behaviour of the other"* [63, p. 24]. As an act of perception, gaze supplements auditory information by monitoring the facial expression, posture and locus of attention of others during conversation. At the same time, the availability of gaze to the other party during face-to-face interaction means that it inevitably assumes a number of communicative functions [3, 67, 86]. Since the human face presents a wide array of information, much research has focused on how or whether gaze is oriented to others' faces during conversation. Looking at a speaker signals attentiveness [33, 37, 63] and speakers seem to systematically structure their sentences to secure hearer's gaze [37]. Gaze direction can also specify to whom an utterance is addressed in the presence of multiple hearers [36]. Sustained mutual gaze can communicate intimacy during face-to-face communication [2, 63]. Another domain of study is the orientation of gaze as an indexical reference to the environment [38, 114]. Speakers in a conversation can use their gaze direction to point to the objects in the environment and thus establish common ground.

The fundamental distinction of the signalling function from monitoring is that signalling requires the other party to register one's gaze. At the same time, the aforementioned observations on signalling attentiveness or soliciting attention suggests that communicative uses of gaze, at least partly, rely on the affordances created by its monitoring function, that is, *they require participants themselves to have some understanding of each others' visual attention*. As such, visual attention as a construct is instrumental not only for analyzing interaction from an external perspective but also to the participants themselves who maintain a model of what the other person is attending to at a given time. Within this communicative framework, however, analytic focus shifts from visual attention as an objective mental state (as observed by the researcher) to visual attention as a witnessable property in social interaction.

2.2.1 Conflicts between the Monitoring and Signalling Functions

The monitoring and signalling functions can provide competing explanations for gaze behavior [13]. For example, various observational studies showed that it is common for participants in a conversation to look away at the beginning of an utterance and then reorient their gaze to the hearer towards the end [4, 63]. The change in gaze orientation can partly be explained in cognitive terms. Speakers are assumed to be less dependent on recipient's visual feedback at the beginning of their utterance and might even want to limit the external stimuli to dedicate their attention to planning their utterance [4]. Conversely, they are more likely to monitor the recipient's response at the end and ahead of a planned change in conversational role. In this regard, gaze behavior during conversation suggests a top-down shift in attention driven by the divergent needs for information and the constraints to information processing. At the same time, the public availability of gaze direction means that gaze behavior can also be explained by the additional *communicative intention of making the other party aware of one's own gaze direction*. Kendon's interpretation of gaze behaviour is in this direction:

> In withdrawing his gaze, p is able to concentrate on the organization of the utterance, and at the same time, by looking away he signals his intention to continue to hold the floor, and thereby forestall any attempt at action from his interlocutor. In looking up, which we have seen that he does briefly at phrase endings, and for a longer time at the ends of his utterances, he can at once check on how his interlocutor is responding to what he is saying, and signal to him that he is looking for some response from him. [63, p. 42]

Competing explanations based on cognitive monitoring and communicative signalling functions present a research challenge for understanding gaze behavior. However, it has been noted that interlocutors themselves can be very much confronted with the practical challenge of balancing between these multiple functions of gazing. This led to an early theorization by Argyle that explains gaze behavior as a combination of “avoidance” and “approach” factors [2]. For instance, gaze aversion can be optimal for reducing cognitive load while planning an utterance, but complete gaze aversion might be socially inappropriate during face-to-face conversation. The lack of complete gaze aversion can thus be explained in terms of speakers compromising on cognitive needs in order to fulfill the communicative functions of gaze [5]. Similarly, interlocutors might want to increase their monitoring (and thus their information gain), but are likely to inhibit their gaze to avoid signalling undue intimacy through prolonged mutual gaze [2, 3, 63]. The availability of gaze in face-to-face communication means that different considerations are practically intertwined, but there have been attempts to isolate the two in experimental settings. Argyle et al. utilized a one-way screen that allowed one of the participants to monitor the other without his or her gaze being registered [3] (thus eliminating the communicational function of gaze for one of the participants). In line with the expectations, the one way screen resulted in less inhibition by the participant that is not seen by the other.

2.3 Parallels in HCI

The section so far introduced the concepts of attention and visual attention through a brief trajectory of the concepts in psychology and pragmatics, but without going into the specifics of HCI. Yet the parallels to HCI should by now be obvious for many readers. Conception of human–computer interaction as a coupling of two information processors is pervasive in HCI (e.g. [16]). However, as with experimental psychology, modeling human beings in terms of sensory, cognitive and motor bandwidths has been non-trivial [103] and relies on data observed during performance. A classical example is work on pointing performance using Fitts’ law, in which information capacity of the human motor system is simplified into a single channel bottleneck that is inferred from the trade-off between speed and accuracy [28, 76]. One group of HCI innovations such as semantic pointing [7] or bubble cursor [39] essentially aim to make better use of this limited capacity by exploiting information redundancies. As such, they work within the boundaries of the single channel bottleneck.

Another group of innovations can be characterized as aiming to expand the information capacity rather than working within the boundaries of the single channel capacity. Work on multimodal interfaces assumes a multiple resource model of human processing and facilitates concurrent use of audio, haptic and visual channels for increased performance and robustness. Ambient, tangible and graspable interfaces aim to shift interaction from focal visual channel to haptic and visual peripheral channels [30, 54, 91]. A general insight from this line of work is the dependence of the final information capacity (as inferred through performance data) on the particular interface employed, as observed in the relative advantage of bimanual interfaces over single-pointers for certain tasks [15, 60].

The limitations posed by visual attention are broadly relevant for the design of any interactive system due to the significance of visual attention for monitoring the interface and the environment. In this sense, the main interest has been cognitive. One exception to this is multi-user interactions, where users' visual attention assumes communicative functions, as documented early on in shared control rooms [46]. Yet such communicative uses in collaborative work mainly concern human–human interactions that occur in parallel with human–computer interactions. The communicative use of visual attention information by an interactive system is rather a later development; unlike interpersonal communication, where interlocutors' head orientation and gaze is often mutually available, human–computer interface historically developed as a one-way screen. The user can monitor the visual feedback shown by the interface, but the user's head and eye movements are beyond the sensing capability of the system, which rules out their use as communicative signals. In this regard, one-way screen describes an interface quality, namely the inability of the system to sense the user's gaze direction. However, it can also be regarded as a quality of how the interface is articulated by the designer, as the *absence of an explicit and continuously updated model of the user's visual attention*.

Various developments are currently contributing to the fall of this one-way screen. They can be viewed under two different approaches, first through developments in user modeling, which led to the emergence of visual attention as a construct that informs system behavior not only during the design phase but also *during the interaction*, and second by sensing information that more closely corresponds to visual attention. The first is being achieved by inferring visual attention from other sources, most notably through existing manual inputs. The second is being achieved through an increase in the system's sensing capabilities, notably by eye and head tracking.

2.3.1 Inferring Visual Attention from Manual Input

The chapter early on noted that visual attention can involve manipulating the environment, such as when certain objects are made more salient by bringing them to the center of the visual field. Similar behaviors can be observed during various manipulation actions in HCI. Actions that change the visual layout of the interface, such as keeping documents on-screen, scrolling and zooming, make certain objects visible and more salient while hiding others.

In most interfaces, these actions are executed through manual actions, which can alternatively function as a record of what has been visually attended to by the user and some previous work in HCI indeed interpreted them as such. For example, a combination of user's scrolling behavior and dwell time (amount of time an interface element is visible on the screen) can function as a proxy for reading behavior [47]. Research in information retrieval provided taxonomies that classified user behavior such as scrolling or opening a document as "examining" [62]. Since accurate positional input on GUIs requires visual attention, mouse movements—among other information such as interface layout—can be used to construct models of user's actual gaze direction [40, 52, 70, 82, 125]. A system can also infer different levels of visual attention based on the type of input. For instance, whether the user performed a command using touch interaction (which requires visual monitoring for accuracy) or a gesture above a screen can indicate different levels of visual attention [88].

2.3.2 Increased Sensing Capabilities

Another development that is relevant for the communicative use of visual attention is the emergence of new sensors that provide a more accurate measure of visual attention. For example, manual input and dwell time (time window during which an interface element is visible on the screen) alone cannot sense whether the user is physically present in front of the screen or not. The shortcomings of inferring visual attention from manual input devices motivated work on using other sensors such as sonar [110] or web cameras [41] to verify user presence. For larger screens, researchers utilized head orientation and face recognition as rough estimates of gaze direction [20, 126, 109].

Perhaps the most remarkable development is the emergence of eye trackers that provide much more detailed data about a user's gaze. Current technical landscape for eye tracking can be described as a plethora of different image-based and electrophysiological sensing technologies [23]. In HCI, the use of eye tracking dates to as early as 1981 when Bolt [9] used gaze to activate content on multiple screens. Over time, eye tracking has been used in tasks as diverse as pointing [e.g., 55, 129] to understand user interests in search interfaces [e.g., 12, 89] and mediate visual attention information between multiple users [117].

A potential use of novel sensors is replacing manual input by gaze actions. For example, many research contributions that use eye tracking for target acquisition aim to decrease the amplitude of motion travelled by hands [e.g., 55, 129, 107]. By replacing manual input by eye tracking, they also decrease the potential of using manual input to infer visual attention.

2.3.3 Competing Functions of Visual Attention in HCI

Inferring visual attention enables using various actions as communicative signals in addition to their monitoring function. At the same time, just like in interpersonal communication, different uses of visual attention-related information can be conflicting. For example, in most human–computer interaction, eye movements are reserved for monitoring and have no communicative function. This creates a division of labor in which eyes are responsible for monitoring (perceiving the system output) while hands are responsible for manipulation (providing input to the system) and tactile feedback. This neat division of labor changes once eye tracking comes into play. The conflict between the monitoring and communicative functions of eye movements has been acknowledged early on in eye tracking research, under the term Midas Touch [55]; a user gazes to a location to gather information, but his or her eye movements inappropriately trigger commands. The problem has originally been observed for selection tasks [55] and motivated the development of various methods that combine eye movements with another input such as a key press or mouse movements as an additional confirmation [69, 108, 129].

The competing functions of gaze also extends to computer mediated communication in multi-user applications. Visualizing players' eye movements in a multiplayer game can lead them to withhold their gaze or intentionally direct it to mislead their opponents about their game strategy [84]. The competing function is not limited to eye tracking input either. For example, it is common for messaging applications to send read receipts to senders if their message is opened by the receiver. In this case, opening a message (which acts as a proxy for visual attention) not only facilitates monitoring (reading the message) but becomes a communicative signal for the other user. Interviews with messaging users has accordingly shown that they can abstain from opening messages to avoid informing the other party of their reading action [50].

2.4 Summary: Different Implications of Visual Attention

Early in the section, I noted the consequence of attention as a "two-edged sword" when it comes to monitoring the environment [73]; attention stands for the information that is of immediate interest to a person but also for the limitation of what can be processed at a given time. Here, I will argue that this two-edged sword characterization of attention also leads to different considerations when it comes to using gaze—or any other input that operationalizes visual attention—for human–computer communication.

First, attention can represent what is of immediate interest to users, which enables inferring what users plan to do at a given time. As such, visual attention is primarily a measure of what users intend to do or, at least, what they might recognize as appropriate system response. This interpretation assumes a top-down model of attentional shift that is driven by task-related factors. One possible use of inferring appropriate system behavior is to decrease the effort required from the user. For example, the information of the particular information items that are being attended by a user during information search can be utilized to infer the user interest and decrease the need for precise queries [12, 62]. Or, users' gaze direction can be used to infer where they might want to point, which in turn can be used to decrease the need for manual motor action as in various methods that employ eye tracking to completely or partly replace mouse or other manually operated input devices [e.g., 55, 87, 104, 107, 116, 129].

Secondly, the selectivity of attention also allows for utilizing visual attention information to infer what users have monitored. As such, visual attention is primarily a measure of what a user is aware of in the environment at a given time. Unlike the previous consideration, visual attention as a measure of awareness is less sensitive to whether the attentional shift occurred in a bottom-up or top-down fashion. Additionally, what is attended to does not necessarily correspond to awareness due to memory decay [95] and changes in the environment. One possible use of visual attention information is thus to adapt the system behavior based on user awareness. In HCI, this relates to the line of research on systems that aim to compensate for the lack of visual attention through notifications [21, 41] or by delegating control to the system [44, 88].

While not necessarily exhaustive of all design considerations, the two major interpretations of visual attention that derive from its monitoring function—as a measure of what the user plans to do and what the user is aware of—are too important to be overlooked.[3] At the same time, the word "measure" can be problematic as it implies some passive measurement without the participation of the user. Yet the very fact that visual attention becomes observable and usable by the system means that users can adapt their behavior by considering how their input is utilized as a signal.

It is thus useful to list another third consideration for the use of visual attention information as a measure of what the user aims to convey. In this case, the focus partly shifts away from visual attention as an objective measure to how visual attention information is interpreted and utilized by the system, and how users adapts their behavior in consideration of this, although one design approach in HCI has been to base this communicative use of visual attention information on its perceptual function, leading to expectations of 'implicitness' [77, 118, 128]. The user is assumed to perform an action for the purpose of mon-

[3] A prior framework for utilizing visual attention information by Vertegaal distinguishes between 1) sensing attention, 2) reasoning about attention, 3) regulating interaction, 4) communicating attention and 5) augmenting attention [120]. These point to different end-goals of attentive systems but can largely be seen as extensions of the two considerations I have outlined.

itoring but the system utilizes this information in ways that are not targeted by users but are beneficial for them. Yet the section also illustrated how the monitoring and communicative functions of visual attention actions can compete during interaction.

The next section will position the contribution of this thesis in relation to these diverse considerations related to visual attention.

3. The Constructive Research Program

This chapter frames the research strategy pursued in this thesis in terms of a constructive research program that is instantiated through a series of prototypes. I first introduce the concept of 'research program' and justify its use for research. I then describe how the individual work within this thesis concretized the research program in different ways.

3.1 Research Program

The body of work that constitutes this thesis can be framed within the constructive research program of *adapting interaction to users' level of visual monitoring during input*.

Before going into the details of this description, it is useful to unpack the concept of constructive research program in HCI and justify its relevance for HCI design. In HCI and design, the concept of constructive research program has been proposed to articulate research contributions in a way that openly acknowledges their theoretical and methodological commitments [68, 94]. The concept and this emphasis is in debt to Lakatos' explanation of the progress in science [71]. Lakatos argued that scientific achievements are the result of a series of theories and heuristics for problem solving, shortly a *research program*, instead of isolated theories. Framing research in terms of a program thus aims to make these commitments—which operate in the background of various research questions—explicit.

A lengthy discussion of research programs and Lakatos' philosophy is beyond the scope of this chapter. Yet it is necessary to state that the transposition of research programs from natural and social sciences to design requires some effort due to the constructive orientation of the latter. Design contributions, while building on empirical facts, do not just aim to explain or predict the world but aim to modify it. Here, designers are confronted with the challenge of establishing the scope of their design activity, that is they need to decide on what is available for modification and what is not. Secondly, they need to choose the particular empirical observations that are relevant to design. The decisions concerning the design scope and the empirical observations consequently lead to different design heuristics. Let's consider a well-known HCI example, tangible computing: Tangible computing takes humans' existing familiarity with manipulating the physical world as its departure point in an effort to bridge the so-called divide between the physical and digital worlds [54]. The scope of design accordingly involves configuring interfaces around these existing familiarities instead of radically changing human behavior. The most relevant empirical observations are existing practices of manipulating objects and observations of human manual dexterity. These in turn inform various design interventions (in this case tangible interfaces) that embody a set of design heuristics such as providing a direct correspondence between the physical form and the computational variable [54].

The constructive research programs in HCI also emerged with a pragmatic and hands-on mindset that emphasize quick iterations and prototyping [68] over more formal and theoretical approaches that presuppose careful analysis of an existing situation [e.g., 83]. This pragmatic justification for constructive research programs can be summarized as follows:

First, constructive research inherently contains a tension between the use habits and other factors that inform design, and the design interventions that aim to transform them [14, 98]; when making design interventions researchers build on existing practices, but these very practices can be invalidated by their design interventions. Conversely, the design space can be unnecessarily constrained by existing use practices, device and service contexts. For instance, I made the case that the communicative use of visual attention information can lead to changes in how users behave. A pragmatic implication for constructive research is that detailed models of existing visual attention behavior may not easily inform new design, since a design intervention can invalidate previous knowledge about such behavior. In short, information about an alternative future is sometimes best gained after changing certain material conditions, which makes prototyping part of the knowledge production process [130].

The second justification relates to the observation that changing the material settings can be a more cost-effective method of generating knowledge when compared to predicting future use from existing interface uses. A parallel can be made with visual attention. The previous chapter discussed the observation that physical manipulation of the environment is partly interchangeable with eye movements and thinking, and in some cases, can be a more economical method for perception and cognition. The constructive design research can be interpreted as a mere implication of this insight on methodology; instead of striving to build extensive models of the world through observation and try to predict the utility of future design interventions, researchers can start by prototyping their own alternative reality. This is particularly relevant for ill-defined problems in complex settings that do not easily lend themselves to being exhaustively represented. A methodological consequence is that constructive research programs can be exploratory and qualitative in nature, since many factors that need to be evaluated are not necessarily known in advance.

Having discussed the rationale for constructive research programs, we can try to interpret some existing HCI work in terms of how selectively they use empirical data about visual attention and how they accordingly propose different design heuristics. Let's take the example of interaction techniques that replace part of the manual interaction with eye tracking for selection tasks [e.g., 87, 107, 116, 129]. This design heuristic emphasizes particular insights and empirical knowledge about visual attention. First, a departure point is that eye movements to a target precede manual action and can thus be faster than hand movements for selection. The use of gaze for selection also emphasizes visual attention as a measure of what the user plans to do. Overall, the design program is oriented towards bypassing the bottleneck posed by hand movements through gaze input. Less central to the program is the bottleneck posed by the limited visual attention.

The research program of this thesis, *adapting interaction to users' level of visual monitoring during input*, aims to fill some of the gaps left by the above research program. First of all, the precedence of eye movements to a visual target (as observed in mouse use [51]) is not treated as a pre-given but the result of various design decisions; users need to visually monitor a visual target because GUIs require them to do so. Secondly, the research program emphasizes the observation that performance decreases in the lack of visual monitoring, since users are less aware of the environment. In doing so, I utilize visual attention information primarily as a measure of what the user is aware of in a given situation. The program accordingly aims to address the bottleneck posed by visual attention (instead of the bottleneck posed by hand movements). By focusing on visual attention, it also emphasizes the main detriment to the performance as the limitation posed by the spatial resolution of the visual acuity instead of the cognitive limitation of having to handle multiple unrelated tasks. These considerations call for an alternative set of design heuristics. Identifying this alternative set has been the aim of the work in this thesis.

Finally, it should be noted that pursuing a constructive research program does not imply a lack of evaluative criteria. What makes a program valuable is its capacity to guide new design work that goes beyond the state of the art for various use scenarios. Every instantiation (i.e., the practical work that embodies the commitments of the program) helps identify its useful scope, which might result in modifications to the original formulation [94]. The work within the scope of this thesis also unfolded as a progression of various design interventions that were guided by the program. Below, I describe how the individual contributions in this thesis fit into the research program. An overview of these publications is provided in Table 3.1.

3.1.1 Publication I: Single User On-Surface Input

Publication I contributes a set of interaction techniques that use eye tracking to support touch interaction with decreased reliance on visual guidance. Touch

	Single-user	**Multi-user**
Example non-adaptive solution	Tactile/audio cuing & Static interface configurations	Predetermined division of labor between users
Example cause for decreased monitoring	Split-attention due to the spatial distribution of interface elements	Users work in loosely-coupled manner on distant interface regions
Unit of visual attention	Whether a user is visually attending to the interface	Whether multiple users are jointly attending to the interface
Cause of uncertainty	Spatial inaccuracy	Consensual inaccuracy
Adaptive solution developed for	**Publication I**: multifocus image exploration, exploring relational data and color switching in paint; **Publication II**: object drawing and manipulation and real-time video manipulation	**Publication III**: project planning, brainstorming, document sharing

Table 3.1. Overview of different publications within the research program.

interaction comes with several advantages when compared to some other input devices that have traditionally facilitated input with low visual monitoring (such as mechanical keyboards and other tangibles): Touch screens allow dynamically changing the motor and visual spaces of the input surface depending on the application context. On the other hand, several factors make touchscreen use more dependent on visual monitoring. The lack of tactile cues requires users to visual monitor their manual actions for positional accuracy and dynamically changing input surfaces make it harder to rely on memory. At the same time, the flexibility that comes with touchscreens provide an opportunity to address some of these drawbacks through novel interaction techniques.

The main strategy in this publication was to employ eye tracking to understand the degree of visual guidance that a manual action is accomplished with and adapt the system interpretation and handling of the user input accordingly. Decreased visual attention was treated as an instance of decreased control in interaction (Figure 3.1). To deal with this decreased control, I proposed novel input handling and visual feedback techniques that aimed to compensate for users' lack of visual monitoring and demonstrated their use through three example applications that required interacting with multiple regions on the interface (image exploration, exploring relational data and color switching in paint). Two user studies have been conducted to guide future design. The first part measured the degree of positional accuracy based on the degree of visual attention and determined a selection range around a touch point. The second part reported the perceived utility and the hand-eye coordination challenges that emerge during the interaction with applications. The empirical research questions posed in this publication were:

RQ1.1 How is touch accuracy affected by decreased visual monitoring?

RQ1.2 What are the particular considerations for touch input without visual monitoring?

3.1.2 Publication II: Single User On- and Above-Surface Input

The initial work on adapting the interaction based on users' level of visual monitoring was prototyped for a touch screen, but some limitations became apparent during evaluation when users' level of visual monitoring was wrongly interpreted in some situations. The limitation can be framed as a sensing limitation: Hand movements that lead to a touch can be accompanied by different levels of visual monitoring between the initiation of the movement and the touch event, but this complex information about hand-eye coordination is not available through touch sensing alone. A potential solution is to expand the system's sensing capability to above-surface space to sense hand posture, position and speed. Creating a more accurate model of user's visual monitoring was the departure point for utilizing above-surface sensing, but during prototyping it became obvious that above-surface sensing can be used for novel interaction techniques that facilitate concurrent interaction with multiple interface regions.

Publication II contributes a set of interaction techniques that combine on- and above-surface sensing with eye tracking. Together, above-surface sensing and eye tracking allows understanding how users' hands and gaze are distributed across the interface and adapt the interaction accordingly. As with Publication I, the techniques have been developed for use cases (object drawing and manipulation, and real-time video manipulation) that require interacting with multiple regions on the interface. The performance of the interaction methods have been evaluated for acquisition and manipulation tasks against a baseline condition. The empirical research question posed in this publication was:

RQ2.1 How does the performance of a gaze-aware interaction technique compare with traditional input for acquisition and manipulation tasks?

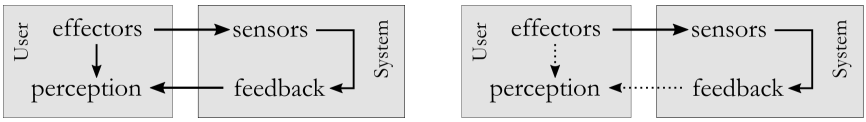


Figure 3.1. The unit of visual monitoring for the publications I and II was the individual human. The interaction with interactive systems often requires users to visually monitor their own actions as well as the system feedback (left). The research program focused on supporting input methods in which the visual monitoring is lower (right).

3.1.3 Publication III: Multi-User Shared Screen Input

Publication III expands the research program to multi-user interaction settings. Publications I and II focused on solitary use cases, in which the interface is manipulated and monitored by the same person. However, some interactive tasks are collaborative and involve multiple users' concurrent input. The coordination can sometimes be accomplished without having to monitor other users' actions—for instance in the presence of established social protocols or predefined divisions of labor. These social protocols are similar to mechanical keyboards in the sense that they allow relying on memory instead of dynamically changing information from the environment. In the absence of such protocols, however, coordination can require participants to monitor each other during collaboration and lack of monitoring can lead to various coordination challenges. A design opportunity to address the challenge of coordination is to adapt the interaction based on multiple users' visual attention.

Similar to publications I and II, publication III treats lack of monitoring as a disruption to the control loop in interaction (Figure 3.2). Yet considerations for limited visual attention in collaborative work differ from that of single-user scenarios. First, the unit of visual attention shifts from the individual monitoring of actions, to the joint attention of multiple users. Secondly, visual attention is limited primarily due to the concurrent input of multiple users, rather than multi-tasking or multi-focus interaction by a single user. Thus, unlike the single-user interactions, the actions that can be attended are not necessarily initiated by the user. Thirdly, adapting the interaction is motivated by avoiding conflicts and maintaining consensus rather than addressing the problem of positional inaccuracy as individual users are assumed to be fully aware of their own actions. In other words, visual monitoring leads to uncertainty about the degree of consensus instead of the spatial position of the input.

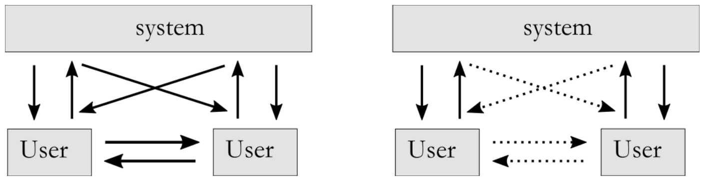


Figure 3.2. The unit of visual monitoring for Publication III was the group level. The publication proposed adapting the system response by distinguishing situations in which the users visually monitor each others' actions (left) or not (right).

Publication III thus investigates input handling techniques based on how multiple users visually attend to the interface and each others' actions on a shared display. During collaboration, users can switch between working on different tasks in parallel to working in tight coordination on the same screen region, leading to different visual attention configurations. In return, actions can require varying degrees of oversight or consensus based on their scope or reversibility. A possible system adaptation is changing the access rights (e.g., who can edit or view a document) based on users' joint attention on a shared display. The framework proposed in the publication presents a framework for visual attention-based access and introduces four different access types based on their availability in solitary and joint attention situations. An exploratory study has been conducted, in which participants were instructed to assign these access types to various actions in three different task scenarios on a large vertical display that tracked their head orientation. The applications (project planning, brainstorming and document sharing) were inspired by existing collocated collaborative scenarios and featured a mix of different action types (such as editing, moving, deleting) and content with varying levels of privacy. Unlike the other publications, the input handling methods were not specified in advance and participants were asked to determine different handling methods for different actions as they perform tasks using these applications. The research questions that guided the evaluation were:

RQ3.1 What are the visual attention-based access preferences for different actions?

RQ3.2 What are the motivations for different visual attention-based access preferences?

3.1.4 Publication IV: Implicit Interaction

I have framed the research program of this thesis as *adapting interaction to users' level of visual monitoring during input*. Treating visual attention information as a measure of their awareness might imply that it is the system that adapts to users and users' participation is somewhat passive in the sense that they do not intentionally target system adaptation. At the same time, the previous section noted that the use of visual attention information as a passive measure becomes problematic once this information is used for communicative purposes as the users can adapt their behavior *in consideration of how the system responds to their input*. One such observation has been made in Publication III, when participants in the study utilized visual attention-based access to direct other users' attention. The mismatch between prior design expectations and actual user practice brings a set of methodological challenges for any system that targets users' unintentional participation.

Publication IV identifies these methodological challenges through an analysis of the concept "implicit interaction" in HCI. The term implicit interaction is often used to describe cases in which user engagement is assumed to be passive. the term has also been used to characterize attentive systems that utilize visual attention information [77, 118, 128]. The publication first reviews the existing meanings of the term implicit and identifies the constructive challenges related to designing for implicit interactions. It then provides a new operational definition of implicit interaction as *user's mental attitude towards an input–effect relationship*.

Input–effect relationships can be used to analyze a diverse set of interfaces and interactions, including interfaces that utilize visual attention information. For example, the communicative use of actions that facilitate visual attention (such as eye movements) can be expressed as situations in which a user action that is sensed by the system (an input) results not only in the monitoring of the interface, but also in additional effects (Figure 3.3). Expectations of implicitness rely on these additional effects being a by-product of a user action; that is, the user has not performed the action in order to achieve this effect.

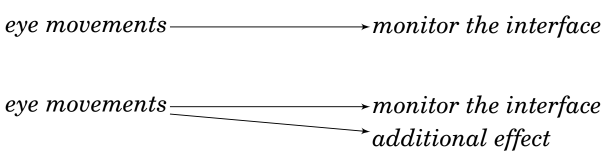


Figure 3.3. The comparison of two cases in which 1) eye movements only facilitate visual monitoring of the interface (above) or their communicative use results in multiple effects (below).

Some applications of eye tracking, such as using eyes as a pointer to trigger commands is usually not considered implicit [77] as the users are assumed to be directing their gaze with the expectation of triggering the commands[1]:

eye movements → *monitor the interface*
eye movements → *trigger a command*

The use of eye movements for communicative purposes can also cause unwanted command triggers as in the Midas Touch problem [55]. A design solution to prevent Midas Touch has been to use additional manual inputs such as mouse or touch confirmation [129, 87], instead of relying on eye movements alone. Since these interactions do not directly trigger a command, their use have been considered implicit [128]. Formally, these are situations in which reaching an effect requires a *complementary* input:

eye movements → *monitor the interface*
eye movements + manual confirmation → *monitor the interface*
eye movements + manual confirmation → *trigger a command*

Note that each interface configuration makes certain action courses easier and others harder. For example, being able to trigger commands by only using eye movements enables users to interact in a hands-free fashion, but it can also prohibit them from monitoring an interface region without triggering a command. Additional manual confirmations remedy the Midas Touch problem, but they *prohibit pointing without visual monitoring*, since the manual input is always used to complement the positional input provided by gaze direction. It can be argued that both design configurations require users to look at an interface location for selection. As such, they aim to address the bottleneck posed by hand movements instead of the bottleneck posed by limited visual attention.

The research program of this thesis, which aims to address the bottleneck posed by limited visual attention, led to the use of visual attention information as a measure of user awareness. The core idea can be illustrated as below:

eye / head movements → *monitor the interface*
eye / head movements + manual action → *monitor the interface*
eye / head movements + manual action → *adaptive input handling*

The rest of the section describes the particular visual attention-based input handling and visual feedback techniques in more detail.

3.2 An Overview of the Interaction Techniques

Below, I provide an overview of how different publications operationalized the visual attention and the interaction methods that have been implemented for different prototypes.

[1] In use cases such as eye typing on a screen-based virtual keyboard, the purpose of eye movements can even be conceptualized as purely communicative, since the user does not aim to gather new information by monitoring the virtual keyboard layout.

3.2.1 Operationalization of Visual Monitoring

Different instruments have been selected for sensing visual attention information depending on the application case. For single user applications (Publication I and II), the input surface was a 10 finger multi-touch screen (27”, 2,560x1,440 pixels) and eye tracking has been used to gather fine-grained data about user’s gaze direction on the interface. Work in Publication I featured an SMI RED eye tracker running at 60Hz and mounted below the touch screen that was approximately 50cm away from users’ eyes. Work in Publication II featured Pupil Labs binocular tracking glasses running at 60Hz (Figure 3.4 left). The multi-user study (Publication III) has been conducted on a larger 2,05 × 1,20 meter vertical interactive surface consisting of three adjacent displays, each with a resolution of 1080 × 1920 pixels. The large size of the display enabled using head orientation as a proxy for visual attention. Head position and orientation of users were tracked by an OpenCV application that detects head-worn markers using a web camera (running at 640 × 480 pixel resolution) mounted at the ceiling (Figure 3.4 right).

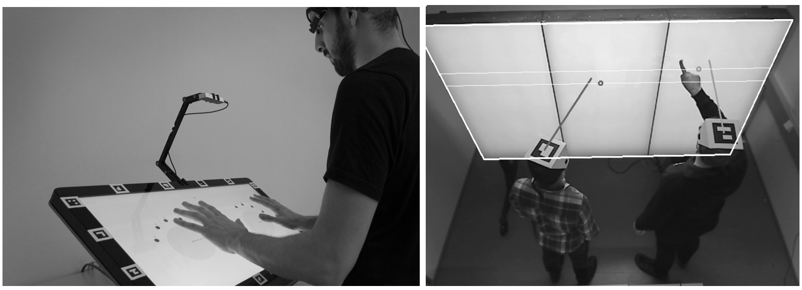


Figure 3.4. Different operationalizations of visual attention, using eye tracking for single users on a 27” touch screen (left, Publication II) and head tracking for multiple users on a wall-sized vertical display (right, Publication III).

In addition to instrumentation, various prototypes differed in regard to how they established whether an action or a system feedback has been visually attended to. For single-user cases, in which the distance between the user and the interactive system is relatively stable (approximately 50cm), distance to the manual input location has been used as the basis for deciding on whether an action is being attended to or not (Figure 3.5 left). On the other hand, multi-user scenarios involve situations in which an action can be viewed from a distance (e.g., if it is performed by the other user). Thus, whether a visual area is attended by a participant has been determined by scoring the visual attention information using visual angle (θ) and distance (d) values between the head and the target on the screen (Figure 3.5 right).

The other considerations for operationalizing visual monitoring were:

- **Continuous, discrete.** Whether an action or a system feedback has been visually monitored can be determined along a discrete (i.e., maintaining a basic

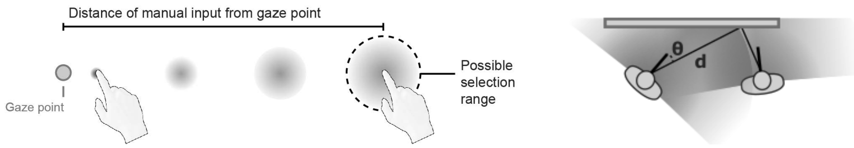


Figure 3.5. For single user touch screen interactions visual monitoring has been operationalized in terms of the distance of the gaze point to the manual input (left, Publication I and II). For users interacting with a wall-size display it has been operationalized in terms of the angle (θ) and distance (d) to the target (right, Publication III).

distinction between visually attended or not) or continuous scale. The choice of the operationalization depends ultimately on how the visual attention information will be utilized by the application. Publication I, for instance, operationalized visual monitoring continuously, based on the distance between the gaze point on the interface and the touch input location. This led to an empirically demonstrated, simple linear relationship between the distance of the gaze to the touch location and the spatial inaccuracy of the touch input. Publication III, on the other hand, utilized visual attention information to discretely distinguish between solitary and joint attention situations (to identify whether an action is attended by a single user or multiple users).

- **Conservative, liberal approaches.** The system's interpretation of users' level of visual monitoring can also update upon eye or head movements (liberal approach) or only upon the movement of the hand (conservative approach). Conservative approach can be more suitable for cases in which the main source of uncertainty is positional inaccuracy, since it is expected that the approximate location of the user's fingers on the surface will persist within the user's short term memory even after the gaze shifts to another location. This approach has accordingly been used in Publication I and II. For single-user cases, the general principle has also been to a) decrease the uncertainty *instantly* when the user increases visual guidance and b) increase the uncertainty *gradually* when the user decreases the visual guidance. The difference is due to the gradual deviation in position with increasing amplitude of movement [8]. The persistence of individual memory does not equally apply to multi-user scenarios. Thus, Publication III updated the visual attention model of the user groups instantly based on their head movements.

3.2.2 Input Handling Techniques

Once the users' degree of visual monitoring has been identified, there remains the question of what type of adaptations can be conducted by the system to compensate for decreased monitoring. The general design approach has been to interpret situations of decreased visual monitoring as cases of uncertain input.

The phrase *uncertain input* stands for an approach to input handling in which the system response to a user input is probabilistically determined through an evaluation of multiple potential interaction outcomes. This is in contrast to many traditional interfaces that abstract user inputs early on into discrete events (such as a mouse click on a Cartesian coordinate or a specific keyboard press), which are then mapped to various user interface commands. As such, the success of the interaction relies on users' ability to provide accurate input. For many input methods such as speech recognition, gestures, touch or physiological sensors, however, the system sensing can be inaccurate or users' situational awareness or capability to provide precise input can be low. A possible design solution is to handle inputs probabilistically by taking various contextual factors into account, instead of immediately abstracting it into discrete events. Many uncertain input handling frameworks in HCI [e.g., 78, 88, 101, 102, 123] follow this approach.

The departure point of the input handling framework in this thesis is that decreased visual monitoring decreases the capacity of users to control interaction. In single-user scenarios, this involves decreased positional accuracy due to the lack of visual monitoring of the action. In multi-user scenarios, this is related to the decreased capacity of users to keep track of the changes on a shared workspace and intervene when another user performs a conflicting action. Below, I describe the various interface adaptations in terms of an input handling framework. This is based on previous work [78, 101] that separates input handling process into successive stages of input modelling and action execution.

Input Modeling

Having identified input with decreased visual monitoring as uncertain, it becomes necessary to identify other input sources that can be utilized by the system to resolve uncertainty. The thesis investigated several inputs for different use cases:

- **Gaze context.** The main use of the visual attention information in this thesis has been to determine the level of visual guidance. Yet visual attention information can also be used to detect task context and resolve uncertainty by prioritizing actions that are related to where the user is visually attending to.

- **Interaction history.** A possible reason for the lack of visual monitoring could be that the user already has some information about the target interface action due to past experience [31]. Thus, decreased visual monitoring can be attributed to the user expectation of repeating a previous action. In this case, users' history of past actions can be an additional input source to resolve uncertainty.

- **Hand gesture.** Various interface actions such as tapping, sliding or rotating can require different finger manipulations and thus different hand postures during manual input. Thus, another potential resource for resolving uncertainty is to supplement the positional information (i.e., where the hands are situated on the interface) with gestural information (i.e., the hand posture and the specific finger that performs the touch).

Action execution

Having compiled different user inputs, the system can proceed to choose an appropriate response. Here, different responses can involve 1) immediate selection of an action, 2) deferring the system response until more information is gathered and 3) inaction.

- **Select action.** One way to handle low visual monitoring is to delegate control to the system. The system can respond to uncertainty in a number of ways for selecting action. Publication I demonstrated various techniques under this category. For example, action selection can involve different actions that are positionally different, such as selecting between different discrete input fields like buttons.

  The selection can also occur between different actions that positionally overlap. For example, a touch action on a text field can be intended for scrolling or text selection [102]. Yet these different actions require different degrees of visual guidance: scrolling has an area effect and does not require exact pointing, while selection requires accurate pointing. Selection among overlapping input fields can also be based on positional and gestural data, as these two components of hand motion are dissimilarly affected by low visual monitoring. Hand posture and relative finger positions are known to the user through proprioception, whereas positional accuracy requires the user to monitor where the hand or finger is located relative to the target. Accordingly, the system can choose the extent it relies on the positional or the gestural component of hand motion based on a user's degree of visual monitoring.

  Finally, if the input field allows range selection, positional uncertainty can be handled by expanding the selection range.

- **Defer action.** Another potential response is to defer action until enough information is gathered for disambiguation. A common example is the press-release sequence for inherently uncertain inputs such as touch [101] or gaze [69]. Publication I utilized this technique by communicating the selected action back to the user as visual feedback upon a touch gesture and deferring the final action execution to a touch release event. Publication II utilized this approach by communicating the widget selection before touch, by taking advantage of the above-surface sensing, and deferring the action execution to the actual touch event.

- **Inaction.** Input without visual guidance can be interpreted as unintentional or unfocused, resulting in the system not taking any action. This approach has been utilized in Publication III to manage access rights on a shared surface. For instance, consensual actions are enabled only if all the users are visually attending to the action, while supervised actions require the attention of a specific user such as the owner of a document. Table 3.2 summarizes the availability of each access type under different attention situations.

	Action can be accomplished	SA	JA
Universal	under any attention situation	●	●
Consensual	only under joint attention	-	●
Supervised	if object owner or supervisor is attending	◐	●
Private	only if the owner is attending and no one else	◐	-

Table 3.2. Types of actions that are available (●), unavailable (-) or only available to a particular user (◐) under solitary attention (SA) and joint attention (JA) situations.

3.2.3 Visual Feedback Techniques

Another way of dealing with uncertainty is to remedy users' lack of visual monitoring through various visual feedback techniques. The two techniques contributed within the scope of this thesis are supporting peripheral awareness and warping information.

- **Support peripheral awareness.** Perception in the periphery of the visual field benefits from larger object sizes [18] as acuity in the peripheral field is lower than that in fovea. A potential system adaptation is thus adjusting the visual feedback size based on the distance of gaze to the target object. The system can increase the visual footprint of the cursor peripheral awareness and to indicate the degree of positional uncertainty as determined by the system. This technique has been utilized in Publication I (Figure 3.6).

- **Warp information** Previous section noted that visual attention is facilitated not only by eye and head movements but can also involve the manipulation of the environment. For instance, instead of redirecting visual attention, a target item can be moved to the center of the visual field through hand movements. A parallel approach developed in this thesis is to overlay the information content near a manual input location to where the user's visual attention is directed to. Publication I and II utilized this technique by showing widget information upon touch (Publication I, Figure 3.6 left) or above-surface hover (Publication II, Figure 3.7 right) to the user.

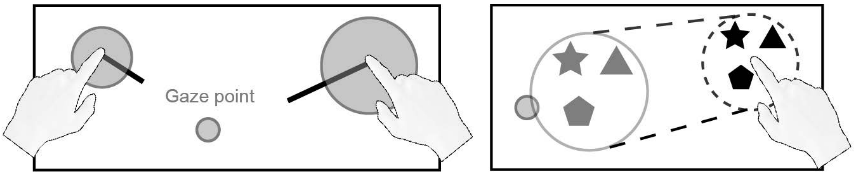


Figure 3.6. Providing peripheral awareness (left) and warping information content around manual input position to gaze point (middle) are two possible visual feedback techniques to communicate system interpretation of user input back to user (Publication I).

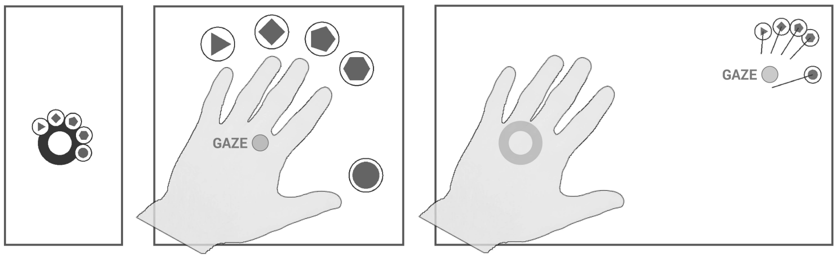


Figure 3.7. An example of warping information: the system adapts where to show visual feedback about a widget based on user's gaze direction (Publication II).

3.3 Summary

The section introduced the concept of constructive research program and framed the work within this thesis in these terms. The constructive program I pursued mainly departs from the consideration of visual attention as a limited resource, which I have contrasted with other approaches that depart from limitations to motor performance. This consideration consequently led to different design solutions than proposed in previous work. Table 3.3 provides an overview of potential input handling and visual feedback techniques and the publications in which they are implemented. Publications I, II and III document the progression of the constructive program from single user situations to multiple users, while publication IV provides a theoretical reflection on the concept of implicitness that often feature in adaptive and attentive systems.

	Technique	Unit	Description
Input Handling	Select action	Single-user	Delegate control to the system when the user is not paying attention (I&II)
		Multi-user	Delegate control to the system when the users are not jointly attending to the action
	Defer action	Single-user	Postpone the execution of an action until the user visually monitors the feedback (I&II)
		Multi-user	Postpone the execution of an action until the other user visually monitors the feedback
	Inaction	Single-user	Do not execute an action if a user is conducting it without visual monitoring
		Multi-user	Do not execute an action if the action is not visually monitored by certain users (III)
Visual Feedback	Warp information	Single-user	Show information near where the user's gaze is directed at (I&II)
		Multi-user	Show information near where the other user's gaze is directed at
	Increase peripheral awareness	Single-user	Increase the visual footprint of an item that is on the periphery of user's visual field (I)
		Multi-user	Increase the visual footprint of an item that is on the periphery of the other users' visual fields

Table 3.3. An overview of various input handling and visual feedback techniques for single- and multi-user adaptations. The roman numerals in parentheses denote the publications that have implemented the technique.

4. Empirical Observations

This section provides an overview of the empirical studies conducted as part of the research program. The studies were conducted to evaluate the interaction techniques described in the previous section based on performance and user experience and are mainly formative as they aim to identify further considerations for design.

	Research Question	Approach	Data
I	RQ1.1: How is touch accuracy affected by decreased visual monitoring?	Explanatory	Data logging
I	RQ1.2: What are the particular considerations for touch input without visual monitoring?	Exploratory	Experimenter Observations, Focus Interviews
II	RQ2.1: How does the performance of a gaze-aware interaction technique compare with traditional input for acquisition and manipulation tasks?	Explanatory	Data logging
III	RQ3.1: What are the visual attention-based access preferences for different actions?	Descriptive	Data logging
III	RQ3.2: What are the motivations for different visual attention-based access preferences?	Exploratory	Experimenter Observations, Focus Interviews

Table 4.1. Overview of different research questions posed throughout different publications (in roman numerals) and the empirical approach and data gathering methods employed for answering them.

A methodological problem for evaluating HCI prototypes is the discrepancy between the current world and a potential future world that is envisioned by a design intervention [98]: The current world might differ in terms of user expectations from interactive systems and available devices. In the context of this research, for instance, a major limitation is the absence of dedicated equipment for sensing visual attention in most systems and people's existing visual attention habits. The discrepancy makes a level of control necessary to recreate the future conditions envisioned by the research. A common way of control is through laboratory studies, which constitute the body of empirical work in this thesis.

Despite the shared laboratory setting, the questions posed throughout the thesis have been approached in different ways (Table 4.1). One way to categorize different empirical studies is based on their degree of open-endedness, or put inversely, how structured they are [61]. Exploratory studies are open-ended as they aim to learn more about a phenomenon and identify considerations that are not anticipated in advance. Descriptive and explanatory studies, on the other hand, aim to document and predict the phenomena under investigation through different variables that are often defined in advance of the study [61]. The research questions 1.2 and 3.2 are thus explorative as they aim to identify different considerations that were observed after deploying the prototypes. The research questions 1.1 and 2.1 on the other hand are explanatory with predefined invariables and variables. Finally, the research question 3.1 is descriptive as it catalogues participant responses into pre-established categories, but without strong prior predictions.

In the rest of the chapter, I describe individual study designs and summarize their main results.

4.1 RQ1.1: How is touch accuracy affected by decreased visual monitoring?

Publication I proposed interaction technique to compensate for users' low visual monitoring during manual input based on the insight that low visual monitoring decreases positional accuracy (which is also observed in previous research [106, 124]). Yet the extent of inaccuracy for touch input surfaces that accommodate bimanual interaction (more particularly the 27" tilted touch screen used in the study) was not established in previous research. Thus a two-part study has been devised with the aim of finding 1) the positional accuracy of touch input with varying degrees of visual guidance and 2) the distance of the gaze point to the touch point for positionally accurate tasks.

The first part of the study treated the degree of visual monitoring as the invariable and the positional accuracy as the variable. An experimental setup has been created to prevent participants from visually monitoring their input (i.e., to keep the visual monitoring as the invariable); the participants had to keep their gaze (controlled by eye tracking) inside a predefined area while tapping on one of the 15 targets (on a 5 × 3 matrix) on the touch screen. The target acquisition tasks were accepted only if the participants kept their gaze within the predefined area. The second part of the study, on the other hand, treated visual monitoring as the variable and the positional accuracy as the invariable, and tasks were accepted only if the participants accurately pointed to the target.

The two stages respectively yielded 1080 and 216 trials from 12 participants (× 90 tasks). The scatter plot in Figure 4.1 shows the relationship between the distance of the gaze point to the target position (invariable) to the positional offset (distance between the touch and target positions).

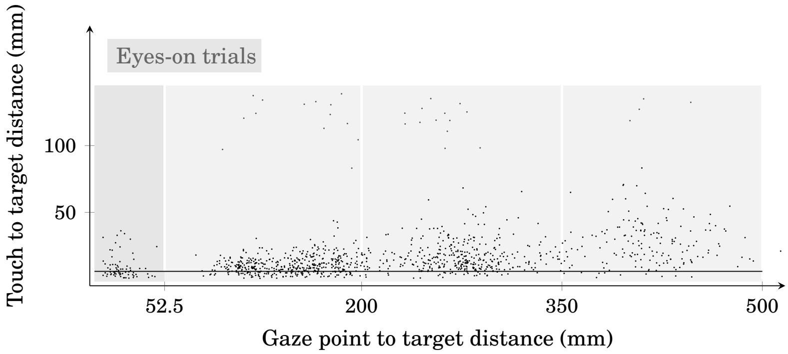


Figure 4.1. Scatter plot of peripheral target acquisition tasks across all participants. The horizontal line indicates the visual boundary of the circle target (rad=5.8mm). The darker background indicates the eyes-on in which the target was within the boundary of the circle the participants had to keep their gaze inside(rad=52.5mm).

The data has been divided into four continuous bins that correspond to varying levels of visual guidance. Figure 4.2 shows the distribution of touch points relative to the target across all users for four chosen intervals of visual monitoring. In line with expectations and previous research, the results showed decreased positional accuracy for increased distance between the touch and gaze points, which can be used to estimate positional uncertainty for input handling.

4.2 RQ1.2: What are the particular considerations for touch input without visual monitoring?

RQ1.1 confirmed the decreased positional accuracy for low visual monitoring, but did not investigate the effect of interaction techniques on user behavior. This was investigated through another set of tasks in the same session. The participants were asked to perform open ended tasks with three different applications until they felt comfortable with the interaction techniques. The sessions were video recorded and participants were interviewed immediately after using each of the applications. The main observations can be summarized as below:

- **Adjustment through use.** Participants often acknowledged the difficulty of "touching without looking" at the start of the session and admitted to force themselves not to redirect their gaze to the touch location. At the same time, later experience has been described as "natural" and "easier" as participants developed a better understanding of how their touch will be interpreted by

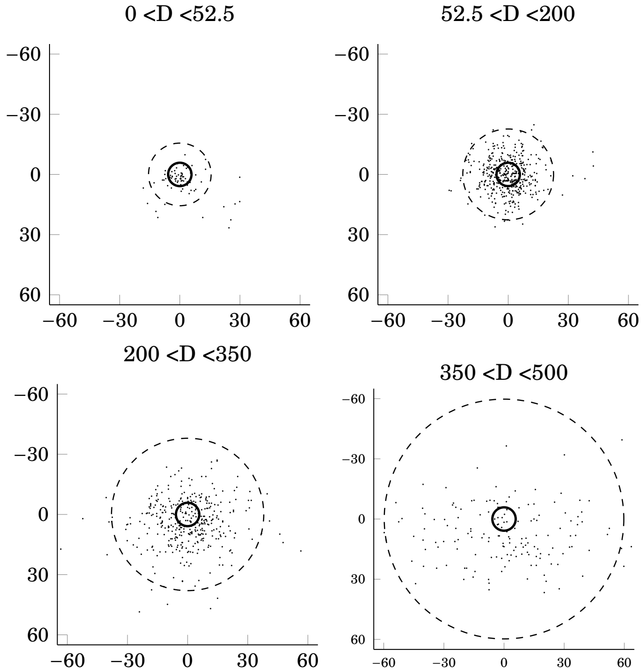


Figure 4.2. The distribution of touch points relative to the target across all users for chosen ranges of distance between gaze point and target. The dashed circles are the 95 % confidence circles. The inner solid circles show the target visual boundary. All units in mm.

the system. Being able to use two hands on the surface while keeping their attention on the work area has also been highlighted as an advantage. An observation from these comments is that input without visual guidance can require a degree of familiarity both with the interface layout and an increased understanding of how the system will handle the input.

- **Misinterpretation of positional uncertainty.** We observed a number of instances in which a manual input action was wrongly interpreted as positionally uncertain due to the system's lack of awareness of the hand and finger movements before the actual touch event. One instance involved a participant keeping his finger just above a specific point on the interface and performing the touch action while looking elsewhere. In these cases, although the participants knew exactly where they were pointing to, the application interpreted it as positionally uncertain and handled the input accordingly. The participant occasionally identified this as a "problem".

- **Screen edge as ambiguous border and tactile guide.** Although a touch screen is an input field with definite boundaries, decreased visual guidance can cause ambiguity for users regarding whether they are *addressing* the system during touch. In some instances, while aiming for the color palette near the edge of the screen, participants touched the insensitive bezel area of the screen. The lack of visual feedback in this case communicated that the system is not addressed, which led the users to a repeated touch action. On the other hand, device borders provide potential tactile cues for eyes-free use. This was observed again when participants *anchored* their left hand on the screen edge for sliding along the input widget with their thumb or index finger while keeping their gaze on another location. So, although the design program targeted flexible input handling the participants still utilized tactile cues to a certain extent to aid eyes-free use.

4.3 RQ2.1: How does the performance of a gaze-aware interaction technique compare with traditional input for acquisition and manipulation tasks?

Publication I evaluated the positional accuracy of input without visual monitoring and qualitatively evaluated different applications, but stopped short of evaluating the performance of various interaction techniques. Publication II has been devised to answer the performance of uncertain input handling with warped visual feedback when compared to a baseline condition. Previous work identifies two interaction stages, namely acquiring and manipulating a control device [29, 115]. A two-part experiment has been prepared that involved a widget 1) acquisition and 2) manipulation task on a touch screen. The two experimental conditions were:

1. Warped visual feedback condition, which facilitates continuous gaze fixation near the stimulus position using a small representation of the user's hand (scaled down by a factor of 0.35 to be visible and less intrusive).
2. Baseline condition, which provides no specific support to facilitate continuous fixation.

Overall, we anticipated time savings by eliminating attention switches under the warped visual feedback condition. However, and in line with previous work in oculomotor coordination [100, 106, 124], we also anticipated a decrease in motor performance in warped feedback condition due to decreased visual monitoring. The study aimed to observe the cumulative effect of these two factors, namely time savings due to eliminating attention switches and losses from motor performance. In addition to two input conditions, the study featured two screen conditions to vary the cost of attention shift. In the first case, the stimuli were shown on the same screen which resulted in a visual angle of around 50° degrees between the input and stimuli positions. In the second case, the stimuli were shown on a separate vertical screen which resulted in a visual angle of around 70° degrees between the input and stimuli positions.

For acquisition tasks, the study showed a significant difference based on whether the action involved lengthy mid-air movements. When the task required users to acquire a widget through midair movements without visually monitoring their hand (Table 4.2, between-widget tasks), the performance decreased, with participants spending significantly more time on the warp condition than on the baseline condition for both the same screen ($r = .42$, $p < .001$) and vertical screen conditions ($r = .54$, $p < .001$). For within-widget tasks, the mean completion times for warped feedback and baseline conditions were similar in both stimuli conditions (Table 4.2). A t-test comparison using within-subject normalized completion times did not show any significant effect for the same screen ($r = .03$, $p = .23$) and vertical screen ($r = .03$, $p = .17$) conditions but the error rates were higher for the warped feedback condition (Table 4.2).

Acquisition task *(within-widget)*

Screen	Technique	Median(ms)	Mean(ms)	Error
Same	Warped	**902.75**	**1031.54**	6.41%
	Baseline	961.75	1068.36	**2.43%**
Vert.	Warped	**885.50**	1055.00	8.16%
	Baseline	986.25	**1021.88**	**3.31%**

Acquisition task *(between-widget)*

Screen	Technique	Median(ms)	Mean(ms)	Error
Same	Warped	2094.00	2305.19	11.33%
	Baseline	**1164.50**	**1406.79**	**1.09%**
Vert.	Warped	2258.75	2342.36	9.09%
	Baseline	**1207**.5	**1415.19**	**2.17%**

Table 4.2. The grand median and grand mean completion times and overall error rates for two interaction and two stimuli conditions for the acquisition tasks. Emphasis (in bold) represents better performance.

The conclusion from this part of the study was that, the gains from not having to shift visual attention did not compensate for the losses of manual coordination due to low visual attention. Our qualitative observations are also in this direction: While participants performed high-speed ballistic movements towards the touch target in the baseline condition, they moved their hand parallel to the screen and kept a tense hand posture in the warped feedback condition. Participants also reported shoulder fatigue for warped feedback condition, which may have been caused by the parallel hand movements.

On the other hand, the warped feedback increased the performance for manipulation tasks; participants spent more time on the baseline condition than on the warped feedback condition (Table 4.3). A t-test comparison of the same and vertical screen conditions using normalized data yielded a larger effect size for the vertical screen condition ($r = .19$, $p < .001$) than for the same screen condition ($r = .10$, $p < .001$), in line with the expectation that the higher cost of redirecting the gaze in vertical screen condition will result in more pronounced benefits when using warped feedback. The error rates were lower for the warped feedback condition in both screen conditions (Table 4.3).

Manipulation Task

Screen	Technique	Median(ms)	Mean(ms)	Error
Same	Warped	**1387.25**	**1576.30**	**4.07%**
	Baseline	1530.25	1746.89	6.46%
Vert.	Warped	**1364.00**	**1509.06**	**3.69%**
	Baseline	1616.50	1770.58	6.54%

Table 4.3. The grand median and grand mean completion times and overall error rates for two interaction and two stimuli conditions for the manipulation task. Emphasis (in bold) represents better performance.

Overall, the performances of the warped feedback and baseline conditions were visibly different based on whether the task was manipulation, within-widget acquisition or between-widget acquisition. The performance of the warped visual feedback condition was higher for manipulation tasks that required no midair motion. The performance between warped feedback and baseline conditions were comparable for within-widget acquisition. However, the performance of the warped visual feedback was significantly worse for between-widget acquisition tasks, in which participants had to acquire the widget through midair motion. Based on these results, I arrive at the following conclusions:

- The warped feedback was successful in decreasing the cost of redirecting the gaze, resulting in the improvement of task completion time for manipulation tasks.
- However, the warped feedback did not facilitate midair hand motion as effectively as direct visual monitoring, which resulted in a decrease in performance for between-widget acquisition tasks. Here, the results are in line with earlier work that reports lower performance and similar observations such as tense hand posture when touch is performed without direct visual monitoring [100].

4.4 RQ3.1: What are the visual attention-based access preferences for different actions?

Publications I and II evaluated applications in which the level of visual monitoring required for each action has been determined in advance during the design phase. In other words, the input handling was specified in advance of the user study as part of the study design. However, the level of visual monitoring required for different applications can itself be the object of empirical inquiry. This is particularly the case in collaborative interfaces, in which a variety of social considerations can lead users to choose different reasons for determining the level of visual monitoring required for an action.

Publications III thus set out to find out participants' input handling preferences for different actions. The input handling preferences corresponded to four access types (consensual, supervised, universal and private) based on their availability in different joint attention conditions. Participants (in pairs) were tasked to decide which actions should belong to different access types as they complete three different scenarios of project planning, brainstorming and document sharing.

The results show salient differences between user preferences across different applications. A general pattern is the use of the access type that poses no restriction (universal access) for actions that do not involve manipulation or are easily reversible such as viewing items (72.3%), moving individual elements (68.6%) or creating new items (95.0%). In contrast, universal access was rarely assigned to element-level delete (7.1%) and never to global delete actions. On the other hand, consensual access that requires joint attention was assigned to actions with global scope such as exiting the session (56.7%), global deletion (76.7%) and aligning elements (35%).

4.5 RQ3.2: What are the motivations for different visual attention-based access preferences?

While the preference data for visual attention-based access types provide a summary of general patterns, it does not directly answer what accounts for the differences between user preferences for the same actions. A separate analysis has been conducted by encoding participant remarks that were recorded as they conducted the tasks and also through interviews. The remarks give insights about participants' externalized reasoning for choosing different access types and the considerations that came into play. Publication III encoded these various considerations into themes. Here, I will summarize the general observations that concern the use of joint attention information for granting access rights on shared workplaces.

- A finding in line with the expectations was participants' assignment of access types that require joint attention to prevent accidents and conflicts. We observed that, in addition to the action type, participants identified content type and the interaction history of an item when deciding on whether an action requires joint visual attention. For example, joint agreement on the content of an item has been highlighted as a reason for requiring joint attention for the item. On the other hand, the participants identified some content as tentative and did not require joint attention for editing or deleting them.

- Participants assigned access types not only for conflict prevention but also for facilitating awareness. In some situations, participants preferred visual attention-based access types not for preventing conflicts but as a means to ensure that the other user is aware of the action or to direct the other user's attention. In these cases, users deliberately utilized access control mechanisms in order to control awareness, providing a counter-example to our conceptualization of access management as an implicit effect of visual attention. Yet the awareness that is achieved through forcing visual attention can come at the expense of flexibility and we observed instances in which participants reverted back to lack of access control when joint attention was impractical.

- Granting access with head-orientation introduced uncertainty. Not having to manually touch the screen for confirmations was highlighted as a convenient feature. At the same time, visual attention-based access introduced uncertainty that was attributed both to a mismatch between head orientation and participants' actual locus of visual attention and also to situations when visual attention does not indicate awareness (i.e., when participants remarked that they may be looking but not paying cognitive attention). In some cases, the participants decreased the uncertainty through a work-around, by using *private* access type that restricts action when another user is looking. By doing so, they precluded giving access by accident.

4.6 Summary

The section presented the main observations conducted within the scope of this thesis. The results of different research questions can be summarized as below:

- **How is touch accuracy affected by decreased visual monitoring?**
  In line with expectations, lower visual monitoring led to a decreased accuracy for pointing tasks on a touch screen, and the study showed a linear relationship between positional inaccuracy and the distance of the gaze point to the target.

- **RQ1.2: What are the particular considerations for touch input without visual monitoring?**
  The qualitative observations gave insight into a number of practical issues and use patterns that emerge during interaction with lower visual monitoring. The main observations are 1) the need for adjustment for pointing with lower visual monitoring, 2) the potential misinterpretations of positional uncertainty and 3) the use of screen edges as a tactile guide.

- **RQ2.1: How does the performance of a gaze-aware interaction technique compare with traditional input for acquisition and manipulation tasks?**
  The design intervention resulted in a performance improvement for manipulation tasks, but a deterioration for tasks that require larger amplitude mid-air motion. This pointed to a trade-off between performance gains achieved through eliminating visual attention shifts and losses due to decreased motor performance that results from lack of visual monitoring.

- **RQ3.1: What are the visual attention-based access preferences for different actions?**
  The logged data showed salient differences between user preferences based on the type (e.g., edit, delete) and scope (individual, global) of actions as well as the content and the individual interaction history of an item. In general, actions that are harder to reverse were assigned more restrictive access criteria. We also observed a number of counter-intuitive preferences that further made the case for the qualitative analysis of interaction and interview data.

- **RQ3.2: What are the motivations for different visual attention-based access preferences?**
  In line with the prior expectations, visual attention-based access control has been used to prevent conflicts. Yet participant interactions and comments also pointed to a number of other motivations such as making it easier to keep track of the workspace and directing others' attention. Visual attention-based access has been perceived as convenient but also uncertain.

The next section positions the design work conducted within the scope of this thesis and empirical observations within the context of more general discussions in HCI research.

5. Discussion

Early in the thesis, I have listed some possible interpretations of visual attention information:

- Visual attention information can be a measure of what users *prioritize* to monitor, providing information about what they plan to do or what they might accept as appropriate system behavior.

- Visual attention information can be a measure of what users have *already monitored*, providing information about the extent they are aware of the interface state, the actions of others or the position of their own body parts.

- Visual attention information can correspond to what users *aim to signal* to the system or to the other users in the environment. In this case, the interpretation of visual attention-related actions depends on how the system and others in the environment utilize this information and the extent a user is aware of these utilizations.

A challenge facing HCI is to design interfaces by taking these diverse considerations into account. In this thesis, I have focused on the second consideration, the use of visual attention information as a measure of users' awareness of the environment, and aimed to address the constraints posed by visual attention due to the limited spatial acuity of the eyes. Through different prototypes, I contributed to the HCI research by proposing new interaction techniques that handle user inputs based on the visual attention, and evaluated these interfaces in formative studies to identify further considerations for design and research. While each prototype and empirical study contributes to their respective domains of single-user interaction techniques and groupware, it is useful to situate the individual observations within the context of more general HCI discussions. Here, I will discuss the observations in terms of the trade-offs between time and spatial multiplexing, and between adaptiveness and predictability in interface design. I will then discuss the work in terms of the tension between adapting to users' existing behavior and transforming this behavior through designing interventions.

5.1 Time and Spatial Multiplexing

The thesis early on noted that human attention can be conceived as a limited resource as observed in the performance trade-offs between multiple time-shared (concurrent) tasks [121]. This observation translates into a design trade-off for interactive systems: In an interface, a designer can choose to devote users' attention to a single task in order to maximize its performance or can parallelize between multiple tasks. The latter could decrease the performance of a single task, but can provide gains through concurrency. The design trade-off is not limited to single-user cases. Research on collaborative systems has long identified a fundamental trade-off between awareness and individual power in groupware design [22, 42]. The ability of individual users to view different parts of the workspace at the same time (as in relaxed WYSIWIS [what you see is what I see] interfaces) provides flexibility, but potentially decreases users' awareness of each others' actions and their general coordination. Different levels of coordination consequently result in a trade-off between performance gains through concurrency and potential losses in the overall group performance due to lack of awareness [53] (i.e., when users do duplicate work or when their contribution is rejected).

The design motivations and the empirical results of this thesis can partly be explained through this trade-off. However, as the thesis mainly focused on the visual attention caused by the *spatial* acuity of the eyes, it is useful to describe the trade-off in spatial terms. A relevant distinction from previous HCI work is that of between time and spatial multiplexing [29]. In time multiplexing, different actions are allocated separate time windows. This allows an individual action to be carried out one at a time and at a single interface location. In spatial multiplexing, actions are conducted in parallel at different locations. While the original work of Fitzmaurice and Buxton [29] limited the scope of spatial multiplexing to manual manipulation actions, I here find it useful to expand the concept to cover *manipulation and perception* on multiple locations. For example, typing on a keyboard while monitoring the screen involves spatial multiplexing, not only due to the concurrent input by many fingers but also due to the spatially distributed input and visual output areas. In this expanded definition, spatial multiplexing can express the distinctions between the execution of a manual action with or without visual monitoring during single-user interaction, or under solitary or joint attention during group work.

An important question for system design is whether the gains in parallel execution make up for the losses in decreased performance (or user comfort) of a single task. The question is all more relevant with the emergence of eye movements as an input; eyes move rapidly but have a single positional focus. Designers face the choice of utilizing eye movements to sequentially point to different interface locations (as in previous work that use gaze as a pointer [e.g., 129]), or as an additional input that complements concurrent positional input from other sources. The first approach allows time multiplexing by pointing to targets one at a time but rapidly. The latter approach targets spatial multiplexing, but the input accuracy can decrease due to the lack of visual monitoring.

The aim of this thesis has been to support spatial multiplexing by addressing the problem of decreased performance in divided attention cases through various interaction techniques. Publication I, in line with previous work, demonstrated that pointing performance indeed decreases when the user is not visually attending to the input. It accordingly proposed various interaction techniques that aim to support spatial multiplexing through uncertain input handling and visual feedback. Publication II showed that the successful trade-off depends on the amount of mid-air motion that needs to be executed without visual monitoring. The study conducted in Publication III was explorative and did not measure performance, but the findings showed in which cases the users would want to allow spatial multiplexing (by making actions available during any attention condition) and in which cases they would want to enforce time multiplexing in order to minimize accidents or conflicts (by assigning access rights so that an action requires joint attention).

5.2 The Uncertainty Introduced by Adaptiveness

The thesis introduced interaction techniques that handle users' input based on their level of visual attention. These were proposed as an alternative to static solutions that target decreasing the need for visual monitoring such as providing tactile cues (for single-user attention) or pre-defined divisions of labor in groupware (for group attention). Here, it should be noted that adaptiveness might come with its own potential drawbacks. The trade-off between adaptiveness and predictability is a long-acknowledged problem in HCI, with some studies reporting a performance advantage for static interfaces [27] and others for adaptive interfaces [34]. A potential interpretation of these different findings is that the performance of adaptive interfaces depends on multiple factors including the particular handling method, the task and the user profile [26]. For example, static interfaces with persistent layouts might better facilitate a spatial memory of an interface, but this advantage might not be as pronounced for novice users or when the number of interface elements are high.

Some observations reported in this thesis can be understood through this trade-off between adaptiveness and predictability. For example, Publication I reported that users required some time to get used to performing touch input without closely monitoring their hand but also to get comfortable with the system handling of the input. Previous work in uncertain input handling promotes providing visual feedback to inform users about how the system interprets their action [102]. The prototypes in Publication I and Publication II used warped visual feedback to inform the user about system interpretation of their action. Visual feedback increases predictability before the actual action execution, but might cause an additional performance bottleneck as the user has to wait for and monitor the system feedback. The trade-off between adaptiveness and predictability in these cases can thus be explained through the need for system feedback: *"As the asymmetry shifts towards feedback-dominated control, the complexity of the model is transferred from the user's mind to the system. This makes the user more dependent on feedback, but requires less training and more efficient use of the input available."* [122, p. 833].

A similar observation has been made in Publication III, when users created work-arounds around to decrease uncertainty. Here, the users welcomed the convenience of not having to perform dedicated manual actions for granting access, but noted that some critical cases might require more certainty. Previous research noted that contextual access management approaches have the drawback of decreased understandability [113] and visual attention-based access is no exception.

Thus, potential decreases in predictability is a consideration that needs to be kept in mind in addition to the trade-off between time and spatial multiplexing when designing for adaptive interaction techniques to support input with low visual monitoring.

5.3 Design Interventions and Adaptiveness

Early in the dissertation, I noted that the communicative uses of visual attention information partly shifts the analytic focus from visual attention as an objective phenomenon to visual attention as something that is perceived and interpreted by other agents. The same insight also applies to the design of interactive systems that adapt their behavior based on a user's visual attention information. As with humans, the system's interpretation of visual attention is determined by its sensing and modeling capabilities. Prior design assumptions about what is visually attended or what is appropriate system behavior do not always match with the subtlety of the natural user behavior. In this case, the success of the interaction partly depends on users behaving in a way that makes their visual attention interpretable by the system. This has been observed both in the context of single-user interactions (i.e., system's misinterpretations of user's visual monitoring) and multi-user interactions (e.g., when the users are looking but are not paying cognitive attention).

The mismatch in sensing visual attention can be addressed through models that more elaborately sense and model pre-intervention (i.e., natural) user behavior. Yet part of the mismatch is inherent to the act of designing interactive systems; the introduction of adaptive technology can ultimately transform the behavior that it aims to adapt to (parallel to the previously identified 'paradox of system design' [14]). This observation has been made within the context of implicit interactions in Publication IV, which proposed asking how different assumptions that guide implicitness or adaptiveness make certain interaction outcomes harder. While this thesis aimed to support some of the use cases that are left by the previous applications of visual attention information (that aimed to address the motor bottleneck), the results showed that the designing for limited visual attention can also make certain interaction outcomes harder. For example, visual attention-based access in Publication III enables access when another user is paying attention, but this interaction mechanism also makes it harder to visually monitor another user without granting access.

5.4 Limitations and Future Work

The contributions of the thesis are primarily constructive and the empirical studies were formative in the sense that they were mostly oriented towards identifying design considerations for future work instead of quantifying the effects of various prototypes. As with every formative study, there are limitations to what can be claimed as final design implications. First, the thesis prioritized utilizing visual attention information in novel ways instead of building precise models of visual attention. Yet, the actual deployment would benefit from more precise models of visual attention information and how it affects awareness. Such models can benefit from the inclusion of additional stimuli-related variables (e.g., color, size and previous knowledge) that influence peripheral salience. This would benefit the selection of input handling and visual feedback techniques employed (e.g., the choice between making peripheral objects larger or warping them to the center of visual attention).

In some cases, visual attention information alone can be an insufficient measure of user awareness and more complex models of memory can be needed to infer user awareness. Additionally, I identified various considerations related to visual attention and adaptiveness (such as the limitation of visual attention, the trade-offs between time and spatial multiplexing as well as between predictability and adaptiveness), but stopped short of providing a complete model that enables their comparison on the basis of performance or other criteria. The fragmentation of attention research in HCI is an acknowledged problem [95] and this thesis does not fully address it. Here, it is useful to discuss a methodological drawback of the constructive research approach. I noted that one advantage of the approach is the ability to gather information about possible design interventions without having to construct detailed models of the problem space. Yet what makes constructive research programs practical can also make their integration into the existing body of knowledge harder. Thus, the consolidation of various considerations that come into play when using visual attention information remains a task for future research.

Another limitation of the thesis is the ecological validity of its observations.

All the studies have been conducted in controlled settings in order to deploy dedicated sensors and the tasks have been selected based on their demands on visual attention. This is a potential limitation when transferring the knowledge to more realistic tasks encountered in daily settings. More informed claims about the utility of the interaction methods require observing a wider range of visual monitoring behavior and conducting additional studies to observe long-term use and habituation. This is especially relevant for collaborative interfaces as user habituation can involve the development of social practices, which can be best observed in longitudinal deployments in the wild.

Finally, the constructive research program defined in this thesis, adapting interaction to users' level of visual monitoring during input, has a wider scope than that could be carried out during the thesis period. For single-user applications, I prioritized pointing due to its general relevance for HCI and also because it provided a good opportunity to compare my own research program with existing work in eye tracking research. Yet the research program can be expanded to more complex tasks such as information seeking or visual analytics. For example, search interfaces typically rely on typed queries. Users type the queries themselves and can thus be safely assumed to be aware of their own input. As entity-based search gains ground (enabling users to input whole documents as search inputs), however, it can become useful to understand what the user has visually attended to in a document before submitting it as a search input. This would, in return, require the use of eye movements or other visual attention information as a measure of user awareness (in contrast to the more extensively researched use of visual attention data as a measure of user interest [e.g., 12, 43]). The work on visual attention-based access (Publication III) expanded the research program to groupware, but this was limited to collocated and synchronous interactions and more work is needed to assess the utility of the research program for remote and asynchronous interactions. Whether the input handling and visual feedback techniques can be applied in these situations, or whether the trade-offs identified within the scope of this work are explanatory beyond the particular application areas remain open questions.

5.5 Conclusions

The increased sensing and inference capability of computers require reevaluating the division of labor between the user and the system as well as between different human actions such as eye and hand movements. This thesis contributed to the on-going HCI discussions on how to utilize visual attention information.

I have laid out how different research insights from research on attention and visual attention lead to different considerations for interface design. I then identified the constraints posed by limitation to human visual acuity as a central consideration for utilizing visual attention information during interaction. This led to a constructive research program of adapting interaction to users' level of visual monitoring during input. The program was instantiated through a series of prototypes developed for single-user and collocated multi-user applications.

The resulting interaction techniques and the observations gained during their evaluation are the main outcomes of this thesis. These involve various input handling and visual feedback methods that compensate for users' lack of visual attention during input with the ultimate aim of allowing concurrent input or maintaining coordination during group work. I have consolidated these methods under an uncertain input handling framework that apply to diverse use cases. The empirical observations gave insights about the particular strengths and drawbacks of these interaction techniques. Particularly, I have quantified the relationship between positional accuracy to the distance between gaze and touch input for pointing tasks and identified the amount of midair motion during manual input as one factor that determines the efficacy of the interaction techniques. The qualitative analysis of the data gathered through observations and interviews point to additional considerations for future system design.

References

[1] Alan Allport. Attention and control: Have we been asking the wrong questions? a critical review of twenty-five years. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, pages 183–218, 1993.

[2] Michael Argyle and Janet Dean. Eye-contact, distance and affiliation. *Sociometry*, 28(3):289–304, 1965.

[3] Michael Argyle, Roger Ingham, Florisse Alkema, and Margaret McCallin. The different functions of gaze. *Semiotica*, 7(1):19–32, 1973.

[4] Geoffrey W. Beattie. Floor apportionment and gaze in conversational dyads. *British Journal of Social and Clinical Psychology*, 17(1):7–15, 1978.

[5] Geoffrey W Beattie. Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics*, 17(1-2):61–78, 1979.

[6] B. Biguer, M. Jeannerod, and C. Prablanc. The coordination of eye, head, and arm movements during reaching at a single visual target. *Experimental Brain Research*, 46(2):301–304, May 1982.

[7] Renaud Blanch, Yves Guiard, and Michel Beaudouin-Lafon. Semantic pointing: Improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 519–526, New York, NY, USA, 2004. ACM.

[8] O Bock and R Eckmiller. Goal-directed arm movements in absence of visual guidance: evidence for amplitude rather than position control. *Experimental Brain Research*, 62(3):451–458, 1986.

[9] Richard A. Bolt. Gaze-orchestrated dynamic windows. *SIGGRAPH Comput. Graph.*, 15(3):109–119, August 1981.

[10] Donald Eric Broadbent. *Perception and communication*. Elsevier, 1958.

[11] Liana E Brown, David A Rosenbaum, and Robert L Sainburg. Limb position drift: implications for control of posture and movement. *Journal of Neurophysiology*, 90(5):3105–3118, 2003.

[12] Georg Buscher, Andreas Dengel, and Ludger van Elst. Query expansion using gaze-based feedback on the subdocument level. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 387–394, New York, NY, USA, 2008. ACM.

[13] Brian Butterworth. Maxims for studying conversations. *Semiotica*, 24(3-4):317–340, 1978.

[14] Graham Button and Paul Dourish. Technomethodology: Paradoxes and possibilities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 19–26, New York, NY, USA, 1996. ACM.

[15] W. Buxton and B. Myers. A study in two-handed input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '86, pages 321–326, New York, NY, USA, 1986. ACM.

[16] Stuart K. Card, Allen Newell, and Thomas P. Moran. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1983.

[17] Marvin M. Chun, Julie D. Golomb, and Nicholas B. Turk-Browne. A taxonomy of external and internal attention. *Annual review of psychology*, 62:73–101, 2011.

[18] Susana T.L. Chung, J.Stephen Mansfield, and Gordon E. Legge. Psychophysics of reading. xviii. the effect of print size on reading speed in normal peripheral vision. *Vision Research*, 38(19):2949–2962, 1998.

[19] Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.

[20] Jakub Dostal, Uta Hinrichs, Per Ola Kristensson, and Aaron Quigley. Spidereyes: Designing attention- and proximity-aware collaborative interfaces for wall-sized displays. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, pages 143–152, New York, NY, USA, 2014. ACM.

[21] Jakub Dostal, Per Ola Kristensson, and Aaron Quigley. Subtle gaze-dependent techniques for visualising display changes in multi-display environments. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, pages 137–148, New York, NY, USA, 2013. ACM.

[22] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, CSCW '92, pages 107–114, New York, NY, USA, 1992. ACM.

[23] Andrew T Duchowski. *Eye tracking methodology: Theory and practice*. Springer New York, 2007.

[24] John Duncan. Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4):501, 1984.

[25] Diego Fernandez-Duque and Mark L. Johnson. Cause and effect theories of attention: The role of conceptual metaphors. *Review of General Psychology*, 6(2):153–165, 2002.

[26] Leah Findlater and Krzysztof Z. Gajos. Design space and evaluation challenges of adaptive graphical user interfaces. *AI Magazine*, 30(4):68, Sep. 2009.

[27] Leah Findlater and Joanna McGrenere. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 89–96, New York, NY, USA, 2004. ACM.

[28] Paul M Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[29] George W. Fitzmaurice and William Buxton. An empirical evaluation of graspable user interfaces: Towards specialized, space-multiplexed input. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 43–50, New York, NY, USA, 1997. ACM.

[30] George W Fitzmaurice, Hiroshi Ishii, and William AS Buxton. Bricks: laying the foundations for graspable user interfaces. In *Proc. CHI'95*, pages 442–449. ACM, 1995.

[31] J. Randall Flanagan, Yasuo Terao, and Roland S. Johansson. Gaze behavior when reaching to remembered targets. *Journal of Neurophysiology*, 100(3):1533–1543, 2008.

[32] Frank O Flemisch, Catherine A Adams, Sheila R Conway, Ken H Goodrich, Michael T Palmer, and Paul C Schutte. The h-metaphor as a guideline for vehicle automation and interaction. 2003.

[33] Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4):694, 2007.

[34] Krzysztof Z. Gajos, Katherine Everitt, Desney S. Tan, Mary Czerwinski, and Daniel S. Weld. Predictability and accuracy in adaptive user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1271–1274, New York, NY, USA, 2008. ACM.

[35] James J Gibson and Anne D Pick. Perception of another person's looking behavior. *The American Journal of Psychology*, 76(3):386–394, 1963.

[36] Charles Goodwin. The interactive construction of a sentence in natural conversation. *Everyday language: Studies in ethnomethodology*, pages 97–121, 1979.

[37] Charles Goodwin. Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological inquiry*, 50(3-4):272–302, 1980.

[38] Charles Goodwin. *Pointing: Where language, culture and cognition meet*, chapter Pointing as situated practice, pages 217–241. Lawrence Erlbaum Associates, 2003.

[39] Tovi Grossman and Ravin Balakrishnan. The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 281–290, New York, NY, USA, 2005. ACM.

[40] Qi Guo and Eugene Agichtein. Towards predicting web searcher gaze position from mouse movements. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3601–3606, New York, NY, USA, 2010. ACM.

[41] Carl Gutwin, Scott Bateman, Gaurav Arora, and Ashley Coveney. Looking away and catching up: Dealing with brief attentional disconnection in synchronous groupware. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 2221–2235, New York, NY, USA, 2017. ACM.

[42] Carl Gutwin and Saul Greenberg. Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, CSCW '98, pages 207–216, New York, NY, USA, 1998. ACM.

[43] Jacek Gwizdka. Characterizing relevance with eye-tracking measures. In *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, pages 58–67, New York, NY, USA, 2014. ACM.

[44] Toshiyuki Hagiya and Tsuneo Kato. Probabilistic touchscreen text entry method incorporating gaze point information. *Journal of Information Processing*, 23(5):708–715, 2015.

[45] Mary Hayhoe. Vision using routines: A functional account of vision. *Visual Cognition*, 7(1-3):43–64, 2000.

[46] Christian Heath and Paul Luff. Collaboration and control: Crisis management and multimedia technology in london underground line control rooms. *Computer Supported Cooperative Work (CSCW)*, 1(1):69–94, 1992.

[47] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. Edit wear and read wear. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 3–9, New York, NY, USA, 1992. ACM.

[48] James E. Hoffman and Baskaran Subramaniam. The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57(6):787–795, Jan 1995.

[49] Eric Horvitz, Carl Kadie, Tim Paek, and David Hovel. Models of attention in computing and communication: From principles to applications. *Commun. ACM*, 46(3):52–59, March 2003.

[50] Roberto Hoyle, Srijita Das, Apu Kapadia, Adam J. Lee, and Kami Vaniea. Was my message read?: Privacy and signaling on facebook messenger. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3838–3842, New York, NY, USA, 2017. ACM.

[51] Jeff Huang, Ryen White, and Georg Buscher. User see, user point: gaze and cursor alignment in web search. In *Proc. CHI'12*, pages 1341–1350. ACM, 2012.

[52] Jeff Huang, Ryen W. White, Georg Buscher, and Kuansan Wang. Improving searcher models using mouse cursor activity. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 195–204, New York, NY, USA, 2012. ACM.

[53] P. Isenberg, D. Fisher, S. A. Paul, M. R. Morris, K. Inkpen, and M. Czerwinski. Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):689–702, May 2012.

[54] Hiroshi Ishii and Brygg Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM.

[55] Robert JK Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI'90*, pages 11–18. ACM, 1990.

[56] Robert J.K. Jacob and Keith S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach, and H. Deubel, editors, *The Mind's Eye*, pages 573–605. North-Holland, Amsterdam, 2003.

[57] William James. *The Principles of Psychology Vol.2*. Holt, New York, NY, USA, 1890.

[58] William A Johnston and Veronica J Dark. Selective attention. *Annual review of psychology*, 37(1):43–75, 1986.

[59] Marcel A Just and Patricia A Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[60] Paul Kabbash, William Buxton, and Abigail Sellen. Two-handed input in a compound task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 417–423, New York, NY, USA, 1994. ACM.

[61] Diane Kelly et al. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1–2):1–224, 2009.

[62] Diane Kelly and Jaime Teevan. Implicit feedback for inferring user preference: A bibliography. *SIGIR Forum*, 37(2):18–28, September 2003.

[63] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.

[64] David Kirsh. Complementary strategies: Why we use our hands when we think. In *Proceedings of the seventeenth annual conference of the cognitive science society*, pages 212–217. Lawrence Erlbaum Associates, Inc., 1995.

[65] David Kirsh. The intelligent use of space. *Artificial Intelligence*, 73(1-2):31–68, February 1995.

[66] David Kirsh and Paul Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4):513–549, 1994.

[67] Chris L Kleinke. Gaze and eye contact: a research review. *Psychological bulletin*, 100(1):78, 1986.

[68] Ilpo Koskinen, John Zimmerman, Thomas Binder, Johan Redström, and Stephan Wensveen. *Design research through practice: From the lab, field, and showroom*. Morgan Kaufmann, 2011.

[69] Manu Kumar, Andreas Paepcke, and Terry Winograd. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proc. CHI'07*, pages 421–430. ACM, 2007.

[70] Dmitry Lagun and Eugene Agichtein. Inferring searcher attention by jointly modeling user interactions and content salience. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 483–492, New York, NY, USA, 2015. ACM.

[71] Imre Lakatos. Falsification and the methodology of scientific research programmes. In Sandra G. Harding, editor, *Can Theories be Refuted? Essays on the Duhem-Quine Thesis*, pages 205–259. Springer Netherlands, Dordrecht, 1976.

[72] Michael Land, Neil Mennie, and Jennifer Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328, 1999.

[73] Peter H Lindsay and Donald A Norman. *Human information processing: An introduction to psychology*. Academic press, 1972.

[74] Yung-Ching Liu. Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics*, 44(4):425–442, 2001.

[75] R Duncan Luce. Whatever happened to information theory in psychology? *Review of general psychology*, 7(2):183, 2003.

[76] I Scott MacKenzie. A note on the information-theoretic basis for fitts' law. *Journal of motor behavior*, 21(3):323–330, 1989.

[77] Päivi Majaranta and Andreas Bulling. Eye tracking and eye-based human–computer interaction. In *Advances in Physiological Computing*, pages 39–65. Springer, 2014.

[78] Jennifer Mankoff, Scott E Hudson, and Gregory D Abowd. Providing integrated toolkit-level support for ambiguity in recognition-based interfaces. In *Proc. CHI'00*, pages 368–375. ACM, 2000.

[79] M-Marsel Mesulam. Attention, confusional states, and neglect. In *Principles of behavioral and cognitive neurology*, pages 125–168. F. A. Davis, 1985.

[80] David E Meyer and Sylvan Kornblum. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, volume 14. MIT Press, 1993.

[81] Neville Moray. Where is capacity limited? a survey and a model. *Acta psychologica*, 27:84–92, 1967.

[82] Vidhya Navalpakkam, LaDawn Jentzsch, Rory Sayres, Sujith Ravi, Amr Ahmed, and Alex Smola. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 953–964, New York, NY, USA, 2013. ACM.

[83] Allen Newell and Stuart K. Card. The prospects for psychological science in human-computer interaction. *Hum.-Comput. Interact.*, 1(3):209–242, September 1985.

[84] Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. Looks can be deceiving: Using gaze visualisation to predict and mislead opponents in strategic gameplay. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 261:1–261:12, New York, NY, USA, 2018. ACM.

[85] Jacob L. Orquin and Simone Mueller Loose. Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 144(1):190–206, 2013.

[86] Miles L Patterson. A sequential functional model of nonverbal exchange. *Psychological review*, 89(3):231–249, 1982.

[87] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 509–518, New York, NY, USA, 2014. ACM.

[88] Henning Pohl and Roderick Murray-Smith. Focused and casual interactions: Allowing users to vary their level of engagement. In *Proc. CHI'13*, pages 2223–2232. ACM, 2013.

[89] Alex Poole and Linden J. Ball. Eye tracking in human-computer interaction and usability research: Current status and future. In C. Ghaoui, editor, *Encyclopedia of Human-Computer Interaction*. Idea Group, Pennsylvania, 2005.

[90] Michael I. Posner. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25, 1980.

[91] Zachary Pousman and John Stasko. A taxonomy of ambient information systems: Four patterns of design. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, pages 67–74, New York, NY, USA, 2006. ACM.

[92] C. Prablanc, J. F. Echallier, E. Komilis, and M. Jeannerod. Optimal response of eye and hand motor systems in pointing at a visual target. *Biological Cybernetics*, 35(2):113–124, Nov 1979.

[93] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372, 1998.

[94] Johan Redström. Some notes on programme-experiment dialectics. In *Nordic Design Research Conference*, 2011.

[95] Claudia Roda. Human attention and its implications for human–computer interaction. In *Human attention in digital environments*, volume 1, pages 11–62. Cambridge University Press, 2011.

[96] Claudia Roda and Julie Thomas. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, 22(4):557 – 587, 2006. Attention aware systems.

[97] Federico Rossano. Gaze in conversation. In *The Handbook of Conversation Analysis*, pages 308–329. Wiley-Blackwell, 2012.

[98] Antti Salovaara, Antti Oulasvirta, and Giulio Jacucci. Evaluation of prototypes and the problem of possible futures. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2064–2077, New York, NY, USA, 2017. ACM.

[99] Andries F. Sanders. Information processing in the functional visual field, 1967.

[100] Dominik Schmidt, Florian Block, and Hans Gellersen. A comparison of direct and indirect multi-touch input for large surfaces. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I*, INTERACT '09, pages 582–594, Berlin, Heidelberg, 2009. Springer-Verlag.

[101] Julia Schwarz, Scott Hudson, Jennifer Mankoff, and Andrew D. Wilson. A framework for robust and flexible handling of inputs with uncertainty. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 47–56, New York, NY, USA, 2010. ACM.

[102] Julia Schwarz, Jennifer Mankoff, and Scott E. Hudson. An architecture for generating interactive feedback in probabilistic user interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2545–2554, New York, NY, USA, 2015. ACM.

[103] Steven C Seow. Information theoretic models of hci: a comparison of the hick-hyman law and fitts' law. *Human-Computer Interaction*, 20(3):315–352, 2005.

[104] Linda E. Sibert and Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 281–288, New York, NY, USA, 2000. ACM.

[105] Elizabeth Spelke, William Hirst, and Ulric Neisser. Skills of divided attention. *Cognition*, 4(3):215 – 230, 1976.

[106] Will Spijkers and Simone Spellerberg. On-line visual control of aiming movements? *Acta psychologica*, 90(1):333–348, 1995.

[107] Sophie Stellmach and Raimund Dachselt. Look & touch: gaze-supported target acquisition. In *Proc. CHI'12*, pages 2981–2990. ACM, 2012.

[108] Sophie Stellmach and Raimund Dachselt. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 285–294, New York, NY, USA, 2013. ACM.

[109] Yusuke Sugano, Xucong Zhang, and Andreas Bulling. Aggregaze: Collective estimation of audience attention on public displays. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 821–831, New York, NY, USA, 2016. ACM.

[110] Stephen P. Tarzia, Robert P. Dick, Peter A. Dinda, and Gokhan Memik. Sonar-based measurement of user presence and attention. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp '09, pages 89–92, New York, NY, USA, 2009. ACM.

[111] Benjamin W. Tatler, Mary M. Hayhoe, Michael F. Land, and Dana H. Ballard. Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5, 2011.

[112] Jan Theeuwes. Top–down and bottom–up control of visual selection. *Acta Psychologica*, 135(2):77–99, 2010.

[113] William Tolone, Gail-Joon Ahn, Tanusree Pai, and Seng-Phil Hong. Access control in collaborative systems. *ACM Comput. Surv.*, 37(1):29–41, March 2005.

[114] Michael Tomasello. Joint attention as social cognition. In Chris Moore and Philip J. Dunham, editors, *Joint attention: Its origins and role in development*, volume 103130, pages 103–130. Psychology Press, Taylor & Francis, 1995.

[115] Philip Tuddenham, David Kirk, and Shahram Izadi. Graspables revisited: Multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2223–2232, New York, NY, USA, 2010. ACM.

[116] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *INTERACT'2013*, pages 170–186. Springer, 2013.

[117] Roel Vertegaal. The gaze groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, page 294–301, New York, NY, USA, 1999. Association for Computing Machinery.

[118] Roel Vertegaal. Designing attentive interfaces. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 23–30, New York, NY, USA, 2002. ACM.

[119] Roel Vertegaal. Attentive user interfaces. *Communications of the ACM*, 46(3):30–33, March 2003.

[120] Roel Vertegaal, Jeffrey S. Shell, Daniel Chen, and Aadil Mamuji. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior*, 22(4):771 – 789, 2006. Attention aware systems.

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

[122] J. Williamson, R. Murray-Smith, B. Blankertz, M. Krauledat, and K.-R. Müller. Designing for uncertain, asymmetric control: Interaction design for brain–computer interfaces. *International Journal of Human-Computer Studies*, 67(10):827 – 841, 2009.

[123] John Williamson. *Continuous uncertain interaction*. PhD thesis, University of Glasgow, 2006.

[124] Jinglong Wu, Jiajia Yang, and Taichi Honda. Fitts' law holds for pointing movements under conditions of restricted visual feedback. *Human Movement Science*, 29(6):882 – 892, 2010.

[125] Pingmei Xu, Yusuke Sugano, and Andreas Bulling. Spatio-temporal modeling and prediction of visual attention in graphical user interfaces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3299–3310, New York, NY, USA, 2016. ACM.

[126] ByungIn Yoo, Jae-Joon Han, Changkyu Choi, Kwonju Yi, Sungjoo Suh, Dusik Park, and Changyeong Kim. 3d user interface combining gaze and hand gestures for large-scale display. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3709–3714, New York, NY, USA, 2010. ACM.

[127] Chen Yu, Linda B Smith, Hongwei Shen, Alfredo F Pereira, and Thomas Smith. Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, 1(2):141, 2009.

[128] Shumin Zhai. What's in the eyes for attentive input. *Commun. ACM*, 46(3):34–39, March 2003.

[129] Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI'99*, pages 246–253. ACM, 1999.

[130] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 493–502, New York, NY, USA, 2007. ACM.

Publication I

Barış Serim, Giulio Jacucci. Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5789-5800, May 2016.

Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input

Barış Serim[1,3], **Giulio Jacucci**[1,2]
[1] Helsinki Institute for Information Technology HIIT, Dept. of Computer Science, University of Helsinki
[2] Helsinki Institute for Information Technology HIIT, Aalto University
[3] Department of Design, Aalto University
firstname.lastname@hiit.fi

ABSTRACT
We propose using eye tracking to support interface use with decreased reliance on visual guidance. While the design of most graphical user interfaces take visual guidance during manual input for granted, eye tracking allows distinguishing between the cases when the manual input is conducted with or without guidance. We conceptualize the latter cases as input with uncertainty that require separate handling. We describe the design space of input handling by utilizing input resources available to the system, possible actions the system can realize and various feedback techniques for informing the user. We demonstrate the particular action mechanisms and feedback techniques through three applications we developed for touch interaction on a large screen. We conducted a two stage study of positional accuracy during target acquisition with varying visual guidance, to determine the selection range around a touch point due to positional uncertainty. We also conducted a qualitative evaluation of example applications with participants to identify perceived utility and hand eye coordination challenges while using interfaces with decreased visual guidance.

Author Keywords
Gaze input; eye tracking; multimodal interaction; uncertain input; interaction techniques; interactive surface

ACM Classification Keywords
H.5.2. Information interfaces and presentation: Input devices and strategies

INTRODUCTION
In HCI, terms such as eyes-on or eyes-free input are used to describe the degree of visual guidance an input action is performed with, in other words, the extent sight is used to guide action. Input actions vary regarding their degree of visual guidance. While typing on a physical keyboard can be conducted with little visual guidance, selecting items from a graphical interface often demands users to look where they are pointing to. Visual guidance of input actions gains particular importance with the use of eye tracking as a real time input for interaction. Examples of gaze input often feature gaze as a pointer for selection [30, 33, 41], assuming user's visual focus at the region of interest [3, 21]. Human visual attention, however, is a limited resource and there are a number of reasons to support input without extensive reliance on visual guidance:

- It can be desirable or necessary to remain visually focused at a certain region of interest without having to redirect gaze to another region in the interface.
- Input accuracy can be uncritical for certain cases, when the user is casual or wishes to delegate a certain level of control to the system.
- The task can require concurrent pointing at multiple regions of interest within the interface.

In this paper, we propose *using eye tracking to support manual input in the absence of or with little visual guidance*. The design of most graphical interfaces takes user's full visual guidance during manual input for granted. System interpretation of input is accordingly definitive; pointing actions on the interface are processed as exact coordinates. As an alternative to the current adoption, eye tracking can be used to understand the degree of visual guidance that a manual action is accomplished with and adapt the system interpretation and handling of the user input. Our main design strategy is to use manual input as a direct input and utilize gaze to increase

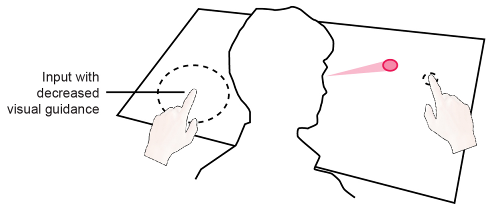


Figure 1. The user's manual input can be handled based on the degree of visual guidance it is conducted with. The input position is interpreted as exact when the action is realized eyes-on, while the system increases the potential selection range around touch and utilizes contextual resources and feedback techniques for input handling in the case of decreased visual guidance.

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

ACM 978-1-4503-3362-7/16/0515.00
DOI: http://dx.doi.org/10.1145/2858036.2858480

its expressiveness. This approach is fundamentally different than many current examples of gaze interaction that complement gaze with indirect manual input [21, 30, 33].

We make a number of contributions to support manual input with varying visual guidance. We conceptualize user input performed with decreased visual guidance as input with uncertainty and adopt an uncertain input handling framework for adapting system behavior. We describe action mechanisms and novel feedback techniques for handling such input and demonstrate their use through three example applications. We conducted a two part study to guide future design. In the first part, we determine the selection range of manual input on a large touch screen by measuring the positional offset with different degrees of visual guidance. In the second part, we report the perceived utility and the hand-eye coordination challenges that emerge during the interaction through a qualitative evaluation of the applications.

BACKGROUND

We motivate our design approach by discussing earlier work on gaze input, hand-eye coordination and input with uncertainty.

Gaze as Input Modality

Our work falls under the design approach that utilizes gaze as an additional modality rather than replacing manual input. Research in this direction aims to compensate the lack of a confirmation mechanism in gaze input (known as "Midas Touch" [16]) by using physical keyboard [19], mouse [41], touch [30] or gesture.

Previous studies conducted in controlled, isolated settings show that gaze can be faster than other pointing devices for target selection [28, 34]. Thus, a strong motivation for most previous work that combine gaze and manual input has been motor performance gains in target acquisition [4, 18, 21, 30, 31, 33, 41]. A pioneering application is Zhai et al.'s MAGIC pointing [41], a manual and gaze hybrid pointing method, that eliminates part of the mouse movement by warping the mouse cursor to the eye fixation coordinates and then accomplishes the selection action through the mouse, thus cascading the two input modalities. Recently, interaction with large and distant displays, where direct input is impractical, has been an application area for utilizing gaze. In these applications gaze is complemented by touch input on a hand-held device [30, 31, 33] or free air gestures [18].

In general, previous work capitalizes on the rapid switching of spatial context afforded by gaze to decrease the amplitude of movement by hand. Thus, a common feature among them is the separation of the hand from the target, namely the indirect and relative use of manual input to complement the absolute coordinates provided by gaze. Two hybrid exceptions are GazeTouch [21] and Gaze-Shifting [22] that utilize manual input both as a direct and indirect input, based on the distance of gaze point to the input position.

Some of the examples cited above are similar to our approach in that they facilitate manual input without visual guidance. This is achieved through different means, such as using touch as a relative, indirect input [21] or in small handheld devices that enable eyes-free interaction [31]. Our approach departs from them by *always using manual input directly*, even in the case of input without visual guidance. We use manual input for selection and use gaze input to qualify manual input. While earlier work advocates the separation of the hand from the target, summarized as *"gaze suggests and touch confirms"* [30] or *"gaze selects, touch manipulates"* [21], we propose an alternative use in which *"gaze qualifies hand input."*

Hand-Eye Coordination

Our approach is partly motivated by the simultaneous use of gaze and manual input on multiple points of interest. Previous work in eye cursor coordination in web search shows that mouse use is not purely *incidental*, (i.e. performed for the purpose of clicking) [24]. Instead, the cursor can be used for other purposes such as keeping track of what is read and as a placeholder on interesting items, while eyes switch to other regions [15, 24]. Additionally, as Bieg et al. [3] argue, one assumption in techniques that aim to decrease the amplitude in target acquisition using gaze is that eye movement precedes pointing actions. Contrary to this, their study reports that pointing behavior is initiated without visual guidance for items whose approximate locations are known.

In the above described situations gaze and pointing accomplish parallel tasks in different regions within interface. However, such parallel use of eye and cursor movements might not be well supported by design approaches that cascade (i.e., sequence) manual and gaze input such as MAGIC pointing [41].

While approaches like MAGIC pointing focus on increasing the performance in a sequential set of actions using gaze, we target supporting concurrent access to multiple regions on the interface, without necessarily redirecting gaze. A usable distinction has been made by Fitzmaurice et al. [9, 10] between spatial and time multiplexing for user input. While time multiplexing refers to sequential and mutually exclusive techniques, spatial multiplexing refers to the concurrent access to dedicated input fields. Their observation of manual interaction with domino bricks is illustrative of spatial multiplexing: *"...Tactile feedback was often used to grab dominos while visually attending to other tasks. The non-dominant hand was often used to reposition and align the dominos into their final resting place while, in parallel, the dominant hand was used to retrieve new dominos..."* [10]. Fitzmaurice et al., accordingly, design for spatial multiplexing through graspable input devices by citing the benefits of tactile confirmation and possible use without visual guidance.

On the other hand, the advent of multi-touch devices enabled spatial multiplexing in graphical interfaces. Even though interaction with tactile interfaces has shown to be more robust and efficient [32, 37], multi-touch input surfaces allow similar benefits like bimanualism. Previous work on touch screens aims to support eyes-free interaction in various ways such as using touch as a gestural input or directing finger to predefined locations using magnetic attraction [36]. In contrast, we support eyes-free interaction through appropriate interpretation and handling of the user input. We interpret lack of visual guidance as situations of inputs with uncertainty.

Input with Uncertainty

Proliferation of inherently uncertain inputs, such as speech recognition, gestures and touch, motivated a number of techniques and frameworks for the flexible handling of user input and communicating system interpretation of input back to the user [20, 23, 25, 26, 38, 39].

A large body of research on input positional uncertainty deals with the "fat finger problem", namely the large touch contact area and visual occlusion caused by the finger [2, 26, 35, 38]. In the context of this paper, the source of uncertainty is the user's lack of exact information about how his/her manual input coordinates map to the visual content on the user interface *due to decreased visual guidance*. The most closely related work in this direction is by Hagiya and Kato [14], who use gaze point information to model touch distribution on a hand-size mobile display. Although particularly focused on text entry, their distinction between accurate and ambiguous touch is parallel to our approach. Different from their work, we consider the overall design space of input handling and demonstrate their use in diverse applications with multi-touch interaction on a large screen. Previous work on the accuracy of target acquisition using arm movements without visual guidance suggests that errors increase in relation to the amplitude of movement [5, 29] and cumulatively [6].

Users' lack of information about their exact input region allows system to interpret the input as positionally ambiguous and less decisive. Conversely, high visual guidance reinforces user's manual input. In HCI various design frameworks aim to adapt system behaviour depending on user's varying degrees of control. In vehicle design, "horse metaphor" [12] refers to a level of delegation of decision making to the system, based on how tight or loose the user's control is. Pohl and Murray-Smith propose design approaches for mobile systems that allow users to vary their level of engagement along a focused-casual continuum [23]. When user input is casual (i.e., lacking in precision and deliberation) the system partly takes control using available personal and contextual information. In the same spirit, we use the degree of visual guidance for a partial delegation of decision making to the system. However, while user attention is inhibited or reserved for another activity in mobile use [23], we are primarily interested in the cases in which user attention is divided between two actions related to the same task and two regions within the same interface. This enables using gaze position on the interface as a resource for interpreting the user's input and providing feedback to the user through various channels.

DESIGN SPACE

Previous work [20, 25] on handling input with uncertainty separates handling process into successive stages of modeling input, event dispatch, interpretation and action, in which the system component "mediator" is responsible for deciding on the action. We used a similar structure and provide an inexhaustive list of considerations and techniques that are particularly relevant for handling input with varying degrees of visual guidance.

User input involves both user manual input position and other contextual information. The system handles user input through various *action mechanisms* (i.e. select, defer or inaction). *Feedback techniques* aim to remedy users' lack of visual guidance by making manual input information and system interpretation of input available to the user.

User input

Manual input position. The primary resource for interpreting user's manual input is the position (such as x,y values) of the input. Potential selection range around a manual input position increases with decreasing visual guidance (Figure 2), due to positional uncertainty. We operationalized visual guidance as the distance between the gaze and manual input position and use it to compute the potential selection range around touch input. The interpretation of manual input is exact up to a certain distance threshold between gaze and input position. Beyond this threshold, the selection range increases in linear relation to the distance between gaze and manual input position. For touch input the selection range is greater than a single pixel, even when it is conducted with visual guidance, due to the inherent uncertainty of touch. During the design process we heuristically defined the threshold and linear relation values. A two stage target acquisition study described further in the paper shows how the threshold and linear relation can be empirically determined.

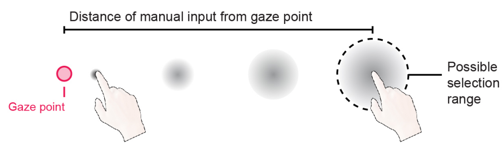


Figure 2. Input selection range increases the further gaze point is located from the manual input position (where the touch or cursor is located).

Additionally, the determination of visual guidance is dependent on a number of design decisions:

- **Continuous, discrete.** For a manual input event, visual guidance can be determined along a discrete (such as only covering the two opposite ends eyes-on and eyes-free) or continuous scale.
- **Conservative, liberal approaches.** The determination of visual guidance and thus the selection range around an input position can change upon eye movements (liberal approach) or only upon the movement of the hand (conservative approach). We borrow the terms from Zhai et al. [41], who used them to distinguish the cases in which a mouse cursor is continuously warped to the gaze point coordinates (liberal) or only upon a cursor movement event (conservative). For many cases, conservative approach can be more suitable, since elements pointed with visual guidance will persist in the user's short term memory even after the gaze shifts to another location.

  When scaling the range of uncertainty upon movement, our general principle is to a) decrease the uncertainty *instantly* when the user increases visual guidance and b) increase the uncertainty *gradually* when the user decreases the visual guidance. The difference is due to the gradual deviation in position with increasing amplitude of movement [5]. It should be noted that for cursor input (e.g., mouse), the uncertainty can always be determined upon movement as the cursor is always present at the interface. On the other hand, touch input involves finger enter and exit events.

Gaze context. Interpersonal interaction can involve "referential gaze" in combination with speech and manual pointing to ground and disambiguate meaning. The use of "dual pointing" to two different regions of interest in the environment, one with gaze the other with the hand, to semantically associate them, has been documented [13]. Similarly, eye tracking information can be used beyond determining the positional uncertainty of manual input. The information of where the user is gazing at the interface can be used to resolve uncertainty in manual input by prioritizing actions that are related to the gaze context.

There are multiple possibilities regarding how the gaze context can be determined in relation to manual input. First, the gaze context position can be determined at the beginning of manual interaction and remain fixed until the user ends the manual input (such as by releasing a mouse or in a touch up event). Second, the gaze context position can be constantly updated upon eye movements. However, during the design process we noticed that continuous synchronization of gaze context with eye movements can be intrusive and unstable. A more viable option is to update gaze context only upon touch or cursor movements.

Interaction history. Another resource for resolving positional uncertainty is the interaction history of the user. A possible reason for the lack of visual guidance and loose hand-eye coordination could be that the location of the item is remembered [11]. Thus, decreased visual guidance can be attributed to the user expectation of repeating a previous action. The system can also resolve uncertainty by keeping track of how user changes the application state. Application functionality makes certain action sequences more probable over others in an interface configuration. As an example, if the user previously opened a dropdown menu, it is more probable that a selection action on one of the menu items will follow.

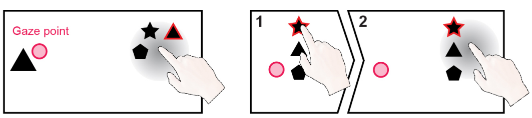


Figure 3. Two possible ways of resolving positional uncertainty are using gaze context (left) and interaction history (right).

Action Mechanisms

Select action. The system can respond to positional uncertainty in a number of ways for selecting action (Figure 4). One is positional selection between different actions, such as selecting between different discrete input fields like buttons. The selection can also occur between different actions that positionally overlap. For example, a touch action on a text field can be intended for scrolling or text selection [26]. These different actions require different degrees of visual guidance: scrolling has an area effect and does not require exact pointing, while selection requires accurate pointing. Decreased visual guidance in such cases can aid the system decision making between various actions types. Finally, if the input field allows range selection, positional uncertainty can be handled by expanding the selection range.

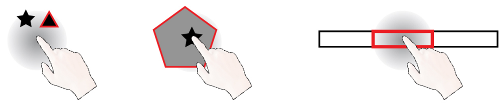


Figure 4. Action selection can involve positional selection (left), selecting between actions that positionally overlap (middle) or range selection (right).

Defer action. Another potential response is to defer action until enough information is gathered for disambiguation. A common example is the press-release sequence for inherently uncertain inputs such as touch [25] or gaze [19]. System interpretation of user command is communicated as a feedback after key or touch press event and the final action is deferred to a key or touch release event.

Inaction. Input without visual guidance can be interpreted as unintentional or unfocused, resulting in the system not taking action.

Feedback Techniques

In addition to possible non-visual notifications such as sound or tactile feedback, eye tracking allows various visual strategies to provide feedback about user actions and the system interpretation of them:

Support peripheral awareness. In the case of manual input without visual guidance the system can remedy the lack of information by supporting peripheral awareness. The system can increase the visual footprint of the cursor a) to support the peripheral awareness of where the cursor or finger is located within interface and b) to indicate the degree of positional uncertainty as determined by the system (Figure 5). The visual footprint of potential targets can also be increased, informing users about the system interpretation of their action.

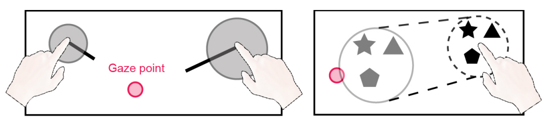


Figure 5. Providing peripheral awareness (left) and warping information content around manual input position to gaze point (right) are two possible visual feedback techniques to communicate system interpretation of user input back to user.

Warp information to gaze point. Warping information to the gaze point is the counterpart of warping cursor to the gaze point location (e.g. [41]). The information content around the user's manual input position or the system interpretation of user action is overlaid to where the user's gaze is directed (Figure 5).

Figure 6 provides a summary of the design space. It should be noted that providing feedback and uncertain input handling are two competing approaches, since feedback techniques decrease the uncertainty by providing information to the user. However, they are not mutually exclusive and can be integrated in various stages of interaction as can be seen in the applications described further in the paper.

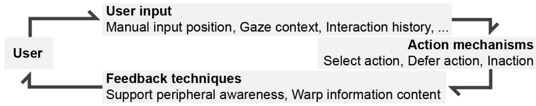


Figure 6. Summary of the design space.

DETERMINING SELECTION RANGE

An important design question related to positional uncertainty is determining selection range for input with varying levels of visual guidance.

We conducted a two stage study of target acquisition to determine the potential selection range around a touch point. 12 participants (4 female) aged between 20 to 34 (m=28, sd=3.95) took part in our evaluation. Each session started with 9 point eye tracking calibration and proceeded if the calibration was successful (below 2° deviation in accuracy). The height of the table on which the screen and eye tracker were mounted was adjusted for each participant and the participants remained standing during the evaluation.

Apparatus

The study has been conducted using a 10 finger multi-touch screen (27", 2,560x1,440 pixels) combined with an SMI RED eye tracker that is positioned below the touch screen running at 60Hz. The screen was tilted 30° to enable easier hand reach for touch input. The screen and eye tracker were positioned respectively 50cm and 70cm in front of the eyes (approximate values).

Study Design

Each participant performed two set of tasks. The first set of tasks aimed to determine positional inaccuracy (variable) for varying distances between gaze and target (invariable). The second task set aimed to determine the distance between gaze and the target (variable) for accurate pointing tasks (invariable).

We logged gaze, touch and target positions for each task. During the study there were brief moments when eye tracking signal was not available due to hand occlusion or head movement. Thus, a task was completed only when the gaze point was available to the system.

Position Inaccuracy for Gaze Distance

Earlier work suggests a decreased positional accuracy for motor target acquisition without visual guidance [5, 29]. At the first stage, we aimed to determine the positional uncertainty for target acquisition in varying degrees of visual guidance, which we operationalized as *the distance between the target and where the gaze is directed.*

To complete a target acquisition task, the participants had to keep their gaze (controlled by eye tracking) inside a circle while tapping on one of the 15 targets (on a 5×3 matrix) on the touch screen (Figure 7). The participants were instructed to tap as correctly and as fast as possible to determine the positional offset for acquisition. We used the 6 lower middle points within the matrix, where the eye tracking is most accurate, as gaze fixation points. While keeping their gaze within the defined area, the participants tapped on all the defined targets on the matrix at a randomized order. The target acquisition tasks were accepted only if the participants kept their gaze within the circle (indicated to the participants by changing the circle area to green).

To prevent giving any visual cues, the entire target matrix was visible during the tasks. However, information of which target to tap on was shown within the circle. The pairing of 6 eye fixation regions with 15 target positions resulted in 90 tasks, that show varying distances between the target and the gaze point. To prevent the misidentification of the target, each column was assigned a different shape (from left to right: circle, cross, triangle, square and pentagon).

The degree of visual guidance as we operationalized in the study is not easily comparable to index of difficulty in Fitts' law. First, the amplitude of motion is not dependent on the distance of the target from the gaze, since the participants initiated the movement from the previous target location in the matrix. Second, the visual boundary of the targets does not accurately represent target width as the system did not require the participants to touch on the exact position.

At the same time Fitts' law has implications for acquisition with restricted visual guidance. It has been shown that Fitts' law is valid for restricted visual guidance on the target or hand [40]. By increasing the selection range for manual input, we increase the target width and thus decrease the index of difficulty. How the increased target width compensates the lack of visual guidance for acquisition performance is a highly relevant question for future research. However, we limit the scope of this study to determining the selection range around touch point.

Figure 7. The experimental screen (left) and the close up of the circle (rad = 52.5mm) in which the participants need to keep their gaze inside (right). The target is shown in red inside the circle, while the position of the circle is shown in green.

Gaze Distance for Accurate Acquisition

While the first stage aimed to understand the positional inaccuracy with varying visual guidance, second stage aimed to understand the distance of gaze from the target for accurate pointing. Participants were instructed to touch circular targets, without any constraints on where they look. This stage forced participants to be accurate since a task was considered complete only when the touch point fell within the circular target (rad = 5.8mm). Each participant completed 3 repetitions of 6 target acquisition tasks (randomized order).

Results

The first stage yielded 1080 trials from 12 participants (× 90 tasks). The scatter plot in Figure 8 shows the relationship between the distance of the gaze point to the target position (invariable) to the positional offset (distance between the touch and target positions). The outliers in the scatter plot refer to the trials in which there was a large positional offset (M=120.8 mm, sd = 13.4), but the touch point was close to an adjacent target on the matrix (M=19 mm, sd = 8.8). We categorize these 31 outliers as cases in which participants misidentified the target, and exclude them from the analysis.

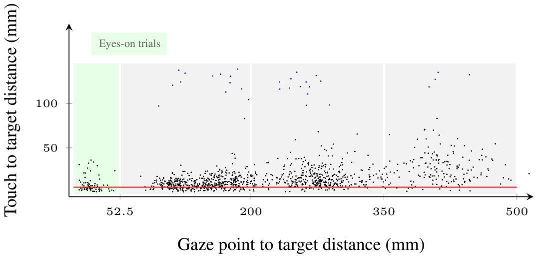


Figure 8. Scatter plot of peripheral target acquisition tasks across all participants. Outliers are shown in blue. Horizontal red line indicates the visual boundary of the circle target (rad=5.8mm). Green background indicates the eyes-on in which the target was within the boundary of the circle the participants had to keep their gaze inside(rad=52.5mm).

We divide the data in four continuous bins that correspond to varying levels of visual guidance. The intervals of the first bin (0-52.5) were determined by eyes-on tasks, in which the distance between the gaze point and the target are smaller than the radius of the circle in which the participants had to keep their gaze inside. We divided the rest of the data in three bins of even intervals. Figure 9 shows the distribution of touch points relative to the target across all users for four chosen levels of visual guidance. The deviation in the distance between gaze point and target is due to the large diameter of circle in which the participants need to keep their gaze within. For varying visual guidance levels, 95% confidence values for positional offset can be used to determine touch selection range. The results (Table 1) suggest an increasing positional inaccuracy with increasing distance between gaze point and target.

An unusual result from the first stage is the very large deviation for eyes-on tasks (0-52.5mm) when compared to a previous study [2] that reports an accuracy rate higher than 95% for 5mm radius target acquisition. We relate the unusual result to the experimental design of the first stage, in which participants did not have to touch within the target to complete the task. This is in contrast to the second stage, in which participants had to touch within the target visual border. In 216 total trials gathered from the second stage, 95% of gaze

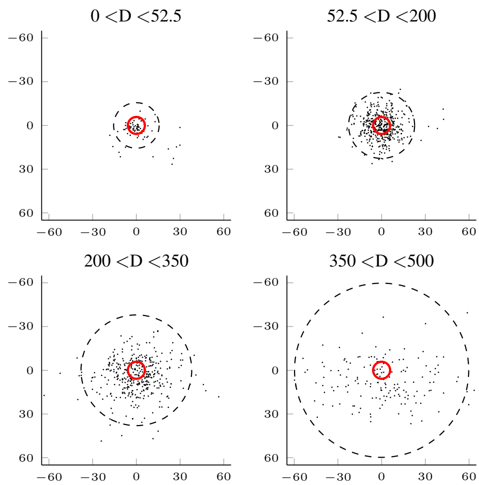


Figure 9. The distribution of touch points relative to the target across all users for chosen ranges of distance between gaze point and target. The dashed circles are the 95 % confidence circles. The red circles show the target visual boundary. All units in mm.

Distance range (Gaze to Target)	Mean distance (Gaze to Target)	95% confidence (Touch offset)
0-52.5mm	20.7mm (sd=9.3)	15.6mm
52.5-200mm	146.0mm (sd=29.6)	22.6mm
200-350mm	272.8mm (sd=33.1)	37.9mm
350-500mm	414.8mm (sd=32.1)	56.7mm

Table 1. 95% confidence values of positional offset for distance between touch and target for different visual guidance ranges. The range 0-52.5mm represent eyes-on tasks.

points were within a 61.0mm radius range around the target (M=26.3mm, sd=20.0).

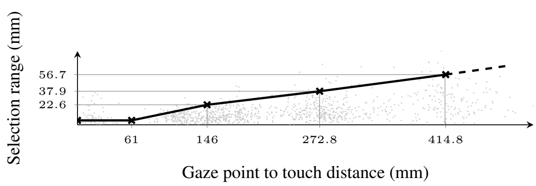


Figure 10. Selection range profile showing 95% confidence ranges from the first and second stages. Note that the minimum selection range on y axis does not need to start from a single pixel width for inputs that are inherently uncertain, such as touch.

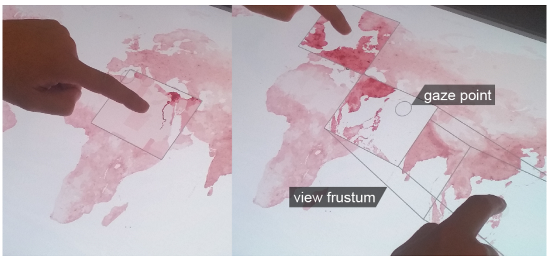
We use the 95% confidence value (61.0mm) for the distance between gaze and target in accurate pointing tasks as a threshold for increasing positional uncertainty due to decreased visual guidance. Together with the values from the peripheral target acquisition tasks at the first stage (Table 1), we plot a tentative profile of selection range with varying degrees of guidance (Figure 10). We should stress that the profile is intended as a provisional design guide. More importantly, not every potential action within the input selection range should be given the same weight for input handling. We consider a discrete confidence threshold to be most useful for cases that require making the selection range visually explicit to the user.

APPLICATIONS

We developed a number of applications to demonstrate the applicability of our design approach for a variety of use cases. The applications feature different combinations of input handling components. Below, we provide a conceptual breakdown of each application in terms of user input, action mechanisms and feedback techniques described above.

Application 1: Multifocus Image Exploration

A potential application case is multifocus interfaces (e.g., [7, 8]) that involve spatial juxtaposition of multiple points of interest. As opposed to time multiplexed methods such as zooming, multifocus interaction utilizes spatial multiplexing to display information [7]. A common aim in juxtaposition is to compare and correlate between multiple foci [8] and avoid redirecting gaze over long distances. The process of declaring multiple foci can be sequential or concurrent (e.g. using multi-touch). Gaze input can be a useful addition to multifocus interaction tasks, both as a focus point and for supporting input with decreased visual guidance.



User input: Manual input position (discrete, conservative), Interaction history (previously zoomed in regions are used to resolve positional ambiguity)

Action mechanisms: Select action (resolve positional uncertainty, increase selection range by zooming out)

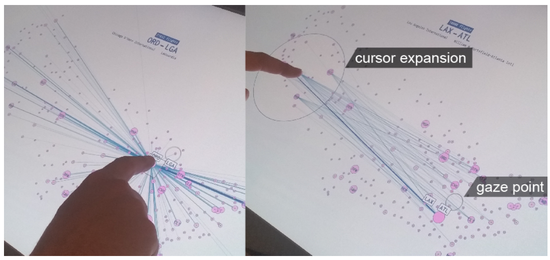
Feedback techniques: Peripheral awareness (show view frustum from finger to the gaze context), warp information content (warp lens near the primary touch point)

Figure 11. The degree of visual guidance is used to determine the position of lenses in image exploration. In the case of a touch event with visual guidance, the lens is shown at the touch position (left). In the case of a touch event with decreased visual guidance, the lens is warped near an existing lens (right).

Here, we demonstrate the use of input without visual guidance for exploring a map image that shows world population density (Figure 11). The application allows creating multiple lenses that are aligned edge to edge and controlled by individual touch points. The degree of visual guidance on different touch events is used to determine the primary touch point near the other lenses are aligned by its edge. The primary touch point is reevaluated with each touch down event. In the case of a touch event with visual guidance, the other lenses are warped to the new touch location. In the case of a touch event with decreased visual guidance, the lens is warped near the primary lens that the user's gaze is directed. In the latter case, positional uncertainty is handled in two ways. First, the lens covers an increased areal range. Second, positional uncertainty can be resolved by using interaction history by zooming into a previously viewed location.

Application 2: Exploring Relational Data

Another multifocus application case is interaction with relational data. We created a geospatial visualization of flight connections in the US (Figure 12). Interaction with the graph allows filtering flight connections based on the airports near the manual input. The degree of visual guidance is used to determine the positional uncertainty of touch points. For touch actions with high visual guidance, the application visualizes the connections from a single node that the user is pointing to. Positional uncertainty is increased and the cursor is expanded for pointing actions with decreased visual guidance. In this case, the gaze context of the user is used to resolve positional uncertainty; only the connections between manual selection range and airports near gaze point are visualized. In contrast to multifocus image application, the gaze context is not associated with an existing touch point and updates continuously with the movement of the touch.



User input: Manual input position (continuous, conservative), Gaze context (updates based on manual movement on touch surface, resolve positional uncertainty using relational data)

Action mechanisms: Select action (resolve positional uncertainty)

Feedback techniques: Peripheral awareness (expand the cursor to show positional uncertainty), Warp information content (show airport code on the gaze context)

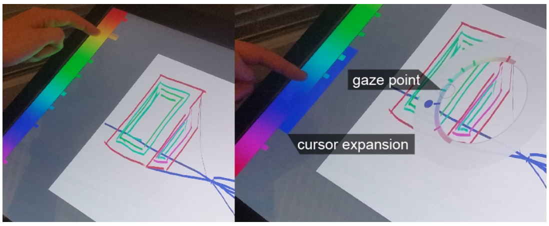
Figure 12. If the manual positions are interpreted as exact, the application visualizes all the connections from a single node (left). If the user gaze is directed elsewhere at the graph, the cursor is expanded to show increased positional uncertainty (right).

Application 3: Color Switching in Paint

Many interfaces involve sequences of tool switching and manipulation actions. Tool switching can be realized using toolbars, keyboard shortcuts or in-place selection techniques such as pie menus. Direct input devices such as touch screens allow bimanual action, where one hand performs tool selection while the other manipulates the target. In these cases, visual guidance can be used to qualify tool selection and manipulation actions. In this application we take color selection and painting on a canvas as an example of pair actions. Our application features two input fields: a continuous color selection bar and a virtual painting canvas.

Input field: Color palette

The color palette is configured as a vertical bar with varying hues along the y axis (Figure 13). The degree of visual guidance is determined continuously, corresponding to increased positional uncertainty for selection. In the case of selection with visual guidance, the system only considers the touch position for selecting a hue. In case of decreased visual guidance, the positional uncertainty is increased and the system uses additional user input resources for selection, namely the previous color selections (interaction history) and the colors in the gaze context of the user. The actual color selection is deferred to a touch release event. In the meantime, feedback of selection is displayed to the user through a radial color palette that appears on the gaze point. The gaze context is fixed at the start of a touch event and remains constant until the touch is released.



User input: Manual input position(continuous, conservative), Gaze context (the colors on the canvas region that the user looks at are used to resolve uncertainty for selection), Interaction history (previously used colors are used to resolve uncertainty)

Action mechanisms: Select action (resolve positional uncertainty), Defer action (defer the actual color selection to touch up event)

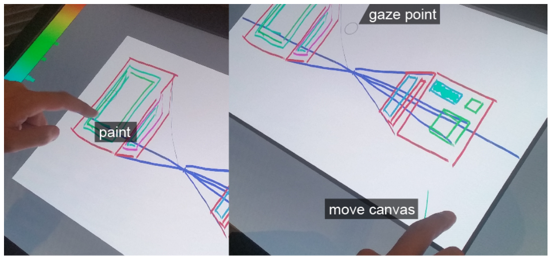
Feedback techniques: Peripheral awareness (increase the cursor size to show positional uncertainty), warp information content (create a color palette around gaze the context)

Figure 13. When conducted with a high degree of visual guidance the manual input is interpreted as exact (left). If the user gaze is directed elsewhere, such as on the canvas, the positional uncertainty is increased. In this case system provides feedback by visualizing a radial color palette at the gaze point (right).

Input field: Canvas
For input on canvas (Figure 14), the degree of visual guidance is determined discretely and is used to select action type. The degree of visual guidance at the moment of touch is used to determine if the action type is intended for painting (requiring fine degree of control) or moving the canvas (has area visual effect, thus requiring less visual guidance). The input type associated with a touch point remains stable until the touch is released.

EVALUATION OF APPLICATIONS
The various applications we developed combined different input handling techniques. We evaluated different applications to investigate 1) possible hand-eye coordination challenges that are general for input with decreased visual guidance and 2) the particular interaction challenges related to the action mechanisms and feedback techniques that vary among applications.

After finishing the target acquisition tasks, the participants proceeded into using the three applications. They were asked to perform open ended tasks with the example applications



User input: Manual input position (discreet)

Action mechanisms: Select action (between types "paint" or "move")

Figure 14. Touch event on a canvas can be interpreted either as a paint (left) or move the canvas function (right) depending on the degree of visual guidance.

until they felt comfortable using the system (approximately 5 minutes). While our design approach does not require gaze point to be made explicit to the user, we still visualized the gaze as a translucent gray ring to inform participants in case the system loses track of their gaze. In this case, the translucent ring turned opaque, warning participants to correct their posture.

We video recorded participant interactions and collected their feedback after using each application. We also interviewed the participants at the end of the study to gather their overall feedback. In this section we report the participant feedback and observational data.

Participant Feedback & Video Analysis

Adjustment through Use
A common reaction among participants was the reported difficulty of "touching without looking" at the start of the session followed by gradual adjustment. Words "unnatural", "unintuitive", "strange" were often used to describe the initial experience, while the participants described their later experience as "natural" and "easier". Deliberately "avoiding drifting" of gaze to the touch location was observed during interaction and was also reported by participants as one reason for initial difficulty. *"At first it was of course quiet strange, pointing to a place where you can't see and your gaze tries to go there and you try to use your peripheral vision, but that also gets easier as you use it, you get used to the feeling of touching somewhere that you don't see."* (P4). The experience has also been compared to typing on a keyboard: *"... you start to write in the keyboard without looking, initially you look but you can try to do without looking..."* (P1)

Some participants described input handling mechanisms as "forgiving" and assistive of eyes-free input. *"I don't trust at all what I am seeing in my peripheral vision... knowing that it (eye tracking) being taken into account I trusted it even more and could predict and had some expectation of what will happen"* (P12). while others highlighted the need to know *"exactly where everything is"* before being able to point without looking. Although participant reaction differed regarding the degree of proficiency needed, individual confidence during eyes-free or peripheral pointing was a common dimension of use experience.

Being able to concentrate on the task such as drawing and *"using peripheral vision to do other specific tasks that are very obvious"* were highlighted as benefits. One participant also reported that gradual adjustment was useful for *"using two hands"*.

Gaze as Additional Pointer

The three applications are different regarding how explicitly they use the gaze context and how gaze updates in relation to touch events. The participant feedback helped identify potential benefits and drawbacks of different ways of using the gaze context. Explicit use of the gaze context has been welcomed as an additional "third hand" in flight visualization application and enabled concurrent access to three different locations (Figure 15). *"So the same feature can be done with two hands... but then I realized we only have two hands so maybe some possibility could be use your gaze as a third hand."* (P1)

Figure 15. Gaze being used as a third pointer in addition to two hands.

On the other hand, the use of gaze as an additional pointer, especially with dynamic update, caused hand eye coordination challenges. An additional challenge is the difference between the system and user interpretation of gaze context. Participants compared between the use of gaze in the multifocus image and flight visualization applications. *"I think it is easier to use it updating in a static way, because there is nothing that constantly change, I can compare more easily, there is nothing unexpected, but in dynamic I had more options, was good that it updates fast where I look, but then it was loosing..."* (P9).

Misinterpretation of Positional Uncertainty

We observed a number of instances in which a manual input action was wrongly interpreted as positionally uncertain due to the system's lack of awareness of the movement before the actual touch event. The instances usually involved the participant keeping his finger just above a specific point on the interface and performing the touch action while looking elsewhere (Figure 16). In these cases, although the participants knew exactly where they were pointing to, the application interpreted it as positionally uncertain and handled accordingly. The participant occasionally identified this as a "problem".

Screen Edge as Ambiguous Border and Tactile Guide

Although a touch screen is an input field with definite boundaries, decreased visual guidance can cause ambiguity for users regarding whether they are *addressing* the system during touch. In some instances, while aiming for the color palette near the edge of the screen, participants touched the insensitive bezel area of the screen (Figure 17). The lack

Figure 16. Interaction sequence leading to misinterpretation of positional uncertainty. The participant placed his right index finger on a region on the image (1). After a brief look, he lifted his right index finger from the touch screen but held it just above the surface (2), while pointing to another location with his left index finger (3). This was followed by a touch on the same point with the right index finger (4).

of visual feedback (color palette warped to the gaze point) in this case communicated that the system is not addressed, which led the users to a repeated touch action. On the other hand, device borders provide potential tactile cues for eyes-free use. This was observed again for selecting colors, when participants *anchored* their left hand on the screen edge for sliding along the color palette with their thumb or index finger while keeping their gaze on the canvas (Figure 17).

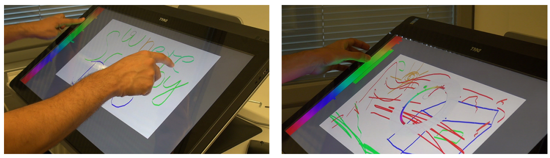


Figure 17. Instances of tapping on the bezel rather than the display area (left) and using the screen edge as a tactile guide (right).

DISCUSSION

We presented a design approach that targets supporting manual input with decreased visual guidance. Informed by previous literature [5, 14, 29], our main assumption was an increased positional uncertainty for input with decreased visual guidance. The first part of evaluation confirmed our assumption and provided a tentative profile for scaling selection range for different levels of visual guidance. The conceptualization of input with decreased visual guidance as input with uncertainty led to the design of various input handling and feedback techniques. The feedback we gathered from the participants during evaluation provides evidence for the viability of input with decreased visual guidance, although cognitive challenges related to hand-eye coordination and confidence during input were reported. Moreover, some such as addressing challenges and misinterpretation problems are mainly communicative challenges that might emerge in sensing or adaptive interfaces [1, 17, 27].

While we formulated manual input with decreased visual guidance as a design motivation, the specific benefits like increased functionality or satisfaction depends on a number of contextual factors such as physical setup and task. Manual input with decreased visual guidance can be forced by application context or preferred by the user. Similarly, various feedback techniques can be necessary when decreased visual guidance is forced by application context while they are not as essential when the users are able to direct their gaze to the input location.

At this point, we would like to discuss what we learned along the design process and evaluation through the analytical lens of spatial multiplexis [9]. Tuddenham et al. [32] limit the use of *bimanualism* to "two-handed one-object interaction", while using the terms *concurrent unimanualism* and *lateral sequential unimanualism* for "two-handed two-object interaction". The distinction is highly relevant regarding the degree of visual guidance during input. We argue that the benefit of supporting manual input with decreased guidance is most valid for multiple object interactions, since these cases involve direct and concurrent access to multiple interface regions. In painting application, this was observed when participants switched between colors with left hand and painted with their right hand (lateral sequential unimanualism) without redirecting their gaze. For multifocus image exploration and flight visualization the main pattern of interaction was concurrent unimanualism.

At the same time, gaze input requires revisiting the scope of space multiplexed user input. In their seminal paper, Fitzmaurice and Buxton investigate space-multiplexis through parallel use of hands [9]. However, the parallel use of manual input and eyes on different regions of interest [3, 15, 24] suggests that the scope of spatial multiplexis can be extended to the concurrent use of visual perception and input actions at different regions. In addition, spatial multiplexis can be extended to the interactions in which gaze is used not only for perception but also as a pointer. Many recent examples that combine gaze with indirect touch can be described as "one-hand+gaze one-object" or "two-hand+gaze one-object" interactions [21, 30, 33]. In contrast, during evaluation sessions, participants interpreted the explicit use of the gaze context as an additional "third hand", and performed "one-hand+gaze two-object", or "two-hand+gaze three-object" interactions.

Finally, gaze input requires revisiting the rationale of multiple object, space multiplexed input. Unlike multi-touch, gaze is a single channel but a very rapid input. Thus, using gaze coordinates for selection might favor sequential interaction over simultaneous selection of multiple targets, as in the case of sequential multiple target acquisition by gaze [21]. In this paper, we presented a design approach that targets multiple object interaction by supporting the use of touch with varying visual guidance as a direct input. However, further research is needed to evaluate the drawbacks and benefits of both approaches for different applications.

Limitations

Determining Visual Guidance

In the experimental design and applications we operationalized visual guidance as the distance of gaze point from the touch position at the moment of touch. This does not account for the complex hand-eye coordination over time that leads to a touch event: the motor movements can be accompanied by different levels of visual guidance between the initiation of the movement and touch. Planar input on the screen alone is not always sufficient for sensing this coordination, occasionally resulting in misinterpretation of positional uncertainty as reported above. A potential solution is over-the-screen sensing of the hand and finger movement to increase resources available to the system. The problem is not as significant for mouse or other cursor based movement, since the cursor position information is always available to the system.

Stability of Input Field

Our design approach assumes the stability of the input field. Keyboard, fixed toolbars and geospatial data are relatively stable input fields, which users either have prior knowledge of or get accustomed to. However, more information is needed to determine the viable scope of performing manual input with decreased visual guidance.

Non-planar Surfaces

It should also be noted that we limited our scope of investigation to a single planar interactive surface. However, our design approach can be applied to a range of settings that are spatially more diverse, such as distributed, on-body or virtual interfaces. These settings raise a number of questions such as determining positional uncertainty and solving the addressing problems described earlier.

CONCLUSION

Potential ubiquity of eye tracking in the near future calls for reevaluating the division of labor between different input modalities and the role of gaze within. In this paper, we contributed to the ongoing discussion on how gaze can be integrated with other modalities. Our proposed design approach utilizes gaze to qualify direct manual input by taking into account the level of visual guidance the input is performed with. In the case of input with visual guidance the system allows familiar interaction, while the input with decreased visual guidance is supported through various action mechanisms and feedback techniques for handling input with uncertainty. Adaptive handling of input, in return, supports concurrent access to multiple locations in an interface. We consider the design space we developed as a starting point for a systematic exploration of interfaces that adapt manual input handling in relation to visual guidance.

ACKNOWLEDGMENTS

This work has been partly supported by MindSee (FP7 ICT; Grant Agreement #611570). We would like to thank the anonymous reviewers for their insightful comments and to the participants of the study.

REFERENCES

1. Bellotti, V., Back, M., Edwards, W. K., Grinter, R. E., Henderson, A., and Lopes, C. Making sense of sensing systems: five questions for designers and researchers. In *Proc. CHI'02*, ACM (2002), 415–422.
2. Bi, X., and Zhai, S. Bayesian touch: a statistical criterion of target selection with finger touch. In *Proc. UIST'13*, ACM (2013), 51–60.
3. Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., and Bülthoff, H. H. Eye and pointer coordination in search and selection tasks. In *Proc. ETRA'10*, ACM (2010), 89–92.
4. Blanch, R., and Ortega, M. Rake cursor: improving pointing performance with concurrent input channels. In *Proc. CHI'09*, ACM (2009), 1415–1418.

5. Bock, O., and Eckmiller, R. Goal-directed arm movements in absence of visual guidance: evidence for amplitude rather than position control. *Experimental Brain Research 62*, 3 (1986), 451–458.
6. Brown, L. E., Rosenbaum, D. A., and Sainburg, R. L. Limb position drift: implications for control of posture and movement. *Journal of Neurophysiology 90*, 5 (2003), 3105–3118.
7. Butscher, S., Hornbæk, K., and Reiterer, H. Spacefold and physiclenses: simultaneous multifocus navigation on touch surfaces. In *Proc. AVI'14*, ACM (2014), 209–216.
8. Elmqvist, N., Riche, Y., Henry-Riche, N., and Fekete, J.-D. Mélange: Space folding for visual exploration. *IEEE TVCG 16*, 3 (2010), 468–483.
9. Fitzmaurice, G. W., and Buxton, W. An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In *Proc. UIST'97*, ACM (1997), 43–50.
10. Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. Bricks: laying the foundations for graspable user interfaces. In *Proc. CHI'95*, ACM (1995), 442–449.
11. Flanagan, J. R., Terao, Y., and Johansson, R. S. Gaze behavior when reaching to remembered targets. *Journal of neurophysiology 100*, 3 (2008), 1533–1543.
12. Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., and Schutte, P. C. The h-metaphor as a guideline for vehicle automation and interaction.
13. Goodwin, C. *Pointing: Where language, culture and cognition meet*. Lawrence Erlbaum Associates, 2003, ch. Pointing as situated practice, 217–241.
14. Hagiya, T., and Kato, T. Probabilistic touchscreen text entry method incorporating gaze point information. *Journal of Information Processing 23*, 5 (2015), 708–715.
15. Huang, J., White, R., and Buscher, G. User see, user point: gaze and cursor alignment in web search. In *Proc. CHI'12*, ACM (2012), 1341–1350.
16. Jacob, R. J. What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI'90*, ACM (1990), 11–18.
17. Jacucci, G., Spagnolli, A., Freeman, J., and Gamberini, L. Symbiotic interaction: a critical definition and comparison to other human-computer paradigms. In *Symbiotic Interaction'14*. Springer, 2014, 3–20.
18. Kosunen, I., Jylha, A., Ahmed, I., An, C., Chech, L., Gamberini, L., Cavazza, M., and Jacucci, G. Comparing eye and gesture pointing to drag items on large screens. In *Proc. ITS'13*, ACM (2013), 425–428.
19. Kumar, M., Paepcke, A., and Winograd, T. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proc. CHI'07*, ACM (2007), 421–430.
20. Mankoff, J., Hudson, S. E., and Abowd, G. D. Providing integrated toolkit-level support for ambiguity in recognition-based interfaces. In *Proc. CHI'00*, ACM (2000), 368–375.
21. Pfeuffer, K., Alexander, J., Chong, M. K., and Gellersen, H. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proc. UIST'14*, ACM (2014), 509–518.
22. Pfeuffer, K., Alexander, J., Chong, M. K., Zhang, Y., and Gellersen, H. Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze. In *Proc. UIST'15*, ACM (2015), 373–383.
23. Pohl, H., and Murray-Smith, R. Focused and casual interactions: Allowing users to vary their level of engagement. In *Proc. CHI'13*, ACM (2013), 2223–2232.
24. Rodden, K., Fu, X., Aula, A., and Spiro, I. Eye-mouse coordination patterns on web search results pages. In *Proc. CHI EA'08*, ACM (2008), 2997–3002.
25. Schwarz, J., Hudson, S., Mankoff, J., and Wilson, A. D. A framework for robust and flexible handling of inputs with uncertainty. In *Proc. UIST'10*, ACM (2010), 47–56.
26. Schwarz, J., Mankoff, J., and Hudson, S. E. An architecture for generating interactive feedback in probabilistic user interfaces. In *Proc. CHI'15*, ACM (2015), 2545–2554.
27. Serim, B. Querying and display of information: symbiosis in exploratory search interaction scenarios. In *Symbiotic Interaction'14*. Springer, 2014, 115–120.
28. Sibert, L. E., and Jacob, R. J. Evaluation of eye gaze interaction. In *Proc. CHI'00*, ACM (2000), 281–288.
29. Spijkers, W., and Spellerberg, S. On-line visual control of aiming movements? *Acta psychologica 90*, 1 (1995), 333–348.
30. Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *Proc. CHI'12*, ACM (2012), 2981–2990.
31. Stellmach, S., and Dachselt, R. Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc. CHI'13*, ACM (2013), 285–294.
32. Tuddenham, P., Kirk, D., and Izadi, S. Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In *Proc. CHI'10*, ACM (2010), 2223–2232.
33. Turner, J., Alexander, J., Bulling, A., Schmidt, D., and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *INTERACT'2013*. Springer, 2013, 170–186.
34. Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. In *Proc. CHI'87*, ACM (1987), 183–188.
35. Weir, D., Pohl, H., Rogers, S., Vertanen, K., and Kristensson, P. O. Uncertain text entry on mobile devices. In *Proc. CHI'14*, ACM (2014), 2307–2316.
36. Weiss, M., Wacharamanotham, C., Voelker, S., and Borchers, J. Fingerflux: near-surface haptic feedback on tabletops. In *Proc. UIST'11*, ACM (2011), 615–620.

37. Weiss, M., Wagner, J., Jansen, Y., Jennings, R., Khoshabeh, R., Hollan, J. D., and Borchers, J. Slap widgets: bridging the gap between virtual and physical controls on tabletops. In *Proc. CHI'09*, ACM (2009), 481–490.

38. Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K., Mazeev, M., and Benko, H. Ripples: utilizing per-contact visualizations to improve user interaction with touch displays. In *Proc. UIST'09*, ACM (2009), 3–12.

39. Williamson, J. *Continuous uncertain interaction*. PhD thesis, University of Glasgow, 2006.

40. Wu, J., Yang, J., and Honda, T. Fitts law holds for pointing movements under conditions of restricted visual feedback. *Human movement science 29*, 6 (2010), 882–892.

41. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI'99*, ACM (1999), 246–253.

Publication II

Barış Serim, Khalil Klouche, Giulio Jacucci. Gaze-Adaptive Above and On-Surface Interaction. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, 115-127, June 2017.

Gaze-Adaptive Above and On-Surface Interaction

Barış Serim[1,3], **Khalil Klouche**[1,4], **Giulio Jacucci**[1,2]
[1] Helsinki Institute for Information Technology HIIT, Dept. of Computer Science, University of Helsinki
[2] Helsinki Institute for Information Technology HIIT, Aalto University
[3] Department of Design, Aalto University, [4]Media Lab Helsinki, Aalto University
firstname.lastname@helsinki.fi

ABSTRACT
We explore the combination of above-surface sensing with eye tracking to facilitate concurrent interaction with multiple regions on touch screens. Conventional touch input relies on positional accuracy, thereby requiring tight visual monitoring of one's own motor action. In contrast, above-surface sensing and eye tracking provides information about how user's hands and gaze are distributed across the interface. In these situations we facilitate interaction by 1) showing the visual feedback of the hand hover near user's gaze point and 2) decrease the requisite of positional accuracy by employing gestural information. We contribute input and visual feedback techniques that combine these modalities and demonstrate their use in example applications. A controlled study showed the effectiveness of our techniques for manipulation tasks against conventional touch, while the effectiveness in acquisition tasks depended on the amount of mid-air motion, leading to our conclusion that the techniques can benefit interacting with multiple interface regions.

Author Keywords
Eye tracking; gaze interaction; above surface interaction; multi-touch

ACM Classification Keywords
H.5.2. Information interfaces and presentation: Input devices and strategies

INTRODUCTION
Large multitouch screens allow designers to create wide workspaces that provide direct and concurrent access to UI widgets: users can access commands or information without any additional interaction steps. However, single-focus human visual attention remains a bottleneck for concurrent access. Although wide or distributed workspaces come with the benefit of displaying many UI elements, they can divide users' visual attention between distant interface regions where the cost of redirecting the gaze is high. This is especially the case for precise pointing tasks, as the positional accuracy of touch input depends on users visually monitoring their own motor actions [12, 25].

DIS 2017, June 10-14, 2017, Edinburgh, United Kingdom.

ACM 978-1-4503-4922-2/17/06...$15.00
DOI: http://dx.doi.org/10.1145/3064663.3064744

Previous research has shown that additional input modalities can decrease the need for positional accuracy by sensing more than the touch position. One such modality is *above-surface sensing* of hand posture, position and gesture. Above-surface information has been used to discriminate between different commands that a touch input could be intended for, thereby expanding the functional vocabulary of touch actions [7, 14]. Above-surface sensing also holds promise for decreasing the difficulty of pointing tasks on large screens by adapting the interface in anticipation of touch [2, 37]. In parallel, *eye tracking* has very recently been employed to address the challenge of limited visual attention on touchscreens. Previous work in this domain compensates the lack of visual monitoring through flexible input handling [12, 25] and gaze-adaptive visual feedback [25].

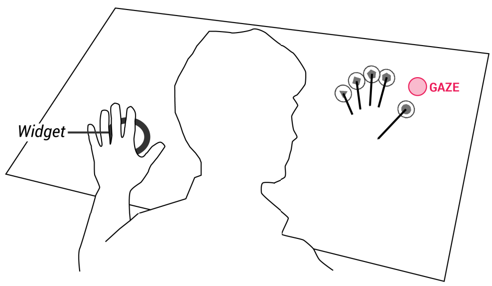


Figure 1. We use the combination of above-surface sensing and eye tracking to facilitate direct input, even when the user in not visually attending to the target. The system determines potential user actions through above-surface sensing, but defers their confirmation to touch input, accompanied by gaze-adaptive visual feedback between these two steps.

In this paper, we explore the combination of above-surface sensing with eye tracking to partly overcome the limitation of single-focus visual attention and facilitate concurrent interaction with multiple interface regions. Though each has shown individual promise, the combination of these modalities has not been studied. Our motivation for their combination is the new design possibilities they lend to supporting concurrent interaction. Together, above-surface sensing and eye tracking allow us to understand how users' hands and gaze are distributed across the interface and adapt the interface when the hands are further from user's gaze. Our interaction techniques address the aforementioned challenges of a) reliance on positional accuracy and b) limited visual attention under the two components of *input handling* and *visual feedback*.

Input handling: The interactive affordances of input widgets predispose hand posture and eye movements prior to the actual touch contact. Above-surface sensing and eye tracking allow the system to capture this pre-touch ([1, 14]) information, which we use to discriminate user actions (e.g. based on hand posture), without extensive reliance on input position.

Visual feedback: We utilize above-surface sensing and eye tracking to enable visual monitoring of multiple interface locations. This is accomplished by warping the interface contents hovered by hands to where the user's gaze is directed at, thus enabling interaction with distant interface locations without having to redirect the gaze. This allows for visual juxtaposition of a UI widget with an interface region.

Our main contributions are as follows:

- Our work contributes, to our knowledge, the first combination of above-surface sensing and eye tracking with touch. We formally describe the design considerations for the combination of these modalities to support concurrent interactions with multiple interface regions.
- We developed novel interaction mechanisms using these modalities. For input, we use hand posture to assign interactive widgets and finger-mapped touch actions for confirmation. These are complemented by visual feedback techniques that adapt the position, timing and visual aspects of the feedback using eye fixation coordinates and above-surface sensing. We demonstrate these techniques in example applications.
- Finally, we tested the efficacy of warped feedback for acquisition and manipulation tasks against a baseline condition of non-warped feedback that requires redirecting the gaze. The results show that for manipulation tasks, the performance of warped feedback was significantly better than baseline, while the results were comparable for within-widget acquisition tasks. On the other hand, the performance of warped feedback significantly suffered when the participants had to switch between widgets through midair movement.

Based on these findings we conclude that the proposed interaction techniques can be used to complement conventional visually-monitored motor actions, with warped visual feedback employed for inputs that require minimal midair movement and visually monitored motor actions employed for large midair movements by hand.

RELATED WORK

Touchscreens, which lack any tactile cues, require users to visually monitor their own action for positional accuracy. This has led to various strategies to support eyes-free input on touch surfaces, such as augmenting them with tactile widgets [34] or directing the finger to predefined locations using magnetic attraction [33]. In contrast, non-haptic solutions take advantage of the dynamic adaptation of motor and visual spaces afforded by touchscreens. In this section, we review the use of above-surface sensing and eye tracking as two modalities to decrease reliance on positional accuracy or visual monitoring during input.

Combining Gaze and Touch Input

The potential of gaze as a real-time input is being investigated in an increasing number of settings and in combination with other modalities. A common strategy is to take advantage of the rapid change of spatial context afforded by eye movements for gains in motor performance during selection. This has been the motivation for using gaze as a cursor on large and distant displays on which direct touch input is impractical [27, 29]. In these situations, eye fixations provide the input position whereas touch input on a handheld device confirms the action.

On the other hand, there has been a recent interest in combining manual input with gaze on the same surface. Common in this work, is the use of the distance between the gaze point and touch (or any other manual input type) for the flexible handling of touch. However, they differ in terms of how the situations in which touch and gaze misalign are handled. One approach is to utilize touch as an indirect and gestural input and use gaze point position for selection instead [18, 19]. This assumes visual attention on the location of input and has been motivated by the need to decrease the amplitude of motion (i.e. distance traveled by the hand), parallel to MAGIC pointing that cascades manual and gaze input [38].

Another approach to combining gaze and touch is preserving touch as a direct input but distinguishing between accurate and ambiguous touch actions by incorporating gaze point information. The main motivation in this approach has been facilitating input with decreased visual monitoring for concurrent access to multiple objects [25] or high-throughput interactions such as typing on a touchscreen while looking at the text field [12]. Both cases exemplify situations in which touch input is performed further from the location of visual feedback. They demonstrate flexible input handling mechanisms based on the level of visual monitoring with which the touch action is conducted. Decreased visual monitoring results in an expanded area for potential selection and subsequent delegation of control to the system for decision-making. The flexible input handling is accompanied by various techniques such as providing visual feedback at the periphery or translating the visual feedback coordinates to where the user's gaze is directed [25].

The aforementioned work also varies regarding how it operationalizes visual monitoring. The distance between the gaze and touch point can be utilized as a discrete distance threshold for classifying direct and indirect touch [18, 19] or as a continuous scale to determine positional uncertainty [12, 25]. However, common among them is input handling and visual feedback at the moment of touch contact. This poses a limitation because touch and gaze information at the moment of touch does not account for the complex hand-eye behavior that leads to a touch event.

Above and On Surface Interaction Continuum

Above-surface sensing promises to extend input handling and visual feedback processes to pre-touch. Earlier work used above-surface sensing in various ways, ranging from deliberate midair input actions [10, 13, 20] to implicit use of pre-touch movements by the system [14, 36]. Within previous work, we focus on a subset that use above-surface sensing in continuum with touch input [17] rather than in isolation. Research in this direction already targets relaxing the requirements of positional accuracy using various strategies.

One strategy for relaxing the requirements of positional accuracy has been adapting the motor space of the interface through target expansion [2, 37]. TouchZoom [37] uses the proximity of fingers to the screen to increase toolbar and icon target sizes before touch. Similarly, the proximity of the finger to the screen has been used to expand targets for in-vehicle interfaces that need to be operated with little visual monitoring, as visual attention is reserved for driving [2].

Above-surface sensing has also been used to extend touch functionality. Touch actions have been mapped to different interface commands based on various pre-touch or after-touch midair gestures [7]. The strategy that is most closely aligned with ours is that of Hinckley et al. [14] who utilize above-surface sensing (including grip) primarily as a pre-touch modality to provide anticipatory visual feedback and distinguish between the different commands a touch action can be intended for on a mobile device. However, instead of small and hand-held devices, we focus on interaction with large touch surfaces, which point to significantly different design considerations. We discuss these differences below.

Comparison to Previous Work

By combining gaze and above-surface sensing, we build on and advance the state of the art for these two input modalities.

When combining gaze and touch, we use touch primarily as a direct input, even in the cases where touch and gaze points misalign. This is in contrast to previous work that combines gaze with indirect manual input when touch is performed further away from gaze [18, 19]. We also further previous work that translates visual feedback near user's focus of visual attention [25]. In previous work, the visual feedback has been limited to the moment of touch contact and more importantly *does not leverage the multitouch capability of the human hand*. We use above-surface sensing to detect the proximity of not only the hand or a single finger tip, but multiple fingertips. In turn, this allows the system to visualize possible commands that can be triggered through touch with different fingers. This is achieved by showing a simplified representation of the hand that shows command-to-finger mapping near where the user is gazing.

In the domain of above-surface sensing, our main contribution is the extension of pre-touch to multifocus interactions on larger touchscreens. These interactions differ from handheld pre-touch [14] in at least three ways. First, hand grip becomes less relevant for larger screens, while the projected position of the hand over the interface (hover) gains relevance. Second, multifocus interaction on large screens causes the user's visual attention to be split between more distant interface regions, which can be detected through eye tracking for the purpose of input handling. Third, larger surfaces can accommodate the entire human hand, thereby allowing pre-touch visual feedback for all fingers.

Much effort in interaction techniques for large screens (e.g. [3, 16]) and gaze pointing [38, 18] has been directed to increasing the pointing performance for distant targets by decreasing the amplitude of motion. In contrast, we target situations in which the required motion by hand is not large but the cost of redirecting the gaze can be high. While the definitions of large touchscreen vary, the considerations of split visual attention and accommodation of both hands define the scope of our work. Indeed, previous work [12, 25] shows that input accuracy suffers when the gaze is away even for screens of 27" or smaller due to the very limited area where the vision is sharp. In the next section, we motivate our combination of these modalities and show how they enable interactions that are not possible with eye tracking or above-surface sensing alone.

DESIGN CONSIDERATIONS

Pointing to interface elements involves gradual alignment of fingertips to a visual target, a closed-loop control process that starts with midair motion and finalizes with the end of touch. Figure 2 summarizes this process and our design space by showing the progression of a command event in three modalities. It should be noted that our focus is on using the above-surface as a pre-command modality for selecting widgets and deferring the confirmation of action to touch. Thus, the figure excludes mid-air gestures that do not yield to a touch event as in gaze and gesture combinations [6] or after-touch gestures [7].

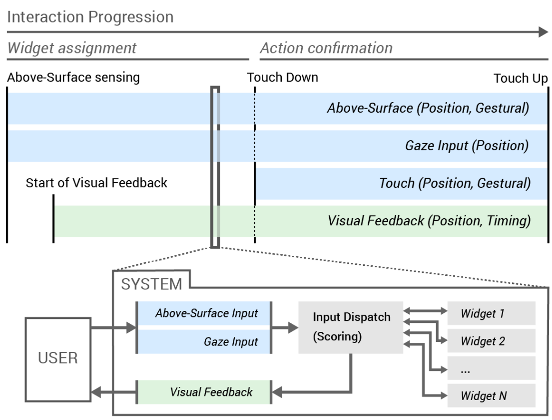


Figure 2. Continuous feedback control loop showing three input modalities during input. UI widgets are continuously ranked before a touch action, while the system provides the visual feedback of the selected widget (top). The widget assignment process (bottom) is modeled after Schwarz et al. [24].

To assign widgets to hands, we employ the uncertain input handling framework developed by Schwarz et al. [24], but adapt it to above-surface interaction. An event dispatcher continuously scores every widget-hand pair ($score_{ih}$) based on pre-touch information. Widgets can have multiple scoring criteria (j), such as position or hand posture, that are all scaled to a uniform range (0-1). Not all scoring criteria is equally relevant to a widget, which results in different scoring *weights*. As a preliminary method, we calculate the overall score of a widget by taking the weighted geometric mean of individual scores and assign the highest scoring widget to a hand (h):

$$score_{ih} = \left(\prod_{j=0}^{n} score_{jh}{}^{weight_j}\right)^{1/\Sigma_{j=1}^{n} weight_j}$$

The information of the assigned widget is communicated back to the user through visual feedback. The visual feedback differs based on the above-surface sensing and gaze conditions (Figure 3). *Minimal* refers to the visual representation in which a widget is unassigned and is minimally represented on the interface, to signify its location and current state. In *hover* condition, a hand or a finger is assigned to a widget at the same time as the user's gaze is directed to the widget. In this state the visual representation of the widget corresponds to the motor space. Finally, in the *warp* condition, a widget is assigned to a hand or finger, but the user's gaze is directed elsewhere. In this case, while the motor space remains the same, the visual feedback is warped to where the user's gaze is directed.

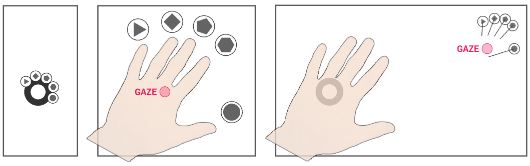


Figure 3. Three different visual states of a widget. Left: Minimal visual feedback in the absence of an assigned hand. Middle: Hover visual feedback when the widget is assigned and user gaze is directed on the widget. Right: Warped visual feedback at the gaze location when hands are above but gaze is directed elsewhere.

We made two design decisions to maintain robust interaction during continuous scoring and visual feedback. First, if a widget is already assigned to a hand or finger, it is scored slightly higher. This eliminates jittery alternations between two widgets in borderline situations. Second, when a hand or finger that is assigned to a widget touches the screen, it remains assigned to that widget until the touch is released regardless of its score.

Above and On-surface Input

Cognitive studies of motor control, distinguish between *proximal* (i.e. getting the hand near the target) and *distal* (i.e. shaping the hand in anticipation of the action) components of manual action [15]. In HCI, they correspond to the *positional* and *gestural* components of the user input. Figure 4 shows their breakdown into above- and on-surface modalities.

It is important to understand that the positional and gestural components are dissimilarly affected by low visual monitoring. Hand posture and relative finger positions are known to the user through proprioception, whereas positional accuracy requires the user to monitor where the hand or finger is located relative to the target. For input handling, the main design principle that guides the interaction techniques is replacing or complementing the positional component of input with the gestural component when possible.

Widget assignment based on hand posture

The interactive affordances of various touch actions favor certain above-surface hand postures, which we use for scoring widgets. For example, a widget that requires a pinch gesture can be scored higher for hand postures that feature the thumb and index as the only extended fingers. As we use

	Positional Component	Gestural Component
Above - surface	Projected position (x, y) Proximity to the surface (z)	Hand posture Hand / finger speed, direction
On - surface	Touch position (x,y)	Digit(thumb, index, ...) Touch gesture, touch duration

Figure 4. Breakdown of above and on-surface modalities into positional and gestural components.

above-surface sensing as a pre-touch modality, different finger configurations form the basis of various hand postures that we have defined (Figure 5). We use the data provided by a skeletal tracking software (Leap Motion) to calculate how bent each finger is and identify the best fitting hand posture (none if there is no fitting posture). The identification is done by assessing each finger for the designated range of the candidate posture.

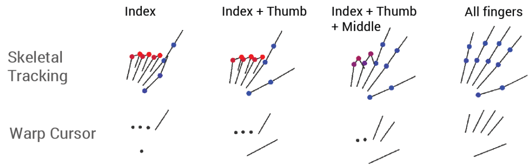


Figure 5. The system tries to match a predefined hand posture using skeletal tracking (above). The warped visual feedback shows a simplified representation of the hand (below).

An interface region can be populated with multiple widgets that favor different hand postures. In these situations, we use hand-posture information in addition to the positional component (i.e. proximity of the widget to the projected hand position) to score widgets (Figure 6). However, the same mechanism can also be applied to purely gestural interactions, in order to score widgets based on the gestural component alone.

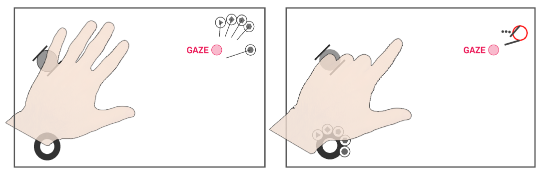


Figure 6. Hand posture can be used to discriminate between widgets in addition to the projected position of the hand or finger. Here, hand posture information is used to select between two widgets that are operated differently.

Finger-mapped touch actions

Another example of replacing positional component with the gestural component is finger-mapped touch actions. This interaction technique is accomplished in two steps. Once a widget is assigned to a hand, the system visually notifies the user by showing available command options that are mapped to individual fingers. At this stage, what determines action is the specific digit (e.g. thumb or index finger) that performs the touch action rather than the touch position (Figure 7).

This technique is related to previous efforts to design touchscreen interactions with the particular physical qualities of the human hand in mind, such as number of fingers [4, 9, 30]. For example, HandMark menus [30] provide access to menu elements upon resting the hand on the screen. The selection is then accomplished by tapping on the relevant item with the other hand. We use the above-surface modality to eliminate the need of confirmation by a second hand. Instead, available options are shown before the actual touch and are then confirmed through touch. More than a single command can be mapped to a digit. In this case, additional commands are accessed by dragging the finger to the additional commands before releasing it.

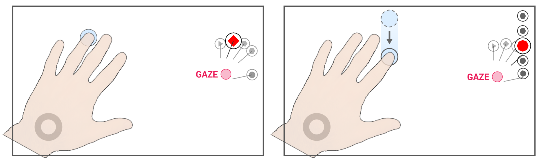


Figure 7. Various commands are associated with individual fingers when a hand is assigned to a widget. Available commands that correspond to fingers are visualized prior to the touch. Multiple commands that are mapped to the same finger can be accessed by dragging the finger (right).

Eye Tracking

We use eye tracking not only for determining the position of the visual feedback, but also for input handling. Eye tracking data is noisy and eye behavior is unstable, so we rely on fixation points as operational gaze points. The fixations are calculated with the commonly used dispersion-threshold identification algorithm, I-DT (implementation in [21]), with the minimum time window set to 90 milliseconds and the dispersion threshold set to 1°.

Determining the weight of the position score

We use gaze point information for scoring the widgets. When scoring widgets, positional and gestural components can *reinforce* each other, for example, when a thumb and index finger hand posture is in the proximity of a virtual knob. However, the two components can also *contradict* each other if the same hand posture is in the proximity of an interactive widget that is operated by tapping.

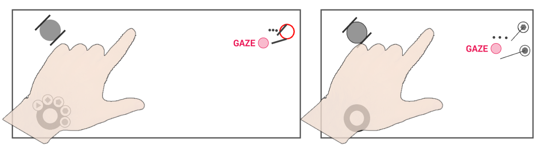


Figure 8. The weight of positional component is manipulated based on the degree the action is visually monitored. The relative weight of the gestural component (hand posture) increases when the user gaze is directed elsewhere.

In these cases, we employ eye tracking to determine the weight of the positional component for calculating the score (Figure 8). The weight is continuously re-evaluated based on the distance between the eye fixation and the projected position of the hand on the screen. The weight of the position is higher when the user's gaze is located near the widget and lower when it is further away.

Gaze as pointer interactions

The design space can also be extended to situations, in which gaze is used as a pointer for selection in combination with touch input (*gaze + touch interactions*). For example, a manipulation action on an object can be performed at a single step by keeping the gaze on the object and determining the action by touching on a widget. In other words, rather than following a sequential order for selection and manipulation, gaze and manual input are used concurrently. What distinguishes this technique from earlier work that combines indirect touch and gaze [18, 19] is the use of touch as a direct input and concurrent pointing by touch and gaze.

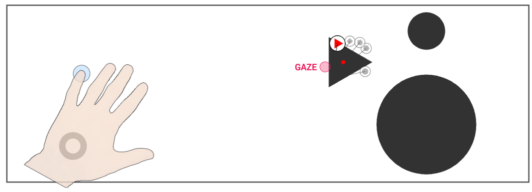


Figure 9. Gaze coordinates can be coupled with manual input to perform *gaze + touch* interactions. In the figure the hand is over a widget and the finger-mapped direct touch is used to modify the shape of an object.

Visual Feedback

Unlike hovering with a cursor, touch input lacks a pre-command notification stage. Above-surface sensing has been proposed as a modality for providing informative visual feedback prior to the actual touch [8, 17]. Input with low visual monitoring can benefit from the timely communication of the system's interpretation of the user input (assigned widget) back to the user. Above-surface interaction and eye tracking call for a reconsideration of design choices regarding the *timing* and *positioning* of the visual feedback.

Timing

While above-surface sensing enables pre-touch visual feedback, feedback that occurs earlier than expected can be intrusive. Similarly, appropriate withdrawal of the visual feedback is necessary when it is not needed. To prevent earlier-than-expected feedback, we use proximity to the surface (z) and hand velocity variables. Visual feedback is shown only when the hand is in the proximity of the screen (z) and the projected hand velocity (on x-y plane) is low.

We additionally rely on widget interaction events to correctly time the visibility of the feedback. As a principle, the system changes the visibility for continuous events gradually and for discrete events instantly (Figure 10). Assignment of a widget to a hand, or a touch-down action on a widget maximizes the visibility of the widget. The feedback starts fading out in case of inaction. Conversely, above-threshold hand movement causes the warped feedback to gradually gain visibility. The warped feedback is fully visible as long as the widget is being touched by a finger, whereas releasing all the fingers from a widget causes a sudden drop in visibility, followed by a gradual fade out. The visibility can be mapped to different visual channels such as opacity or size.

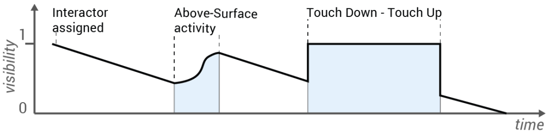


Figure 10. The relative visibility of a widget in relation to the interaction events. The visibility gradually decreases during episodes of inaction.

Position

We use eye fixation coordinates to determine the position of the warped visual feedback. However, eye fixations shift rapidly, and constant warping of the visual feedback can be intrusive. Thus, we update the position of the warped visual feedback upon interaction events (such as when a widget is assigned), upon touch actions and above-threshold mid-air movement. Also, we update the position of the feedback only when the distance between the current position of the visual feedback and the most recent fixation is above a certain threshold ($\approx 10^{\circ}$ of visual angle from eyes).

Summary

We described how gestural component can partially replace position during widget assignment and action confirmation. We also described how eye tracking can be used both for input handling and visual feedback. The next section describes example applications that demonstrate various combinations of the aforementioned techniques.

APPLICATION EXAMPLES

Apparatus

The applications were prototyped for a 10 finger multitouch screen (27", 2,560 × 1,440 pixels) that was tilted 30° to enable easier hand reach for touch input. For above-surface sensing, a commercial short range infrared sensor for skeletal hand tracking (Leap motion) was used. We mapped the coordinate system of above-surface depth sensing to screen coordinates using a 4 point calibration procedure. The eye tracking was performed by Pupil Labs binocular tracking glasses running at 60Hz. The applications consistently ran at 60fps, with the latest gaze data point synchronized at every frame.

Object Drawing and Manipulation

We prototyped an application that enables parametric manipulation of a group of objects or adding new objects on a canvas. Objects have multiple parameters, namely shape, fill color, border color, size and orientation. The application demonstrates how these parameters can be modified without redirecting the gaze between the toolbar and canvas areas through above-surface sensing and warped visual feedback. Besides allowing continuous fixation on the canvas, warped feedback allows for the juxtaposition of a widget with a specific region on the canvas. For example, warped feedback of the color selection tool on the canvas can facilitate easier visual comparison of the selected color with the already existing colors on the canvas. The application features the following widgets:

1) The shape selector allows switching between different geometric forms. Basic shapes such as triangle, square, pentagon and circle are selected through tap actions with different fingers. Additionally a larger set of polygons can be accessed by dragging the index finger on the surface. Given that all the fingers can be used for interaction, the widget is scored higher when all the fingers of the hand are extended.

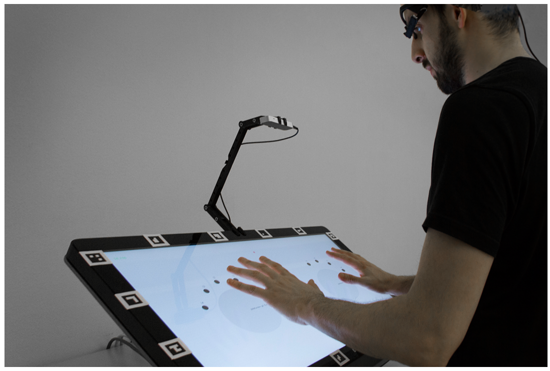


Figure 11. The hardware setup consisting of a touch screen, depth sensor (Leap Motion, attached to the upper screen edge) and eye tracking glasses.

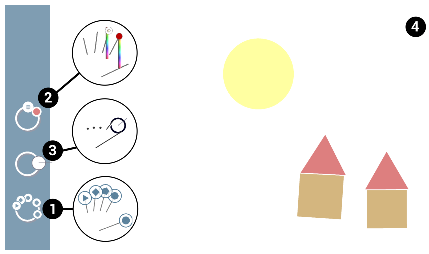


Figure 12. Drawing application featuring the shape selector (1), the color selector (2), the physical modifier (3) and the canvas (4). Warped feedback visuals of the widgets 1,2 & 3 are annotated.

2) The color selector selects or modifies object colors by dragging a finger on a two-dimensional color space. Index and middle fingers are mapped to fill and stroke colors, respectively. The widget is scored higher for hand postures that feature all or the first three (thumb, index, middle) fingers extended and when the hand is in the proximity.

3) The physical modifier manipulates the size and orientation of a group of objects through pinch and two-finger (thumb + index) rotation. Besides positional proximity, the widget is scored higher for hand postures that only feature extended thumb and index fingers.

4) The canvas allows users to add new objects to the scene. When the user's hand is hovering over the canvas, the system displays the shape that will be drawn upon touch. The canvas is scored based on the projected position and orientation of the hand. The widget is scored higher if the thumb is located to the left of the palm center. This allows for discrimination between other widgets on the toolbar in borderline conditions.

Real-time Video Manipulation

We prototyped a simplified VJ (video jockeying) tool that enables the user to apply a variety of visual filters to a video loop. Real-time video editing offers a relevant challenge, as the main feedback for most inputs results in real-time visual changes to the video picture away from the input location. In that case, being able to display essential interaction information at the gaze location allows users to keep their attention on the target when it is most critical. The different widgets either appear as soon as the user's hand is detected, or inconspicuously suggested on the sides of the display as perceived affordances. The application features the following widgets:

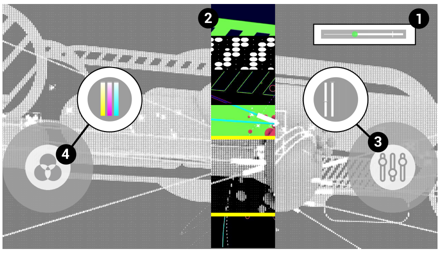


Figure 13. Vj application featuring loop control (1), the filter selector (2), the filter modifier (3) and the color modifier (4). Warped feedback visuals of the widgets 1,3 & 4 are annotated.

1) Loop control is a global widget that can be accessed from anywhere on the screen, as its score is solely based on hand posture. A pinch posture assigns the widget to the hand. Visual feedback of the playback information is shown near the gaze fixation. The loop ranges are then selected by dragging the index finger or the thumb.

2) The filter selector is scored based on the position information and appears when the user extends her hand towards the middle of the screen. The video picture then gets divided into multiple horizontal areas, each displaying a preview of available visual filters. The selection is then accomplished by touch.

3) The filter modifier and *the color modifier (4)* are located on the right and left sides of the screen and are scored based on position. The parameters are mapped to one specific finger each – i.e. index for parameter 1, middle for parameter 2, ring for parameter 3, and pinkie when a fourth parameter is needed. Approaching the widget with the hand will show gaze-adaptive feedback for the parameters, as vertical gauges, with cursors indicating the current value between the maximum and minimum allowed for each gauge. The locations of color and filter modifiers allow concurrent bimanual modification of both widgets.

STUDY: WARPED FEEDBACK AND DIVIDED ATTENTION

We set out to test the viability of warped visual feedback for situations in which attention needs to be divided between multiple interface regions. Previous work identifies two interaction stages, namely acquiring and manipulating a control device [11, 28]. Thus, we prepared a two-part experiment that involves acquiring a target on a touch surface and manipulating a widget through on-surface interaction. We were interested in seeing how the task completion times in these two stages vary for the conditions of:

1. Warped visual feedback condition, which facilitates continuous gaze fixation near the stimulus position using a smaller representation of hand (scaled down by a factor of 0.35 to be visible and less intrusive).
2. Baseline condition, which provides no specific support to facilitate continuous fixation.

Apparatus

The same hardware setup described earlier has been used, with a 27", tilted touchscreen (set to 1920×1080) for input. Additionally, The experimental setup involved a second, vertically positioned monitor of the same size placed on the upper edge of the first monitor (Figure 14). The combination of a horizontal input surface with a vertical monitor have been investigated before [5, 31, 32], and has the advantage of showing visual feedback at the eye level while providing arm support on the horizontal surface. In the experiments, the two monitor setup was used to display the target stimuli either in one of the following locations:

1. Near the left edge of the same tilted screen on which the input is performed (same screen condition). This resulted in a visual angle of around 50° degrees between the input and stimuli positions depending on the distance of eyes to the screen.
2. Near the upper-left corner of the vertical screen. This resulted in a visual angle of around 70° degrees between the input and stimuli positions.

The main motivation for using two screens was to compare the effect of different distances between target stimuli and input area expressed in visual angles. Different distances between the target stimuli and input area corresponds to different costs for redirecting the gaze, which can be be generalized to other display setups.

The experimental setup also controlled the warped feedback location by always displaying the warped feedback near the stimulus. This was implemented to prevent users from shifting their gaze to the input area (for warped feedback condition). However, eye tracking data was collected from participants for later analysis.

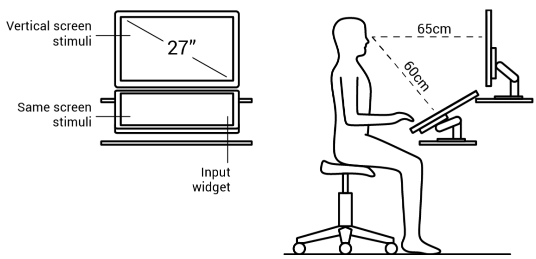


Figure 14. The front and right views of the experimental setup. The screens were positioned approximately 60 and 65 cms away from the eye.

Participants
Twelve participants (2 female), aged 20 to 33 (m=26.1, sd=3.4) were recruited for the study. Of the participants, eleven were right handed and all reported extensive familiarity with touch devices. The participants have been compensated with one cinema ticket and their informed consent has been collected for data logging.

Procedure
We performed a within-subject study of two interaction conditions (warped feedback and baseline) and two stimulus location (same or vertical screen) conditions. The order of the interaction and stimulus conditions alternated between participants using a Latin square design to counterbalance the potential effects of learning and fatigue. The participants were seated in front of the setup, with the input and vertical screens positioned approximately 60 and 65 cms away from the eye. The participants performed the acquisition and manipulation experiments in sequence. The participants had a chance to rest between experiments. The sessions lasted 40 to 50 minutes.

Analysis
The initial trials of each condition (1 block for acquisition tasks, 2 blocks for manipulation) were excluded from the analysis. The metric for evaluation is the task completion time, which we report as grand mean and the grand median, the latter being more robust to outliers. The baseline and warp conditions have been compared using Welch's t-test and normalized time completion values based on subtracting the means of each participant from the data divided by standard deviation. Error percentages represent the ratio of the erroneous touch releases to all touch releases.

Experiment 1: Widget Acquisition
With this task, we aimed to compare the combination of warped feedback and finger-mapped input versus regular direct touch actions. The experimental interface (Figure 15) consisted of 3 vertically arranged widgets (each with 4 targets "1,2,3,4", "A,B,C,D" and "+,-,*,/" from left to right).

Each widget covered a 400*120 pixel area of the screen. In the baseline condition this area was vertically divided into 4 pixel buttons that are each 100*120 pixels and selected through positional input. In contrast, for the warped feedback condition, the participants acquired a widget when their middle finger hovered within the 400*120 pixel area and selected any of the 4 targets respectively through thumb, index, middle and ring fingers rather than touch position. Thus, the spatial footprints of the widgets were the same for both input conditions, but the interaction differed regarding the selection of 4 targets within the widget (positional or finger-mapped) and the position of feedback (no warped feedback and warped feedback).

Each experimental block involved alternating between three widgets in randomized order and acquiring targets in the same widget in randomized order, which resulted in 12 trials (3 widgets $\times$ 4 targets). With this procedure, we aimed to observe the performance of acquisition task between-widgets (the first acquisition task after the widget is alternated) and for acquisition within-widgets.

We anticipated time savings by eliminating attention switches under warp condition. However, we also anticipated a decrease in motor performance in warped feedback condition due to lack of visual monitoring. Previous work reports a decrease in accuracy when the hand movement is not visually guided [26, 35]. More specifically, Schmidt et al. report decrease in performance for large amplitude bimanual tasks when visual feedback is separated from the motor space [23]. However, the cumulative effect of redirecting the gaze and motor performance has not been studied. Even we target relaxing the need for positional accuracy, we were interested in seeing the viability of positional input through warped visual feedback, which corresponds to between-widget acquisition tasks.

Each trial was completed upon a touch release on the correct target. The combination of 2 input methods $\times$ 2 stimuli positions $\times$ 10 blocks (excluding the training block) $\times$ 12 targets resulted in 480 trials per participant. Before the experiment the participants practiced the finger-mapped input condition with 60 trials.

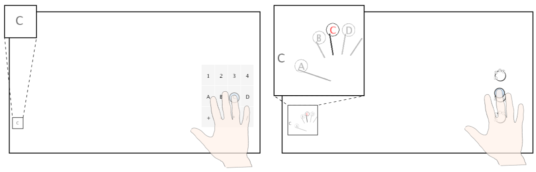


Figure 15. In the baseline condition (left) the target was acquired through touch position, while in the warped feedback condition (right) it was acquired through finger mapped input after the widget is assigned. Details are magnified. The stimuli character height was 9.3mm.

Results
Early in the analysis, we noticed time completion differences for between within-widget and between-widget tasks and analyzed them separately. For within-widget tasks, the mean completion times for warped feedback and baseline conditions were similar in both stimuli conditions (Table 1). A t-test comparison using within-subject normalized completion times did not show any significant effect for the same screen ($r = .03$, $p = .23$) and vertical screen ($r = .03$, $p = .17$) conditions. The error rates were higher for the warped feedback condition (Table 1). In few instances participants reported system failures for identifying the wrong finger, which might have affected the error rate for warped feedback condition.

However, the results were significantly different for between-widget acquisition tasks (Table 1). On average, participants spent significantly more time on the warp condition than on the baseline condition for both the same screen ($r = .42$, $p < .001$) and vertical screen conditions ($r = .54$, $p < .001$). This was accompanied by a much greater error rate for warped feedback condition (Table 1). Participant interviews suggested a few possible explanations for the difference in performance between the two input conditions. One is the mismatch between the distance traveled by the hand and its scaled down visual representation in warped visual feedback. During between-interaction tasks, we informally observed that in the baseline condition participants performed high-speed ballistic movements towards the touch target, while in the baseline condition,

Acquisition task *(within-widget)*				
Screen	Technique	Median(ms)	Mean(ms)	Error
Same	Warped	**902.75**	**1031.54**	6.41%
	Baseline	961.75	1068.36	**2.43%**
Vert.	Warped	**885.50**	1055.00	8.16%
	Baseline	986.25	**1021.88**	**3.31%**

Acquisition task *(between-widget)*				
Screen	Technique	Median(ms)	Mean(ms)	Error
Same	Warped	2094.00	2305.19	11.33%
	Baseline	**1164.50**	**1406.79**	**1.09%**
Vert.	Warped	2258.75	2342.36	9.09%
	Baseline	**1207.5**	**1415.19**	**2.17%**

Table 1. The grand median and grand mean completion times and overall error rates for two interaction and two stimuli conditions for the acquisition tasks. Emphasis (in bold) represents better performance.

they moved their hand parallel to the screen, keeping a tense hand posture. Participants also reported shoulder fatigue for warped feedback condition, which may have been caused by parallel arm movements.

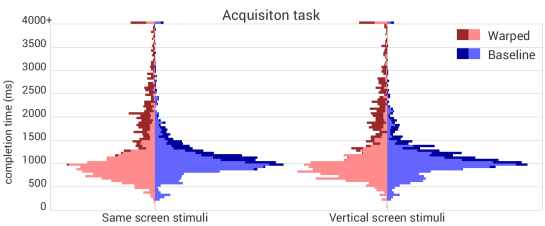


Figure 16. Distribution of acquisition task completion times (50ms bins) across all participants for two interaction and two stimuli conditions. The shades and tints respectively indicate between-widget and within-widget acquisitions tasks.

Experiment 2: Widget Manipulation

The manipulation task required participants to match the value of an interactive slider to that of the stimulus. To isolate manipulation from acquisition, the interactive slider was always assigned to the participant's hand. Each trial required the participants to manipulate the slider by dragging the index finger on the touchscreen. If the touch was released when the value of the slider matches with that of the stimulus, the task was counted as complete. The slider consisted of 5 steps (the targets being the numerals "1,2,3,4,5"), and 15 pixels of movement resulted in 1 step. Unlike the acquisition task, the only difference between the baseline and warped feedback conditions was the location of the visual feedback. In the warped feedback condition the feedback was displayed near the stimulus, whereas in the baseline condition it was displayed near the widget. We expected that eliminating visual attention switches in gaze-adaptive condition would result in shorter task completion times. The combination of 2 input methods × 2 stimuli positions × 20 blocks (excluding training blocks) × 5 targets (shown in random order) resulted in 400 trials per participant.

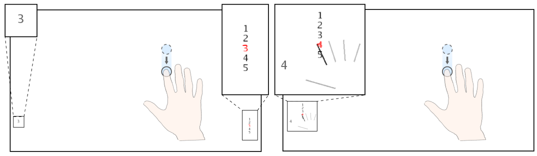


Figure 17. In the baseline condition (left) the visual feedback of the manipulated slider was shown at the widget location, while in the warped feedback condition (right) it was shown near the target. Details are magnified. The stimuli character height was 9.3mm.

Results

For both stimuli conditions participants spent more time on the baseline condition than on the warped feedback condition (Table 2). A t-test comparison of the same and vertical screen conditions using normalized data yielded a larger effect size for the vertical screen condition ($r = .19$, $p < .001$) than for the same screen condition ($r = .10$, $p < .001$), in line with the expectation that the higher cost of redirecting the gaze in vertical screen condition will result in more pronounced benefits when using warped feedback. The error rates were lower for the warped feedback condition in both screen conditions (Table 2).

Manipulation Task				
Screen	Technique	Median(ms)	Mean(ms)	Error
Same	Warped	**1387.25**	**1576.30**	**4.07%**
	Baseline	1530.25	1746.89	6.46%
Vert.	Warped	**1364.00**	**1509.06**	**3.69%**
	Baseline	1616.50	1770.58	6.54%

Table 2. The grand median and grand mean completion times and overall error rates for two interaction and two stimuli conditions for the manipulation task. Emphasis (in bold) represents better performance.

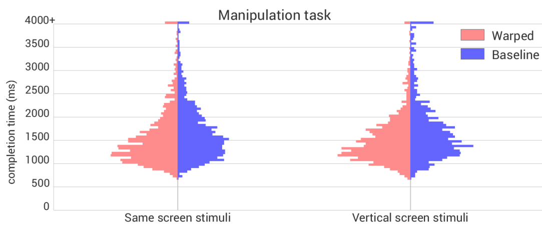


Figure 18. Distribution of manipulation task completion times (50ms bins) across all participants for two interaction and two stimuli conditions.

DISCUSSION

We combined above and on-surface modalities with eye tracking to decrease the need for positional accuracy for input on touchscreens. This is enabled by discriminating between potential user commands through the gestural component of input and visual feedback the near gaze point location. The controlled study evaluated a subset of the interaction techniques but gave us valuable insights about the viable scope of warped visual feedback. Here, we would like to discuss the strengths and weaknesses of the interaction techniques based on the outcome of the user study.

Midair Motion Amplitude

The performances of the warped feedback and baseline conditions were visibly different based on whether the task was manipulation, within-widget acquisition or between-widget acquisition. The performance of the warped visual feedback condition was higher for manipulation tasks that required no midair motion. The performance between warped feedback and baseline conditions were comparable for within-widget acquisition. However, the performance of warped visual feedback was significantly worse for between-widget acquisition tasks, in which participants had to acquire the widget through midair motion. Based on these results, we arrive at the following conclusions:

- The warped feedback was successful in decreasing the cost of redirecting the gaze, resulting in the improvement of task completion time for manipulation tasks.
- However, the warped feedback did not facilitate midair hand motion as effectively as direct visual monitoring, which resulted in a decrease in performance for between-widget acquisition tasks. Our results confirm earlier work by Schmidt et al. who reports lower performance and similar observations such as tense hand posture and parallel-to-screen hand movements when touch is performed without direct visual monitoring [23].

These conclusions also suggest that the interaction techniques described are strong in cases where the required midair motion is minimal and the task requires visual attention at a distant interface region. On the other hand, the conclusions suggest a weakness for inputs that require large amplitudes of midair motion. These have the following implications. First, in its current state, we see the role of warped feedback as complementary to conventional touch input. In these cases, the interface should allow for effortless transitions between warped feedback and conventional input modes by, for example, allowing acquisition of the widget through visual monitoring, but then transitioning to warped feedback for manipulation or within-widget acquisition. Second, for the interaction techniques presented, we see the main advantage of large touch surfaces in accommodating both of the user's hands to allow finger-mapped and bimanual concurrent input instead of expansive widget configurations.

Warped Feedback and Visual Search

Our primary reason for implementing warped feedback was decreasing the redirection of gaze and juxtaposing two different regions. Warping a color palette to a canvas area is a good example of this, as it enables comparison of the selected color with a region on the canvas. However, there can be cases in which it is more beneficial to see multiple input widgets rather than warping the interface widget. Visual search of a command in the toolbar area is a good example of such a use case in which the user needs to see widgets side by side. To support these use cases, the minimal representation of the widgets should facilitate visual search and interface learning.

Spatial and Motor Memory

Replacing positional accuracy with the gestural component (e.g. hand posture, or finger-mapped touch actions) also suggests an increased reliance on motor rather than spatial memory during interface learning. Although spatial and motor memories can be related (e.g. finger-mapped touch actions also correspond to different positions) the extent to which they are related is an open question in the research community [22]. Thus, the effects of spatial and motor memory for proficient use should be further investigated. Another important consideration is the number of interface position, hand posture and finger combinations that are both discernible and memorable.

Gesture Input

The interaction techniques we presented rely on both positional (proximity of hand to the widget) and gestural information (hand posture and digit). However, we see potential application cases for pure gestural interactions. Above-surface sensing can be used to anticipate the action in advance and provide visual confirmation at the gaze location to inform the user of the possible action a touch event can lead to, thereby addressing the problem of learning and memorization in gesture-based interfaces.

CONCLUSIONS

The limited nature of our visual attention stands in contrast to the concurrent action possibilities afforded by bimanual action. These concurrent action and sensing possibilities have been our motivation for supporting interactions that are distributed around multiple interface regions. We presented the design considerations for combining eye tracking with above- and on-surface modalities, in order to decrease reliance on positional accuracy during interaction. The system determined potential user actions midair, but deferred their confirmation to on-surface touch, accompanied by gaze-adaptive visual feedback between these two steps. The preliminary results from the controlled study suggest that warped visual feedback can complement conventional direct touch for manipulation tasks and acquisition tasks that require minimal midair motion, although future work can target improvements for increasing the viable scope. Besides immediate design implications, we identify these three modalities as a promising area for investigation, as their combination provides a fine-grained understanding of hand-eye coordination.

ACKNOWLEDGMENTS

This work has been partly supported by MindSee (FP7 - ICT; Grant Agreement #611570). We would like to thank Oswald Barral for his help with the data analysis, the anonymous reviewers for their comments and to the participants of the study.

REFERENCES

1. Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2011. Medusa: A Proximity-aware Multi-touch Tabletop. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 337–346. DOI: http://dx.doi.org/10.1145/2047196.2047240

2. Ilhan Aslan, Alina Krischkowsky, Alexander Meschtscherjakov, Martin Wuchse, and Manfred Tscheligi. 2015. A Leap for Touch: Proximity Sensitive Touch Targets in Cars. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 39–46. DOI: http://dx.doi.org/10.1145/2799250.2799273
3. Patrick Baudisch, Edward Cutrell, Dan Robbins, Mary Czerwinski, Peter Tandler, Benjamin Bederson, Alex Zierlinger, and others. 2003. Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch-and pen-operated systems. In *Proceedings of INTERACT*, Vol. 3. IOS Press, 57–64.
4. Hrvoje Benko, T. Scott Saponas, Dan Morris, and Desney Tan. 2009. Enhancing Input on and Above the Interactive Surface with Muscle Sensing. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '09)*. ACM, New York, NY, USA, 93–100. DOI:http://dx.doi.org/10.1145/1731903.1731924
5. Xiaojun Bi, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2011. Magic Desk: Bringing Multi-touch Surfaces into Desktop Work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2511–2520. DOI: http://dx.doi.org/10.1145/1978942.1979309
6. Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, USA, 131–138. DOI: http://dx.doi.org/10.1145/2818346.2820752
7. Xiang Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott E. Hudson. 2014. Air+Touch: Interweaving Touch & In-air Gestures. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 519–525. DOI: http://dx.doi.org/10.1145/2642918.2647392
8. Victor Cheung, Jens Heydekorn, Stacey Scott, and Raimund Dachselt. 2012. Revisiting Hovering: Interaction Guides for Interactive Surfaces. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 355–358. DOI: http://dx.doi.org/10.1145/2396636.2396699
9. Daewoong Choi, Hyeonjoong Cho, and Joono Cheong. 2015. Improving Virtual Keyboards When All Finger Positions Are Known. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 529–538. DOI: http://dx.doi.org/10.1145/2807442.2807491
10. Bruno De Araujo, Géry Casiez, and Joaquim Jorge. 2012. Mockup Builder: Direct 3D Modeling On and Above the Surface in a Continuous Interaction Space. In *GI'12 - 38th Graphics Interface conference*. Canadian Information Processing Society, Toronto, Canada, 173–180. https://hal.inria.fr/hal-00687891
11. George W. Fitzmaurice and William Buxton. 1997. An Empirical Evaluation of Graspable User Interfaces: Towards Specialized, Space-multiplexed Input. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM, New York, NY, USA, 43–50. DOI: http://dx.doi.org/10.1145/258549.258578
12. Toshiyuki Hagiya and Tsuneo Kato. 2015. Probabilistic Touchscreen Text Entry Method Incorporating Gaze Point Information. *Journal of Information Processing* 23, 5 (2015), 708–715. DOI: http://dx.doi.org/10.2197/ipsjjip.23.708
13. Otmar Hilliges, Shahram Izadi, Andrew D. Wilson, Steve Hodges, Armando Garcia-Mendoza, and Andreas Butz. 2009. Interactions in the Air: Adding Further Depth to Interactive Tabletops. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology (UIST '09)*. ACM, New York, NY, USA, 139–148. DOI: http://dx.doi.org/10.1145/1622176.1622203
14. Ken Hinckley, Seongkook Heo, Michel Pahud, Christian Holz, Hrvoje Benko, Abigail Sellen, Richard Banks, Kenton O'Hara, Gavin Smyth, and William Buxton. 2016. Pre-Touch Sensing for Mobile Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2869–2881. DOI: http://dx.doi.org/10.1145/2858036.2858095
15. M. Jeannerod. 1984. The Timing of Natural Prehension Movements. *Journal of Motor Behavior* 16, 3 (1984), 235–254. DOI: http://dx.doi.org/10.1080/00222895.1984.10735319
16. Azam Khan, George Fitzmaurice, Don Almeida, Nicolas Burtnyk, and Gordon Kurtenbach. 2004. A Remote Control Interface for Large Displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 127–136. DOI: http://dx.doi.org/10.1145/1029632.1029655
17. Nicolai Marquardt, Ricardo Jota, Saul Greenberg, and Joaquim A. Jorge. 2011. *The Continuous Interaction Space: Interaction Techniques Unifying Touch and Gesture on and above a Digital Surface*. Springer Berlin Heidelberg, Berlin, Heidelberg, 461–476. DOI: http://dx.doi.org/10.1007/978-3-642-23765-2_32
18. Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 509–518. DOI: http://dx.doi.org/10.1145/2642918.2647397

19. Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 373–383. DOI: http://dx.doi.org/10.1145/2807442.2807460

20. Henning Pohl and Roderick Murray-Smith. 2013. Focused and Casual Interactions: Allowing Users to Vary Their Level of Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2223–2232. DOI: http://dx.doi.org/10.1145/2470654.2481307

21. Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 71–78. DOI: http://dx.doi.org/10.1145/355017.355028

22. Joey Scarr, Andy Cockburn, and Carl Gutwin. 2013. Supporting and Exploiting Spatial Memory in User Interfaces. *Found. Trends Hum.-Comput. Interact.* 6, 1 (Dec. 2013), 1–84. DOI: http://dx.doi.org/10.1561/1100000046

23. Dominik Schmidt, Florian Block, and Hans Gellersen. 2009. A Comparison of Direct and Indirect Multi-touch Input for Large Surfaces. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I (INTERACT '09)*. Springer-Verlag, Berlin, Heidelberg, 582–594. DOI: http://dx.doi.org/10.1007/978-3-642-03655-2_65

24. Julia Schwarz, Scott Hudson, Jennifer Mankoff, and Andrew D. Wilson. 2010. A Framework for Robust and Flexible Handling of Inputs with Uncertainty. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 47–56. DOI: http://dx.doi.org/10.1145/1866029.1866039

25. Baris Serim and Giulio Jacucci. 2016. Pointing While Looking Elsewhere: Designing for Varying Degrees of Visual Guidance During Manual Input. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5789–5800. DOI: http://dx.doi.org/10.1145/2858036.2858480

26. Will Spijkers and Simone Spellerberg. 1995. On-line Visual Control of Aiming Movements? *Acta Psychologica* 90, 1 - 3 (1995), 333 – 348. DOI: http://dx.doi.org/10.1016/S0001-6918(95)90000-4 Discrete and Continuous Information Processing.

27. Sophie Stellmach and Raimund Dachselt. 2013. Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 285–294. DOI: http://dx.doi.org/10.1145/2470654.2470695

28. Philip Tuddenham, David Kirk, and Shahram Izadi. 2010. Graspables Revisited: Multi-touch vs. Tangible Input for Tabletop Displays in Acquisition and Manipulation Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 2223–2232. DOI: http://dx.doi.org/10.1145/1753326.1753662

29. Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. 2013. *Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch*. Springer Berlin Heidelberg, Berlin, Heidelberg, 170–186. DOI: http://dx.doi.org/10.1007/978-3-642-40480-1_11

30. Md. Sami Uddin, Carl Gutwin, and Benjamin Lafreniere. 2016. HandMark Menus: Rapid Command Selection and Large Command Sets on Multi-Touch Displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5836–5848. DOI: http://dx.doi.org/10.1145/2858036.2858211

31. Simon Voelker, Andrii Matviienko, Johannes Schöning, and Jan Borchers. 2015. Combining Direct and Indirect Touch Input for Interactive Workspaces Using Gaze Input. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 79–88. DOI:http://dx.doi.org/10.1145/2788940.2788949

32. Malte Weiss, Simon Voelker, Christine Sutter, and Jan Borchers. 2010. BendDesk: Dragging Across the Curve. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/1936652.1936654

33. Malte Weiss, Chat Wacharamanotham, Simon Voelker, and Jan Borchers. 2011. FingerFlux: Near-surface Haptic Feedback on Tabletops. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 615–620. DOI: http://dx.doi.org/10.1145/2047196.2047277

34. Malte Weiss, Julie Wagner, Yvonne Jansen, Roger Jennings, Ramsin Khoshabeh, James D. Hollan, and Jan Borchers. 2009. SLAP Widgets: Bridging the Gap Between Virtual and Physical Controls on Tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 481–490. DOI: http://dx.doi.org/10.1145/1518701.1518779

35. Jinglong Wu, Jiajia Yang, and Taichi Honda. 2010. Fitts' law holds for pointing movements under conditions of restricted visual feedback. *Human Movement Science* 29, 6 (2010), 882 – 892. DOI: http://dx.doi.org/10.1016/j.humov.2010.03.009

36. Haijun Xia, Ricardo Jota, Benjamin McCanny, Zhe Yu, Clifton Forlines, Karan Singh, and Daniel Wigdor. 2014. Zero-latency Tapping: Using Hover Information to Predict Touch Locations and Eliminate Touchdown Latency. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 205–214. DOI: http://dx.doi.org/10.1145/2642918.2647348

37. Xing-Dong Yang, Tovi Grossman, Pourang Irani, and George Fitzmaurice. 2011. TouchCuts and TouchZoom: Enhanced Target Selection for Touch Displays Using Finger Proximity Sensing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2585–2594. DOI: http://dx.doi.org/10.1145/1978942.1979319

38. Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 246–253. DOI: http://dx.doi.org/10.1145/302979.303053

Publication III

Barış Serim, Ken Pfeuffer, Hans Gellersen, Giulio Jacucci. Visual Attention-Based Access: Granting Access Based on Users' Joint Attention on Shared Workspaces. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2, 3, Article No. 133, September 2018.

Visual Attention-Based Access: Granting Access Based on Users' Joint Attention on Shared Workspaces

BARIŞ SERİM, Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland and Department of Design, Aalto University, Finland
KEN PFEUFFER, Bundeswehr University Munich, Germany
HANS GELLERSEN, Lancaster University, United Kingdom
GIULIO JACUCCI, Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland

During collaboration, individual users' capacity to maintain awareness, avoid duplicate work and prevent conflicts depends on the extent to which they are able to monitor the workspace. Existing access control models disregard this contextual information by managing access strictly based on who performs the action. As an alternative approach, we propose managing access by taking the visual attention of collaborators into account. For example, actions that require consensus can be limited to collaborators' joint attention, editing another user's personal document can require her visual supervision and private information can become unavailable when another user is looking. We prototyped visual attention-based access for 3 collaboration scenarios on a large vertical display using head orientation input as a proxy for attention. The prototype was deployed for an exploratory user study, where participants in pairs were tasked to assign visual attention-based access to various actions. The results reveal distinct motivations for their use such as preventing accidents, maintaining individual control and facilitating group awareness. Visual attention-based access has been perceived as more convenient but also less certain when compared to traditional access control. We conclude that visual attention-based access can be a useful addition to groupware to flexibly facilitate awareness and prevent conflicts.

CCS Concepts: • **Human-centered computing → Collaborative interaction**; **Synchronous editors**; **Collaborative and social computing devices**; **Interactive whiteboards**;

Additional Key Words and Phrases: Visual attention, joint attention, access control, awareness, CSCW, single display groupware, design, qualitative study

ACM Reference Format:
Barış Serim, Ken Pfeuffer, Hans Gellersen, and Giulio Jacucci. 2018. Visual Attention-Based Access: Granting Access Based on Users' Joint Attention on Shared Workspaces. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 133 (September 2018), 23 pages. https://doi.org/10.1145/3264943

Authors' addresses: Barış Serim, Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland, Department of Design, Aalto University, Finland, baris.serim@helsinki.fi; Ken Pfeuffer, Bundeswehr University Munich, Germany, ken.pfeuffer@unibw.de; Hans Gellersen, Lancaster University, United Kingdom, h.gellersen@lancaster.ac.uk; Giulio Jacucci, Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland, giulio.jacucci@helsinki.fi.

2474-9567/2018/9-ART133 $15.00
https://doi.org/10.1145/3264943

1 INTRODUCTION

Concurrent input of multiple users on shared workspaces comes with the benefit of task parallelization, but also introduces the challenge of maintaining coordination between multiple users, in particular, ensuring that the work done by individual users is relevant to the joint activity and that individual users' actions do not interrupt others. The need for coordination led to research on "awareness", which emerged both as a conceptual tool to understand how multiple users coordinate their activities and as a design goal for groupware design [47]. A design implication of awareness is making the actions and the state of individual users publicly available [17, 23].

At the same time, *awareness depends on the active monitoring of actions by individual users* as much as it depends on their public availability [47]. Because human attention is limited, the public availability of information does not guarantee individual users' awareness of others' actions. The disparity between what is visible and what is monitored can be significant during episodes of attentional disconnect [21], or when the joint activity is conducted on larger workspaces, where users visually attend to different regions of the workspace at a given time [8, 32]. This has potential implications for design. Rather than assuming users' awareness of public information, the system can track users' locus of visual attention and adapt the interaction accordingly.

In this paper, we investigate adapting the access rights based on how multiple users visually attend to the interface and each other's actions on a shared display. During collaboration, users can switch between working on different tasks in parallel to working in tight coordination on the same screen region [29, 32, 53], leading to different visual attention configurations. In return, actions can require varying degrees of oversight or consensus based on their scope or reversibility. These demands have traditionally been satisfied by access control models that are based on *who performs the action* (such as when editing or viewing rights are restricted to particular users). In contrast, we investigate how traditional access control models can opportunistically be relaxed if users are visually attending to each other's actions. Consider a document that can only be edited by the document owner under traditional access control. Access rights for the document can be relaxed to enable editing by other users if the owner of the object is paying visual attention as visual monitoring increases owner's capacity to intervene and keep track of the changes. Similarly, certain commands that cause global changes in the workspace can be restricted to input with joint attention to prevent interruptions to other users.

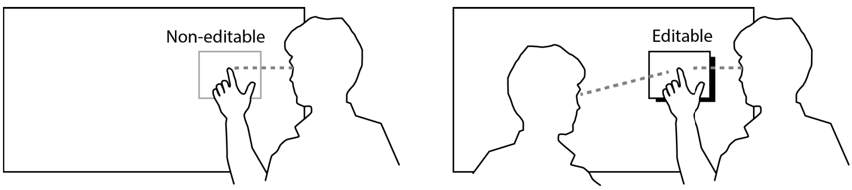


Fig. 1. An example of visual attention-based access: Actions can be configured to require the joint attention of multiple users.

In the rest of the paper, we motivate visual attention-based access by reviewing previous work on shared workspaces and discuss how our work relates to the long-acknowledged ([17, 22]) trade-off between awareness and individual power in groupware design. Then, we present a framework for visual attention-based access and introduce 4 different access types based on their availability in solitary and joint attention situations. *Universal* actions are available under any attention situation, *consensual* actions require the joint attention of all the relevant users, *supervised* actions become available to other users during joint attention (under the supervision of a particular user such as owner/supervisor) and *private* actions become unavailable during joint attention. We conducted an exploratory study with 20 participants in pairs, where participants were instructed to assign these access types to various actions (such as editing, moving, deleting) in 3 different task scenarios on a large vertical display that tracked their head orientation. The data from the study showed salient differences between user preferences based on the type and scope of an action, as well as distinct reasons for assigning the same access type. We draw on these findings to generate a list of design implications for future work. Overall, our contributions include the novel concept of visual attention-based access, an initial framework for visual attention-based access and empirical findings and design implications derived from the user study.

2 BACKGROUND

2.1 Visual Attention and Coordination on Shared Workspaces

To facilitate coordination, researchers often contrast two, sometimes competing approaches, namely allowing users to coordinate their actions through dynamic self-organization or structuring the interaction through access control, or "role-restriction", mechanisms [17, 41]. Here, we outline these two approaches and discuss how they relate to visual attention.

As a design approach, self-organization aims to increase users' own ability to avoid conflicts and keep their actions relevant. Because users' capacity to self-organize depends on their knowledge about other users' actions and intentions, maintaining workspace awareness has been identified as an important goal [17, 22]. Most systems rely on "passive awareness" mechanisms [17] that automatically make actions available to other users. On the other hand, users' awareness of publicly available information depends on their ability to monitor the workspace, which is ultimately influenced by various design decisions. Here, designers of groupware systems are often confronted with a trade-off between individual power and group awareness: The more flexibility and power individual users have the less aware they are of each other's actions [17, 22]. For example, the ability of individual users to view different parts of the workspace at the same time (as in relaxed WYSIWIS [what you see is what I see] interfaces) provides flexibility, but potentially decreases their awareness of others' actions as they can visually attend to different parts of the workspace. This led to various design interventions to work around the trade-off and facilitate awareness while maintaining individual power [22]. Some of these interventions address the challenge of divergent visual attention by indicating the visual attention of other users. For example, the system can provide visual feedback about other users' field of view through view windows [7] or multi-user scroll bars [24]. These representations act as a rough estimate of other users' visual attention, but more recent work has employed eye tracking to make collaborators' gaze points available to each other for remote [11, 13, 14] and collocated (on a large shared display) collaboration [59]. Overall, these examples address the challenge of divergent visual attention by facilitating awareness about other users' attention, but do not alter the way user actions are handled by the system.

Another design approach for coordination has been access control. Access control for groupware determines the conditions under which computational resources become available. A well-known example of access control is edit/view permissions for specific users or roles in online collaboration tools (e.g., [1, 2]). Such "access-matrix" or "role-based" models, however, are not the only means for managing access (for a review, see [54]). For ubiquitous computing applications, role-based models have been extended to take various contextual information (such as the time of day or the location) into account, resulting in context-aware access models [12, 34]. Overall, an important motivation for utilizing access control has been addressing security concerns during collaboration [54], but as Dourish and Bellotti argue, access control also contributes to a heightened workspace awareness due to decreased uncertainty about other users' actions [17].

To summarize, while both design approaches foster awareness, they put different demands on users' visual attention. Relying on self organization, especially in the absence of established social protocols, introduces uncertainty and requires situational awareness, which for many interfaces relies on visual monitoring. In contrast, awareness through access control can be achieved without users having to monitor each other's actions, since role-restriction rules out the possibility of another user performing unwanted actions or viewing private information. On the other hand, the certainty provided by access control comes at the expense of individual power.

With visual attention-based access we aim to work around the trade-off between awareness and power by relying on self-organization when users visually attend to an action and restricting access when they do not. In doing so, we address what we identify as a gap in research: While previous work acknowledged the role of access-control in fostering awareness [17], *how traditional access control models can opportunistically be relaxed if users are visually attending to each other's actions* has not been explored. Our approach is similar to context-aware access models [12], but instead of location, presence or time, which has been the focus of earlier work on ubiquitous environments, we focus on visual attention due to its significance for awareness.

2.2 Coordination on Single Displays

Single display groupware reflects the trade-off between awareness and power described above, but also comes with particular considerations for coordination. Single display applications are technically WYSIWIS, since the interface state is uniform across users. The uniformity of the interface and the visibility of off-screen actions provide a common reference for users [22, 27, 53]. Thus, awareness on single displays has been treated as less problematic when compared to remote collaboration on shared workspaces [22]. However, *what is actually seen* by different users at a given time can diverge significantly due to the larger display size and users working in a "loosely coupled" manner (i.e., independently on different interface regions without coordinating their actions [29, 32, 53]). Coupling has been observed through different metrics such as task focus [29], spatial arrangement around the display [53] and, most relevant for our study, visual attention area as measured by head orientation tracking [32]. A general observation from these studies is the ease with which users fluidly switch between loosely coupled individual work to tightly coupled joint work [29, 32, 53]. As with remote collaboration, different levels of coupling on single displays come with various trade-offs. While loose coupling can facilitate fast, concurrent work, it can also result in duplicate efforts due to lack of awareness [29] or decreased concern for building consensus [8]. In a competitive use scenario, Birnholz et al. observe that concurrent input by multiple users on a single display enabled fast task execution but also resulted in users acting more in their own interests (when compared to single input) [8]. They explain this through the decreased likelihood of users to scrutinize others' actions due to their preoccupation with their own input that has led to decreased awareness. Along the same lines, Mayer et al., observe that competitive tasks result in an increased effort by participants to monitor the whole display space [39]. The studies discussed so far focus on semi-public settings in a confined space within a small group, but similar considerations have informed the design of public displays in urban environments [19, 42, 44]. For example, Fischer and Hornecker regard public visibility of actions as one of the reasons that people refrain from posting inappropriate content when interacting with urban media facades, but otherwise do not report to what degree appropriate behavior relies on the presence, and thus the monitoring, of other people [19].

Lack of customization for individual users is an additional source of conflict for single display interactions. Previous work has documented instances when navigation and object manipulation actions performed by one user interrupted another or when users competed for interaction area on public or semi-public displays [30, 44]. Users can to a certain extent avoid conflicts by assigning territories, employing turn-taking procedures or other social protocols [19, 44, 48, 55]. Importantly, successful self-organization such as turn-taking assumes users' awareness of contextual factors (e.g., who else is interacting with display or whether someone else is waiting in the line [19, 44]). Even so, conflicts can still occur when users knowingly interrupt others [44].

Another line of research has taken a more structured approach to conflict avoidance on single displays. This led to work on interaction mechanisms for conflict resolution [41] and providing access control in a way that is similar to access-matrix models in remote groupware but without having to rely on device-based login information. Thus, previous work utilized hand gestures [58], touch capacitance [15, 40], fingerprint patterns [26] and proximity to an interactive surface [4, 33, 56] to identify users and manage access. For example, manipulation of objects can be restricted to touch actions by authorized users [26] or global changes can be programmed to require the parallel input of all the users in the form of "cooperative gestures" [40]. Some previous work also utilize user input as a proxy for visual attention to control view access permissions. For example, personal information can be shown only when users gets close enough to the screen so that their body acts as a visual obstacle against information voyeurism [56]. Additionally, various view-dependent projection [38], shutter glass [3, 36, 51] and proximity aware [16] techniques have been developed to build multi-view tabletops that provide personalized views for individual users on the same display. However, while various access mechanisms have been devised to determine what items are editable or viewable by different users, managing access based on who *visually attends* to an action or an object at a given time has not been explored. The closest work is the cooperative game GazeArchers [45], which requires both users to look at moving targets when shooting. In this case, visual attention input is used to add additional challenge, rather than to address awareness and conflict issues.

3 VISUAL ATTENTION-BASED ACCESS

Visual attention-based access determines what actions are available based on who visually attends the action. In this section, we distinguish between solitary and joint attention situations during synchronous collaboration and describe 4 different access types that are generated from their combination.

3.1 Visual Attention Situations

In contrast to previous observational studies that categorized attention patterns at the workspace level (how different users visually attend to different areas and to each other) [29, 32, 53], our basic unit is a single visual target; we are interested in how a specific visual target or action is visually attended to at a given time. This leads to a basic distinction between solitary and joint attention situations (Figure 2).

Solitary attention refers to situations in which an item or action is visually attended to by a single user, while other users are away or attending to another area of the workspace. In some cases, the attending user can be the owner of the object, which requires special consideration for access management.

Joint attention refers to situations in which an item or an action is visually attended to by all the relevant users.

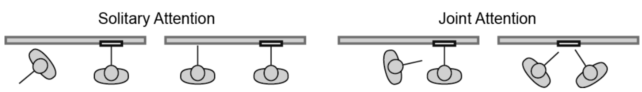


Fig. 2. Examples of solitary and joint attention situations on a visual target at a given time.

3.2 Visual Attention-based Access Types

Actions can be assigned 4 different access types based on their availability in solitary and joint attention situations:

Universal actions can be accomplished under both attention situations. This access type can also be defined as lack of any role restriction. Potential use cases are actions that are not very critical, that can be easily reversed without serious consequences or when users' self-organization alone is sufficient for coordination.

Consensual actions require joint attention to be accomplished. They expand on the concept of cooperative gestures ([9, 10, 37, 40]), which require coordinated input from multiple users for the realization of certain actions (such as actions that affect the whole workspace [40]). However, while cooperative gestures require synchronous

Table 1. Types of actions that are available (●), unavailable (-) or only available to a particular user (◐) under solitary and joint attention situations.

	Action can be accomplished	Solitary Atten.	Joint Atten.
Universal	under any attention situation	●	●
Consensual	only under joint attention	-	●
Supervised	if object owner or supervisor is attending	◐	●
Private	only if the owner is attending and no one else	◐	-

manual input, consensual actions are based on visual attention. This is based on the assumption that monitoring alone increases the capacity of users to intervene or prevent conflicts. Consensual actions potentially require less effort when compared to manual consent mechanisms (e.g., cooperative gestures [40] or confirmation buttons), since users can already be visually attending to the action. However, they are more restrictive than universal access.

Supervised actions require the attention of a particular user (instead of all users as in consensual). They are enabled for joint attention situations and also for solitary attention situations as long as the attending user has special rights. Potential use cases are actions that can benefit from the awareness of a particular user such as the supervisor of a session or owner of a document. When compared to traditional access rights management, which strictly restricts actions to a specific user, supervised access relaxes access by enabling actions by others if the particular user is monitoring. Both consensual and supervised actions are positive access criteria, because they require the visual attention of certain user(s) for access.

Private actions can only be accomplished if a particular user (such as the owner) is attending and no one else. The private access category is the counterpart of consensual access; it is a negative access criterion as the action becomes unavailable during joint attention. While consensual and, to a certain extent, supervised actions enforce awareness, private actions enforce privacy. Private access aims to limit "information voyeurism" of private information or actions on public displays [52]. This makes it similar to other solutions that provide personalized views for individual users (e.g., [3, 36, 51]), but instead of dedicated hardware, such as shutter glasses, private access relies on visual attention information.

Table 1 summarizes the availability of each access type under different attention situations. Note that we so far defined solitary and joint attention situations, respectively, as involving single and all relevant users. The distinction is straightforward for two users, but it would be more appropriate to view solitary and joint attention as a continuum for larger groups. In intermediary situations, such as when a subset of users is attending to an action, the system can be either strict or flexible regarding how it grants access. For example, consensus can strictly be interpreted as requiring the attention of all relevant users or, flexibly, a subset of users. We discuss these different approaches later in the paper in terms of scalability.

3.3 Input Handling

When an initiated action is available for a given visual attention situation, the system grants immediate access by executing the action promptly. A mismatch between the visual attention situation and the access type, however, can be handled in different ways:

Restricting access entirely, without any further interaction, is the least complicated approach. It is also the handling method we employed in the user study due to its straightforwardness.

Deferring action execution until additional input is another method. For example, an action initiated by a solitary user can be later confirmed by another user. When compared to full access restriction, deferring allows for more individual power, but also comes at the expense of real-time awareness and associated risks such as duplicate or irrelevant contributions.

Notifying other users is a real-time alternative to deferring. If a consensual action is initiated under solitary attention, the system can evaluate whether other users are visually attending to the workspace. If so, the system can provide a visual notification of the action near the users' locus of visual attention to inform them about the ongoing action (similar to previous work on gaze-adaptive visual feedback [49, 50] for single user applications). Notifications work around the trade-off between power and awareness even further when compared to restricting the action. On the other hand, they can introduce another long-acknowledged trade-off in groupware design, namely the trade-off between awareness and disruption [28], especially when users would prefer to remain focused on their individual tasks.

4 STUDY: VISUAL ATTENTION-BASED ACCESS PREFERENCES

Having defined different visual attention-based access types, the question remains how the framework can be put into use in collaborative applications, specifically, which access type makes sense for a particular action (such as editing or deleting an item) and on what grounds. We conducted an exploratory study to answer this question and find out user preferences and motivations related to different access types. Participants (in pairs) were tasked to decide which actions should belong to universal, consensual, supervised and private access types as they completed three different use scenarios of project planning, brainstorming and document sharing. The scenarios and their applications have been prototyped for a large vertical interactive display that was situated in a meeting room.

While visual attention-based access can be prototyped using different hardware setups, we chose the interactive whiteboard setup due to a number of practical reasons. First, we expected the collocated, single-display setup to increase participants' awareness of whether their input is performed under joint or solitary attention. Secondly, the setup was chosen to facilitate verbal feedback and deictic references to elements on the workspace (e.g., pointing to different elements while explaining their preferences) for data gathering purposes. In addition to providing a clear solitary–joint attention distinction, the 2-user design ensured equal centrality for the participants, as participants could position themselves to the left and right of each other.

The assignment of access types had instant effect, allowing the participants to immediately observe and test the effects of their preference. Additionally, the joint nature of the assignment task enabled us to observe the agreement process between the participant pairs (i.e., the participants' externalized reasoning and discussions about why a certain action should be available or unavailable for different attention situations). Overall, the research questions that motivated the study were:

(1) What are the access type preferences for different actions?
(2) What are the motivations reported by participants as they assign access types to different actions?
(3) How do the particular qualities of visual attention-based access manifest in participant preferences and interactions?

4.1 Apparatus

The study has been conducted on a 2,05 × 1,20 meter vertical interactive surface consisting of three adjacent displays, each with a resolution of 1080 × 1920 pixels. The displays were able to distinguish between touch and pen input using IR image recognition. In all of the applications, pen input was mapped to editing content while touch input was mapped to moving the elements on the screen. Both could be used for other actions that are accomplished by buttons. Head position and orientation of users were tracked (Figure 3), by an OpenCV application that detects head-worn markers using a web camera (running at 640 × 480 pixel resolution) mounted at the ceiling. Both touch and head tracking data were transmitted to the web-based whiteboard applications through web sockets. We determined whether a visual area is attended by a participant by scoring the visual attention information using visual angle (θ) and distance (d) values between the head and the target on the screen (Figure 3). The participants' gaze points were made visible on the screen to indicate how the system sensed their visual attention.

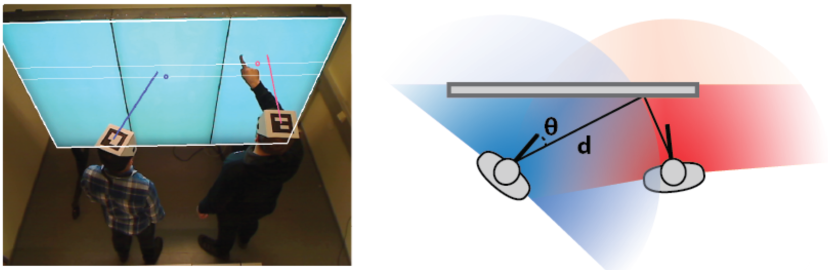


Fig. 3. Detection of head position and orientation in relation to the vertical interactive screen. We determined whether a participant is attending to a target by scoring the visual angle (θ) and distance (d) values between the head and the target on the screen.

4.2 Scenarios

The multiple-application design of the user study was intended to observe the potential commonalities in user preferences across actions in 3 different scenarios (Table 2). The tasks have been created so that participants could accomplish them without any need for specific knowledge, by drawing on their daily experiences.

4.2.1 Project Planning (PP). The first scenario involved creating different "to-do" elements for a hypothetical project that aims to "decrease the energy consumption of households in Helsinki". The project planning scenario and related actions was inspired by the use of whiteboards in agile and scrum software development methods, in which different tasks are represented as cards that can be assigned to different individuals [18, 20]. The application, in total, supported 11 actions to which participants were instructed to assign access types. 8 of these actions were element based including *adding*, *editing*, *viewing*, *moving*, *changing the owner of* and *deleting* to-do elements and *editing* and *viewing* a personal calendar that was assigned to each participant. The remaining 3 actions were global and included *aligning all to-do* elements (that vertically positioned them based on their owners), *deleting all to-do* elements and *exiting the session.*

4.2.2 Brainstorming (B). The second scenario required participants to brainstorm for content ideas for a website about "life in the city" by writing their ideas on post-its. In contrast to to-do elements in the project planning task, post-its did not have any owners. The application supported 7 actions. Five of these actions were element based including *adding*, *editing*, *viewing*, *moving* and *deleting* post-its. The remaining 2 actions were global and included *deleting all post-its* and *exiting the session.*

4.2.3 Document Sharing (DS). The last scenario involved sharing personal documents for a magazine layout project. This scenario builds on previous research that investigated access management for personal media on shared displays (e.g., [30, 40, 46]). The application allowed participants to place different personal elements from a menu including two article drafts (the participants were told that they had composed the articles), one personal

Table 2. Overview of use scenarios and the list of object-level and global actions for each application. In each of the use scenarios the participants were instructed to "Discuss and assign attention based access rights for different items and action".

Task description	Action list
Project planning. Imagine that you are tasked to make a plan and different "to-do" items for a project that aims to decrease the energy consumption of households.	**To-do (object):** add, draw/edit, view, move, change owner, delete **Personal calendar (object):** edit, view **Global:** align all to-dos, delete all to-dos, exit session
Brainstorming. Imagine that you are in a brainstorming session about gathering new ideas to make a website about life in the city of X.	**Post-it (object):** add, draw/edit, view, move, delete **Global:** delete all post-its, exit session
Document sharing. Imagine that you are working on a layout project that require you to go over your own documents.	**Document (object):** annotate, view, move, remove **Global:** pile all media, remove all media, exit session

bookmark element and one personal note element. While the other two applications required participants to create content from scratch, all of the elements in the document sharing scenario had an owner and were pre-configured but could be annotated on. Overall the application supported 7 actions. Four of these actions were element-level including *annotating*, *viewing*, *moving* and *removing* documents. The remaining 3 actions were global and included *piling all media* (moving all the documents to the left-hand side of the screen), *removing all media* and *exiting the session*.

Note that, the three applications featured elements with different levels of ownership. To ensure the relevance of supervised and private access types for global actions and elements with no owners, the participants could assign a general "session master" who could act as a substitute for the owner. Overall, the actions across different applications can be analyzed through these two dimensions:

Element-level/Global: Previous work distinguished between element-level (e.g., editing a single element) and global actions (e.g., piling up all the elements on the workspace) on shared workspaces [40]. We were interested in observing whether participants would select higher attention demand access types (consensual and supervised) for actions with global scope, in other words, whether previous insights from cooperative gestures ([40]) would extend to visual attention-based access.

Action typology: While scenarios involved different elements, the actions they support can be grouped under different typologies such as editing, viewing, moving or deleting. Crucially, different actions types had different levels of reversibility; moving or creating new elements could easily be reversed, but deletion was irreversible. We expected irreversible actions such as deletion to be assigned more restrictive access types (instead of universal). Additionally, some actions such as moving or drawing result in gradual changes that enable the other participant to intervene, while others such as deletion result in discrete and sudden changes. To make these two types of actions comparable, discrete actions required a continuous press on the button for 1 seconds, accompanied by a horizontal progress bar for visual feedback.

4.3 Access Type Control Panel

Participants could select an access type for an action anytime during the session through a contextual control panel (Figure 4) by pointing at one of the 4 access types. We had initially implemented a central menu to control the access types of all the instances of the same action, but during pilot studies this proved to be too abstract and not flexible enough to match the evolving and element-specific participant preferences. Thus, we opted for an object-based control interface that contextually appeared under the related interface element.

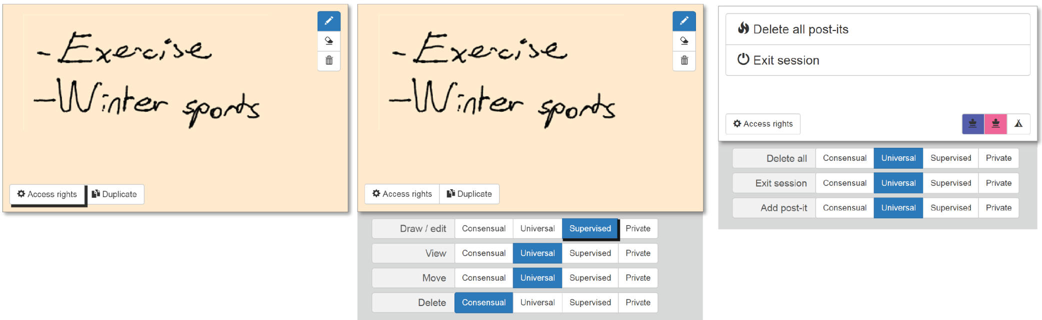


Fig. 4. Each element (such as the post-it element on the left) or the global action menu (right) had a collapsible "Access rights" menu that allowed participants to assign and modify their access preferences for different actions by pointing at one of the 4 access types using touch or a pen. The "duplicate" option allowed the creation of an item with the same access preferences. By default, the access configuration for all actions were set to universal. Other buttons allowed switching between drawing and erasing modes, as well as deleting the element.

4.4 Procedure

Each session started by introducing the purpose of the study, the interaction basics and the access type control panel to the participants. The participants were then given a demonstration of how different access types (universal, consensual, supervised and private) behave differently in each of the visual attention situations. After participants felt comfortable with interaction basics and understood access types (after 5–10 minutes), they proceeded to complete the tasks for each scenario in the order of 1) project planning, 2) brainstorming and 3) document sharing. The scenarios were respectively allocated 30, 15 and 15 minutes, based on their complexity and the amount of actions that need to be assigned different access types (Table 2). Apart from the task and time constraint, the sessions were unstructured. The participants could perform different actions and change their access type preferences anytime during the task, allowing them to experience and discuss different configurations before making a final decision. Additionally, each application featured two draggable information sheets as a reminder of different access types and the task. The participants were instructed to assign access types jointly (except for personally owned items) in order to make them take different considerations into account and observe their verbal reasoning.

Toward the end of each task, the participants were asked to finalize their preferences for actions and explain why they assigned the particular access types for different actions. After completing all of the tasks, the participants were interviewed about their general impression of visual attention-based access. The sessions approximately took one and a half hours.

4.5 Data Collection

The system logged participants' access type assignment actions and their final preferences. Additionally, the sessions were video-recorded to observe participants' interactions and discussions. The videos were later transcribed to link participants' interactions and discussions with their access type assignment actions. The semi-structured post-study interview inquired about participants' overall impression of visual attention-based access by drawing on their experiences and how they compare it with traditional access control.

4.6 Participants

20 participants (9 female), aged 18 to 35 (m = 26.4, sd = 4.5), were recruited for 10 sessions through university email channels and bulletin boards. The participants included 11 undergraduate and master level students, 7 researchers (PhD candidate and post-doc) and 2 designers (1 front-end developer and 1 interaction designer). Participants, on average, reported moderate previous experience with brainstorming (m = 3.3, in a scale of 1–5), project planning (m = 3.1) and sharing personal media (m = 3.2). They were compensated with two cinema ticket vouchers and their informed consent was collected for data logging and video recording.

5 RESULTS

We present the quantitative results based on log data and qualitative results derived from participant comments and actions. The qualitative results are denoted by the session number and the initials of the scenario during which the comment has been recorded (a participant remark recorded in session 4 during project planning scenario is indicated as "S4, PP" and only as "S4" for data from the final interview).

5.1 Access Type Assignment Process

In many cases, the assignment of a specific access type for an action was the result of participants' ongoing interactions with each other and the prototype. Although the sessions were unstructured, we observed that the assignment process was often influenced by 3 different factors:

Discussions among participants were observed in the form of references to prior experiences (e.g. workplace meetings or student committees) and verbal reasoning about why a specific action should be assigned a particular access type. Some of the arguments were presented as practical considerations. For example, in one session a participant questioned the other participant's idea of assigning a supervised access type: *"But this supervised still gives some possibility of misuse that one will change or two will change secretly... I think maybe it should be mostly consensual"* (S1, PP). In other cases, they were presented as personal preferences. For example, in the following exchange, participants (S2, PP) were expressing their priorities (in this case awareness vs. privacy) when deciding on whether viewing personal calendars should be private or supervised:

A: *It is good that you can look my calendar so you know.*

B: *No, I don't want anyone to look my calendar it is my own private stuff.*

Practical implications of assignments were another resource for deciding on a particular access type, as the prototype allowed participants to immediately experience the effects of their assignment. In some cases, the practical implications led to a revision of the initially assigned access type. In one session (S3, B), participants assigned consensual access to moving a post-it, but the assignment later proved to be impractical. This first led to a help request from the other participant and then to a revision of the original assignment:

B: *Can you look this way?*

A: *Ok* [A comes near B]

B: *The consensual moving is really...* [B reverts moving back to universal]

Progression of the task led to changes in participants' access preferences for some elements. For example, for to-do and post-it items it was common for participants to start editing the element with the default universal access type setting (that provides the least restriction) and then modify the access type once the element content was filled in.

5.2 Quantitative Results: Distribution of Final Access Preferences

We logged 1002 access type assignments for different actions from the final state of the application interfaces across the sessions. The number of assignments varied between sessions (min = 82, max = 166, sd = 24.3), as participants created different amounts of to-do (m = 5.8, sd = 2.3) and post-it (m = 6.3, sd = 3.4) elements. Thus, to calculate the overall access type distribution, we report the grand means that are aggregated within each session. Table 3 provides an overview of how access types were assigned to different actions.

We analyzed the distribution of access preferences for different kind of actions based on element-level–global distinctions and action typology. The data show salient differences between user preferences for these two dimensions across different applications. Universal access, which provides the least restriction, was rarely assigned to element-level delete (7.1%) and never to global delete actions, but it was overwhelmingly common for creating new items (95.0%). Universal access was also frequently assigned to view (72.3%) and element-level move actions (68.6%). On the other hand, the results show a high preference for consensual access for actions with global scope such as exiting the session (56.7%), global deletion (76.7%) and aligning elements (35%). For comparison, the preference for consensual access was lower for element-level delete (35.4%) and move (7.3%) actions. Instead, we recorded a higher incidence of supervised and private access for element-level actions (Table 3). These results are in line with previous design insights that propose joint control for actions with global scope [40] and our expectation that elements such as delete will be assigned more restrictive access types.

We also observed a degree of co-occurrence between different action types that belong to the same element, when we collectively analyzed to-do (n = 58), post-it (n = 63) and document (n = 80) elements that all supported editing, viewing, moving and deletion actions (Figure 5). For example, for elements that had editing actions set to private access (n = 35), a high proportion of them would also be assigned private access for deletion (n = 27), viewing (n = 23) and moving (n = 15). On the other end of the spectrum, for elements that had deletion action set to universal access (n = 13), a majority of them would also be assigned universal access for moving (n = 12), viewing (n = 9) and editing (n = 8). Thus, participants' access preferences for certain actions can help infer their preferences for other actions that belong to the same element.

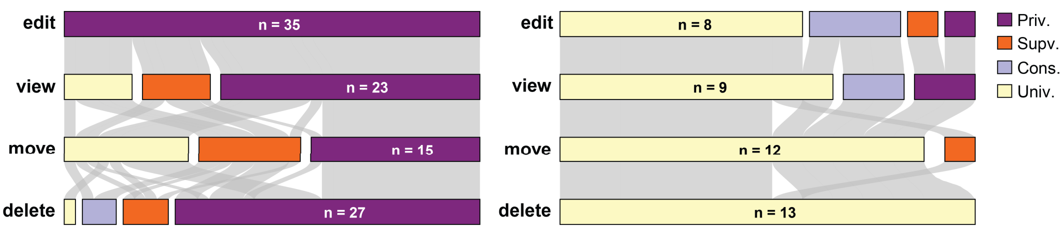


Fig. 5. Two examples of co-occurrence: The distribution of access types for different actions of the same element when editing is set to private (left) and when deletion is set to universal (right). The data is across to-do, post-it and document elements. Each continuous line represents a group of elements that have the same access control settings for 4 action types.

The distribution data provide us with a summative view of overall patterns. However, it does not answer what accounts for the differences between user preferences for the same actions and interesting patterns in the data. For example, while the assignment of private access to view actions is foreseeable, we also recorded many instances in which private access has been assigned to manipulation actions and in rare cases even to actions with global scope such as global delete (3.3%). Below we report qualitative data that give insights about participants' reasoning when assigning different access types.

Table 3. Access right preferences for different actions in different scenarios across all sessions. The number of individual to-dos and post-its varied between different sessions. Thus, the results show grand means that are aggregated within each session. The bottom block shows the aggregated grand means based on action type.

	Action	Level	Type	Access Preferences (%) Univ.	Cons.	Supv.	Priv.
Project planning	Add to-do	Element	Create	90.0	10.0	0.0	0.0
	Draw / edit to-do	Element	Edit	22.8	26.8	45.9	4.5
	View to-do	Element	View	89.5	4.0	0.8	5.7
	Move to-do	Element	Move	60.8	11.3	27.0	0.8
	Change owner to-do	Element	Transfer	19.0	11.3	58.1	11.7
	Delete to-do	Element	Delete	8.0	34.3	51.2	6.5
	Edit calendar	Element	Edit	15.0	10.0	25.0	50.0
	View calendar	Element	View	45.0	0.0	25.0	30.0
	Align all to-dos	Global	Move	60.0	30.0	10.0	0.0
	Delete all to-dos	Global	Delete	0.0	80.0	20.0	0.0
	Exit session	Global	Exit	30.0	60.0	10.0	0.0
Brainstorming	Add post-it	Element	Create	100.0	0.0	0.0	0.0
	Draw / edit post-it	Element	Edit	64.8	15.8	15.4	4.1
	View post-it	Element	View	94.6	0.0	1.2	4.1
	Move post-it	Element	Move	87.3	8.2	1.2	3.2
	Delete post-it	Element	Delete	8.2	63.0	25.0	3.8
	Delete all post-its	Global	Delete	0.0	70.0	30.0	0.0
	Exit session	Global	Exit	20.0	50.0	30.0	0.0
Document sharing	Annotate document	Element	Edit	30.0	2.5	32.5	35.0
	View document	Element	View	60.0	0.0	16.2	23.8
	Move document	Element	Move	57.5	2.5	10.0	30.0
	Remove document	Element	Delete	5.0	8.8	27.5	58.8
	Pile all media	Global	Move	30.0	40.0	30.0	0.0
	Remove all media	Global	Delete	0.0	80.0	10.0	10.0
	Exit session	Global	Exit	0.0	60.0	40.0	0.0
		Element	**Create**	**95.0**	**5.0**	**0.0**	**0.0**
		Element	**View**	**72.3**	**1.0**	**10.8**	**15.9**
		(Element)	**Move**	**68.6**	**7.3**	**12.8**	**11.4**
		(Global)	**Move**	**45.0**	**35.0**	**20.0**	**0.0**
		Element	**Edit**	**33.1**	**13.8**	**29.7**	**23.4**
		Element	**Transfer**	**19.0**	**11.3**	**58.1**	**11.7**
		Global	**Exit**	**16.7**	**56.7**	**26.7**	**0.0**
		(Element)	**Delete**	**7.1**	**35.4**	**34.6**	**23.0**
		(Global)	**Delete**	**0.0**	**76.7**	**20.0**	**3.3**

Table 4. Summary of participant comments for different motivations for assigning different access types.

	Motivation	Example comment
Universal	Unimp./Reversible	"...but this one the edit is universal just because it is not that as important as such and somebody can expand on it." (S10, PP)
	Convenience	"...but while we are setting up this plan because we are alone in this room we can keep this as universal for usability." (S9, PP)
	Trust / Cooperation	"This should be a list that everyone should be able to add to when they figure something out so I let it as universal. I trust my colleagues." (S5,PP)
	Awareness (view)	"I also want everybody in this room to see what I have written here because it is some common data useful for project." (S9, PP)
Consensual	Prevent accidents	"Consensual actions is really good because you would accidentally you know destroy stuff from the screen if you are not both looking." (S6)
	Agreement	"I would say consensual makes sense in the sense that we all agree that we save and quit unless there is some hierarchy..." (S9, PP)
	Group awareness	"Consensual is pretty nice for something like this ... like we are planning to be there make sure that everybody is aware of what is happening." (S3)
Supervised	Permission	"yes you can not remove it without my permission and yeah that is fine right." (S2, DS)
	Owner's awareness	"If you made a mistake in your work it can be supervised and you can see the person modifying your content." (S7, DS)
	Scalable	"This is more like a topic which can be debated upon a lot so instead of making it consensual... I think it is up to the chair it should be." (S4, B)
Private	Privacy	"If you have ideas that you are not certain of yet and you don't want others to bother you about them..." (S5, B)
	Non-attention acc.	"For me editing calendar is private I want to do it myself." (S4, PP)

5.3 Qualitative Results

We analyzed the video recordings and post-study interviews through an open coding process to classify observations and participants' statements for each access type. This was followed by the grouping of participant statements to identify distinct motivations within each access type (Table 4). The analysis particularly focused on how participants referred to the specific affordances of visual-attention based access control.

5.3.1 Visual Attention-based Access for Preventing Conflicts. As expected, a common motivation for assigning consensual and supervised access types was conflict prevention in the absence of joint attention. Consensual access has been frequently assigned to actions with global scope (Table 3). Participants stressed that they *"have to both agree to end it [the session]"* (S7, PP) or stated that *"if I want to delete all to-dos it means that everyone agrees on something"* (S4, PP). Besides maintaining agreement, another motivation for consensual access was preventing accidents: *"I don't think anyone should be able to delete all of them easily so we should all be there to see"* (S2, B). In a few instances, consensual access was assigned after the action of one participant interrupted the other (to prevent further interruptions): *"If it troubled you we should do it... so they can pile only if it is consensual"* (S2, DS).

Participants similarly used supervised access type to prevent unwanted actions and described it as a "permission" mechanism: *"I don't want you to move this without my permission"* (S5, PP). Another participant stated that *"... it is better to have this supervised access types so the other could not delete it [personal task] or change it as he wants"* (S6, PP).

Conversely, lack of need for conflict prevention can partially explain the reasoning behind universal access. Universal access has frequently been preferred for creating, moving and viewing elements (Table 3). Actions such as creating new objects or moving are easily reversible, and this was reflected in participant comments: *"Anyone in principle can add stuff because we can always delete stuff"* (S2, PP). A related motivation was the cooperative setting of the task: *"If it is a competition, I would understand using them [private access] but this is just like brainstorming and creating to-do together"* (S8, PP). Another participant stated that both of them *"are on the same side it does not make sense that there is some guy with malicious intent"* (S2, B).

5.3.2 Visual Attention-based Access for Maintaining Awareness. In some cases, however, consensual and supervised access types have been motivated not through conflict prevention but as a means for facilitating awareness. In these cases, the participants utilized visual attention-based access restrictions as a way to ensure that the action performed will be attended by themselves or the group. For example, in one session participants set "change owner" to consensual so that *"we know who is actually in charge of this"* (S10, PP). During the task and later in the interview, consensual access was suggested as an explicit tool for enforcing the attention of the group, to help the *"group to focus on the single thing when needed"* (S10, PP). In another example, one participant assigned supervised access to moving: *"I would like to know where this goes... so it [moving] should be supervised by me."* (S5, DS). Even though moving is easily reversible, supervised access makes tracking the changes easier by increasing the restrictions on the other user.

In these instances participants traded individual power with awareness. On the other hand, restrictions to individual power were also perceived as inconvenient. For example, one participant described consensual moving as a *"big hassle to have every one look at same place even if it is just two of us. I noticed that when I was not being able to move stuff"* (S3). Accordingly, a lack of restrictions to individual power can explain the preference for universal access, which was perceived as convenient: *"This is like nothing right now ... Because universal is the most easiest setting I guess ... this is not an idea that is necessary for this ... like random stuff"* (S8, PP). One participant stated that he would be willing to prioritize easy editing at the expense of awareness: *"Okay, I changed it to universal because at the moment I want to add ideas I am putting an input and if the people are not aware at the time of putting the input it is okay"* (S4, PP).

5.3.3 The Influence of Element Content. Our observations and participant remarks during the session revealed that the content of individual items can partly account for the variance of access types within element-level actions. In project planning and document sharing tasks, we observed that newly created items (with no content) were often left in their default universal access and were assigned another access type only after participants filled in some content. In some cases, access rights remained universal even after editing due to the unimportance of the content, such as such as for *"temporary ideas"* (S9, B) or *"random stuff"* (S8, PP). As the task progressed, participants set aside certain items as their *"general schedule"* (S1, PP), *"best ideas"* (S7, B) or things *"they both agreed on"* (S5, B) and assigned consensual access to the actions associated with the item.

On the other hand, supervised access type was often motivated by the personal content of the element: *"And those ones are more supervised because it is our actual personal tasks"* (S7, PP). In these cases, participants stressed the need for their visual attention for accessing the document: *"Because they are [pointing to his articles] articles if anyone wants to make change to this I have to be there to see what is happening"* (S4, DS). Supervised access has also been assigned to view actions (Table 3). In one instance, this was motivated by preventing other people from viewing personal information first: *"... cause sometimes in the office you come to the calendar and people have already seen how your days are going to look like..."* (S10, PP). The same consideration also explains the assignment of private access. Participants referred to certain items as *"personal stuff"* or *"personal notes"* and made viewing these elements private: *"my bookmarks should also be private, no one can view it"* (S4, DS).

5.3.4 Private Access as a Proxy for Traditional Access. Even though personal content of an item was a reason for assigning private access, we also found counter-examples to this motivation, particularly when private access has been assigned to actions other than viewing. For example, one participant explicitly ruled out privacy as a concern when assigning private access type: *"It is not other person should not see it... because it is my bookmarks so I want to be the only person who can delete it from the screen."* (S7, DS). We also recorded other statements that emphasize limiting the access to the owner: *"I don't want anyone else to be able to edit them, because this is my own personal text"* (S5, DS). Unlike supervised actions, private actions are strictly restricted to a single person as they become unavailable if another person is visually attending. This also minimizes the risk of giving unwanted access. One participant highlighted this aspect of private access as a reason for selecting it over supervised access: *"talking about private stuff that you don't want to have anyone else access to it seems like you can accidentally give someone access, but if it is private then it is really private"* (S3).

We classify this use of private access as attempts by participants to manage permissions in a way that is similar to traditional access control, namely based on who performs the action rather than who attends to the action. In other words, participants assigned private access not due to privacy concerns but to strictly restrict access to themselves. While private access can be used in this way, it comes with disadvantages that we observed during the session. When another user is visually attending to the same area, private access either requires him or her to look away or it unnecessarily restricts the access of the owner. This has been highlighted as a shortcoming of visual attention-based access during one post-task interview: *"... so there is private you can edit the document when no one is looking... in supervised everyone else can when you are looking... but there should be like that kind of private if someone is looking you can edit but he can't"* (S7). We, therefore, interpret this use of private access as a shortcoming of visual attention-based access control.

5.3.5 Scalability. Even though the study involved two users, participants reflected on the scalability of joint attention for consensual access, particularly the difficulties that can occur for larger groups. One participant described the situation as *"hard to get everyone involved at the same time then it is hard to make any decisions"* (S5). Another participant stated *"If there is two of us, consensual is easy but if there is tens of like 20 coworkers it is hard everyone has to be present to do all of this and it is kind of unnecessary"* (S8, PP).

The concerns about scalability can also explain the assignment of supervised access to actions with global scope (such as global deletion or exiting the session) when there is a session master. Instead of requiring the visual attention of all users (as in consensual), participants proposed using supervised access for global actions: *"If it is consensual then when we exit session everyone should still be here but if supervised someone can leave little early"* (S1, PP). While supervised access can be used to address scalability issues, participants also suggested alternative solutions for larger groups during the post-task interviews, such as dividing the users into subgroups (*"I think there should be groups, you can include group 1, 2, 3 and put there some people."*, S2) or employing degrees of consensus among the user group (*"if we are five people we choose who is consensual, it can be consensual between two of us or all five."*, S10).

5.3.6 The Trade-off between Convenience and Certainty. During the final interview, participants were also asked to give their general impression of visual attention-based access and compare it to dedicated confirmation mechanisms (to make this more concrete participants were given the examples of hitting confirmation buttons on personal devices or a shared screen). Not having to provide manual input was described as convenient: *"It is very convenient because we don't have to move our hands just look."* (S6). This was related to avoiding another step: *"It may be more natural, I think, visual based [access]. When you have to push a button or receive something on your phone there is another step that adds to it and if you are doing it with visual based access right it can be quicker."*

(S7). On these occasions visual attention-based access was compared favorably with explicit confirmations: *"You have freedom everyone looks at the screen and I am gonna delete this and do this... if everyone is confirming then it is a lot bigger process."* (S6).

On the other hand, visual attention-based access was at times described as uncertain. This was related to the perceived uncertainty of measuring attention (*"It seems a bit less exact to be... measuring attention does not seem quite that exact way to manage the right..."*, S3) and potential cheating by users (*"or if you tried to spy on, is it based on head position right? You can kind of cheat."*, S6). Participants also stated that *"explicit confirmation can be more legally suitable"* (S4) and for some situations *" there should be a second layer of authorization"* (S5). One participant's comment directly pointed to a trade-off between convenience and certainty: *"When I am touching it, requires that I go and touch something. It is much more certain kind of manifestation of my attention so it requires effort from the user then the level of certainty increases also. Whereas this gaze-based attention inference it decreases this certainty thing so it could be that I am looking there, but I might just be thinking and not paying attention. But on the other side, it also helps fluidity of the interaction, and it does not enforce users to make explicit actions but is somehow like... could be much more blended in the interaction"* (S9).

6 DISCUSSION

6.1 Main Findings and Design Implications

In this section, we summarize the main findings from the study from which we derive various design implications that can be explored in future work.

6.1.1 Participants Employed Visual Attention-based Access Types to Maintain Agreement and Owner's Control. Consensual and supervised access were chosen over universal and private access types when participants were willing to grant access as long as they were visually attending to the action. In these cases, visual attention-based access was perceived as a sufficient means for conflict prevention.

- This finding suggests that manual confirmation mechanisms proposed in earlier work such as collaborative gestures for global-level actions [40] or touch confirmations [26] can partly be substituted with visual attention-based access. On the other hand, designers should be aware of the uncertainties introduced by visual attention-based access when deciding on to what extent and how they can replace traditional access control.

6.1.2 Granting Access with Head-orientation Was Found Convenient but Uncertain. Not having to manually touch the screen for confirmations was highlighted as a convenient feature. At the same time, visual attention-based access introduced uncertainty that was attributed both to a mismatch between head orientation and participants' actual locus of visual attention and also to situations when visual attention does not indicate awareness (i.e., when participants remarked that they may be looking but not paying cognitive attention). One strategy that participants employed to overcome uncertainty was the use of private access as a proxy for traditional access control. However, as we observed, this strategy comes with the disadvantage of unnecessary restrictions either to the owner's access or to other users' awareness. Thus, future work should consider alternative design options like those proposed below.

- One design approach is decreasing the level of uncertainty by employing more accurate measures of visual attention such as eye tracking [35]. Another approach would be to develop input handling techniques that account for the uncertainty when inferring visual attention from behavioral data. For example, the required level of certainty about users' visual attention can vary depending on the action type; that is, deleting an item can require a higher threshold of certainty when compared to editing. Uncertainty can also be managed by letting the users know about how the system perceives their visual attention. In the current study, this was achieved by showing users their head orientation on the screen as a circle, but insights from

eye movement research warn that salient representations of visual attention can be intrusive [59]. A more acceptable solution would be providing subtle cues on items that indicate who is visually attending to an item and what actions are available. Finally, the interface can provide additional means for overriding or undoing an action (e.g. cancel buttons) to handle situations where awareness does not equate consent.
- An alternative approach could be using visual attention-based access complementary to traditional access models instead of a complete replacement. For example, private manipulation actions such as editing can be managed using traditional access control models if the system is able to determine who performs the action (e.g., through touch identification, proximity or personal input devices).

6.1.3 Participants Assigned Access Types Not Only for Conflict Prevention but Also for Facilitating Awareness. In some situations, participants preferred supervised and consensual types not for preventing conflicts but as a means to ensure that the owner/supervisor is aware of the action or to direct other users' attention. This is an interesting finding as it provides a counter-example to our conceptualization of access management as an implicit effect of visual attention. Instead, *users can deliberately utilize access control mechanisms in order to control awareness.*

- When implementing visual attention-based access designers and researchers should be sensitive to user adaptations that can emerge once detailed measures of awareness becomes part of the interaction. Although not observed in this study, of interest is how actionable use of awareness information in groupware would affect users' privacy concerns, their methods for creating accountability and "plausible deniability" (the ability to perceive information without being held accountable [25, 43]).

6.1.4 Both Action Type and Interaction History Were Relevant for Visual Attention-based Access. The study featured fine-grained access controls for data-gathering purposes, but this level of fine grained control can be too demanding considering that most online collaboration tools make a basic edit and view access distinction. However, we observed that access type assignments were to a certain degree influenced by the action type and interaction history of the item, such as the assignment of consensual access once the item content is agreed by both users.

- A possible design implication of this is potential automation of access types based on interaction history. The system can keep track of the visual attention situations in which an item is created or edited and adapt the access type accordingly. For example, an item that is created under joint attention can automatically be assigned universal view access, or an item that is edited under joint attention can require the visual attention of the same users for further modifications. Similarly, the access rights for viewing an item can be extended to other users who had already viewed the item under the owner's supervision. Furthermore, access types can be clustered based on the user preference data to simplify user choice (e.g., if editing an object is set to private, then the system can automatically assign the same access to deletion).

6.1.5 Strict Consensus Is Likely to Be Challenging in Larger Groups.

- Another take-away is that although visual attention-based access can be more convenient when compared to manual confirmations, strictly requiring the joint attention of multiple users can be arduous. One possible solution is to configure consensual action to work with lower thresholds of joint attention such as the majority of users or with a minimum number of supervising users, such as when an action requires the attention of at least this/these person(s). The threshold can be configured dynamically based on the action type, since actions like deleting can be more critical than editing or moving. Another possible solution that is not explored in the current study is to employ alternative input handling techniques such as notifying other users instead of denying access.

6.2 Limitations and Directions for Future Work

Although our study was limited to two user interactions, we envision the use of visual attention-based access for larger groups and other physical settings. Previous studies suggest that increased number of individuals and multiple groups working in parallel can result in even more fragmented visual attention [31, 57]. Horizontal and vertical displays surfaces can come in various sizes and configurations that facilitate different levels of collaboration and visual monitoring [5]. The current study setting (single-wall display in a meeting room) ensured that the two participants were always in the proximity of each other, and could peripherally monitor the shared space and reorient their attention with relative ease. These conditions might not be valid for larger groups working on distributed or larger shared spaces. Even though the identified user motivations can help inform design work for other settings, the generalizability of user preferences observed in the current study remains an issue for future work, and partly depends on how flexibly joint attention is implemented for larger groups.

We also implemented visual attention-based access for collocated and synchronous interactions. Thus, our exploration of visual attention-based access only covers a single quadrant of the traditional groupware time/space matrix. However, visual attention situations and the access types framed in the paper equally apply to remote collaborations. For example, edit and view rights are often used in remote collaboration tools [1, 2], but we know of no tool that allows extending editing rights to other users based on whether the owner(s) is paying attention. Implementing similar mechanisms, however, would require taking additional factors into account. First, unlike collocated collaboration where the physical constraints of the space naturally limit the amount of users, many people can act on the same workspace area during remote collaboration. Keeping track of massively concurrent input that can be the case for remote collaboration is a challenge even when the user looks precisely at the edited workspace area. Second, as users' arm or body movements are not observable, remote collaboration gives fewer visual cues about users' actions, although past research has shown that this can be remedied to a certain extent [6, 22]. Third, remote groupware would require other means for sensing visual attention, which can range from crude measures of user presence to fine-grained eye tracking data. However, less straightforward is the extension to asynchronous collaboration, as the access types we described are based on concurrent visual attention situations.

In fact, the concepts of groupware and collaboration alone can imply sustained interactions within a defined set of users with pre-established roles, an assumption that might not hold valid in other settings such as spontaneous, brief and anonymous interactions with urban displays in public space (e.g., [19, 31, 42, 44]). These settings often involve larger displays, many transitory users and ephemeral content that might not be as personal or persistent as in groupware. At the same time, previous studies acknowledge the role of public visibility of actions in preventing inappropriate content from anonymous users [19] and observe that users can still feel ownership of the content they create for public displays [44]. Yet we are not aware of any prototype that flexibly handles user input based on the visual attention or the presence of others in the public space. Applications in this domain would require alternative technical solutions for identifying users and sensing visual attention as instrumented solutions (e.g., markers, eye trackers) can be impractical.

We also limited our scope to awareness as facilitated by visual perception. This was informed by the fact that the interactions with our applications relied on visual monitoring to be perceived; manual (touch and pen) input is visually observable through arm movements and the output provided by the system is visual. However, other interfaces can enable interaction using different modalities such as speech input or audio feedback. Input and output in other modalities would require different measures of awareness. For example, the system can utilize proximity information to infer whether a user hears another user's speech input.

Finally, as with every exploratory study, there are limitations to what we can claim. The user study we conducted was oriented towards gathering data about user preferences and motivations for visual attention-based access, rather than quantifying its performance. Further work will be required to assess visual attention-based access in comparison to traditional access models in realistic use settings. Nonetheless, our work points to a number of user motivations and design considerations for future work to build on.

7 CONCLUSION

We proposed the concept of visual attention-based access as a form of a contextual access control model that manages access based on who visually attends to an action or an interface element. Based on insights from previous work, we highlighted that different approaches to awareness, namely relying on self-organization and access control, place different demands on users' visual attention. Visual attention-based access provides a contextual switch between these two approaches by relying on self-organization when users visually attend to an action and restricting access when they don't, in a way working around the trade-off between individual power and awareness [17, 22].

The user study provided us with data about access preferences for different types of actions and uncovered different motivations for utilizing different access types. We observed that in addition to preventing conflicts, visual attention-based access can be employed to ease keeping track of the workspace and direct others' attention. Visual attention-based access has been perceived as convenient but also uncertain, a finding that calls for various design measures to be explored in future work. Overall, our work contributes to a more detailed understanding of utilizing visual attention and awareness data as a real-time input in groupware.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their feedback and Hendrik Schneider for his help with setting up the interactive screens. This work has been partly supported by the Academy of Finland project EMOEMP (Grant Agreement #305577).

REFERENCES

[1] 2018. Google docs. https://www.google.com/intl/en/docs/about/. Accessed: 2018-02-10.

[2] 2018. Office 365. https://support.office.com/en-us/article/collaborate-in-word-2016-b3d7f2af-c6e9-46e7-96a7-dabda4423dd7?ui=en-US&rs=en-US&ad=US. Accessed: 2018-02-10.

[3] Maneesh Agrawala, Andrew C. Beers, Ian McDowall, Bernd Fröhlich, Mark Bolas, and Pat Hanrahan. 1997. The Two-user Responsive Workbench: Support for Collaboration Through Individual Views of a Shared Space. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 327–332. https://doi.org/10.1145/258734.258875

[4] Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2011. Medusa: A Proximity-aware Multi-touch Tabletop. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 337–346. https://doi.org/10.1145/2047196.2047240

[5] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. 2015. Interaction with Large Displays: A Survey. *ACM Comput. Surv.* 47, 3, Article 46 (Feb. 2015), 38 pages. https://doi.org/10.1145/2682623

[6] Ignacio Avellino, Cédric Fleury, Wendy E. Mackay, and Michel Beaudouin-Lafon. 2017. CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6718–6729. https://doi.org/10.1145/3025453.3025604

[7] Michel Beaudouin-Lafon and Alain Karsenty. 1992. Transparency and Awareness in a Real-time Groupware System. In *Proceedings of the 5th Annual ACM Symposium on User Interface Software and Technology (UIST '92)*. ACM, New York, NY, USA, 171–180. https://doi.org/10.1145/142621.142646

[8] Jeremy P. Birnholtz, Tovi Grossman, Clarissa Mak, and Ravin Balakrishnan. 2007. An Exploratory Study of Input Configuration and Group Process in a Negotiation Task Using a Large Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 91–100. https://doi.org/10.1145/1240624.1240638

[9] Lauren J. Bricker, Marla J. Baker, and Steven L. Tanimoto. 1997. Support for Cooperatively Controlled Objects in Multimedia Applications. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems (CHI EA '97)*. ACM, New York, NY, USA, 313–314. https://doi.org/10.1145/1120212.1120410

[10] Xiang Cao, Clifton Forlines, and Ravin Balakrishnan. 2007. Multi-user Interaction Using Handheld Projectors. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 43–52. https:

//doi.org/10.1145/1294211.1294220

[11] Jean Carletta, Robin L. Hill, Craig Nicol, Tim Taylor, Jan Peter de Ruiter, and Ellen Gurman Bard. 2010. Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods* 42, 1 (2010), 254–265. https://doi.org/10.3758/BRM.42.1.254

[12] Michael J. Covington, Wende Long, Srividhya Srinivasan, Anind K. Dev, Mustaque Ahamad, and Gregory D. Abowd. 2001. Securing Context-aware Applications Using Environment Roles. In *Proceedings of the Sixth ACM Symposium on Access Control Models and Technologies (SACMAT '01)*. ACM, New York, NY, USA, 10–20. https://doi.org/10.1145/373256.373258

[13] 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. https://doi.org/10.1145/3025453.3025573

[14] 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. https://doi.org/10.1145/2858036.2858499

[15] Paul Dietz and Darren Leigh. 2001. DiamondTouch: A Multi-user Touch Technology. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST '01)*. ACM, New York, NY, USA, 219–226. https://doi.org/10.1145/502348.502389

[16] Jakub Dostal, Uta Hinrichs, Per Ola Kristensson, and Aaron Quigley. 2014. SpiderEyes: Designing Attention- and Proximity-aware Collaborative Interfaces for Wall-sized Displays. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. ACM, New York, NY, USA, 143–152. https://doi.org/10.1145/2557500.2557541

[17] Paul Dourish and Victoria Bellotti. 1992. Awareness and Coordination in Shared Workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work (CSCW '92)*. ACM, New York, NY, USA, 107–114. https://doi.org/10.1145/143457.143468

[18] Morten Esbensen, Paolo Tell, Jacob B. Cholewa, Mathias K. Pedersen, and Jakob Bardram. 2015. The dBoard: A Digital Scrum Board for Distributed Software Development. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 161–170. https://doi.org/10.1145/2817721.2817746

[19] Patrick Tobias Fischer and Eva Hornecker. 2012. Urban HCI: Spatial Aspects in the Design of Shared Encounters for Media Facades. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 307–316. https://doi.org/10.1145/2207676.2207719

[20] Yaser Ghanam, Xin Wang, and Frank Maurer. 2008. Utilizing Digital Tabletops in Collocated Agile Planning Meetings. In *Agile 2008 Conference*. 51–62. https://doi.org/10.1109/Agile.2008.13

[21] Carl Gutwin, Scott Bateman, Gaurav Arora, and Ashley Coveney. 2017. Looking Away and Catching Up: Dealing with Brief Attentional Disconnection in Synchronous Groupware. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2221–2235. https://doi.org/10.1145/2998181.2998226

[22] Carl Gutwin and Saul Greenberg. 1998. Design for Individuals, Design for Groups: Tradeoffs Between Power and Workspace Awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/289444.289495

[23] Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Comput. Supported Coop. Work* 11, 3 (Nov. 2002), 411–446. https://doi.org/10.1023/A:1021271517844

[24] Carl Gutwin, Gwen Stark, and Saul Greenberg. 1995. Support for Workspace Awareness in Educational Groupware. In *The First International Conference on Computer Support for Collaborative Learning (CSCL '95)*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 147–156. https://doi.org/10.3115/222020.222126

[25] Jeff Hancock, Jeremy Birnholtz, Natalya Bazarova, Jamie Guillory, Josh Perlin, and Barrett Amos. 2009. Butler Lies: Awareness, Deception and Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 517–526. https://doi.org/10.1145/1518701.1518782

[26] Christian Holz and Patrick Baudisch. 2013. Fiberio: A Touchscreen That Senses Fingerprints. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2501988.2502021

[27] Eva Hornecker, Paul Marshall, Nick Sheep Dalton, and Yvonne Rogers. 2008. Collaboration and Interference: Awareness with Mice or Touch Input. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 167–176. https://doi.org/10.1145/1460563.1460589

[28] Scott E. Hudson and Ian Smith. 1996. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, New York, NY, USA, 248–257. https://doi.org/10.1145/240080.240295

[29] P. Isenberg, D. Fisher, S. A. Paul, M. R. Morris, K. Inkpen, and M. Czerwinski. 2012. Co-Located Collaborative Visual Analytics around a Tabletop Display. *IEEE Transactions on Visualization and Computer Graphics* 18, 5 (May 2012), 689–702. https://doi.org/10.1109/TVCG.2011.287

[30] Shahram Izadi, Harry Brignull, Tom Rodden, Yvonne Rogers, and Mia Underwood. 2003. Dynamo: A Public Interactive Surface Supporting the Cooperative Sharing and Exchange of Media. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and*

Technology (UIST '03). ACM, New York, NY, USA, 159–168. https://doi.org/10.1145/964696.964714

[31] Giulio Jacucci, Ann Morrison, Gabriela T. Richard, Jari Kleimola, Peter Peltonen, Lorenza Parisi, and Toni Laitinen. 2010. Worlds of Information: Designing for Engagement at a Public Multi-touch Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 2267–2276. https://doi.org/10.1145/1753326.1753669

[32] Mikkel R. Jakobsen and Kasper Hornbæk. 2014. Up Close and Personal: Collaborative Work on a High-resolution Multitouch Wall Display. *ACM Trans. Comput.-Hum. Interact.* 21, 2, Article 11 (Feb. 2014), 34 pages. https://doi.org/10.1145/2576099

[33] Ulrike Kister, Patrick Reipschläger, Fabrice Matulic, and Raimund Dachselt. 2015. BodyLenses: Embodied Magic Lenses and Personal Territories for Wall Displays. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 117–126. https://doi.org/10.1145/2817721.2817726

[34] Devdatta Kulkarni and Anand Tripathi. 2008. Context-aware Role-based Access Control in Pervasive Computing Systems. In *Proceedings of the 13th ACM Symposium on Access Control Models and Technologies (SACMAT '08)*. ACM, New York, NY, USA, 113–122. https://doi.org/10.1145/1377836.1377854

[35] Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. 2015. GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 395–404. https://doi.org/10.1145/2807442.2807479

[36] Roman Lissermann, Jochen Huber, Martin Schmitz, Jürgen Steimle, and Max Mühlhäuser. 2014. Permulin: Mixed-focus Collaboration on Multi-view Tabletops. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3191–3200. https://doi.org/10.1145/2556288.2557405

[37] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. 2017. CoReach: Cooperative Gestures for Data Manipulation on Wall-sized Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6730–6741. https://doi.org/10.1145/3025453.3025594

[38] Mitsunori Matsushita, Makoto Iida, Takeshi Ohguro, Yoshinari Shirai, Yasuaki Kakehi, and Takeshi Naemura. 2004. Lumisight Table: A Face-to-face Collaboration Support System That Optimizes Direction of Projected Information to Each Stakeholder. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 274–283. https://doi.org/10.1145/1031607.1031653

[39] Sven Mayer, Lars Lischke, Jens Emil Grønbæk, Zhanna Sarsenbayeva, Jonas Vogelsang, Pawel W. Wozniak, Niels Henze, and Giulio Jacucci. 2018. Pac-Many: Movement Behavior when Playing Collaborative and Competitive Games on Large Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 539, 10 pages. https://doi.org/10.1145/3173574.3174113

[40] Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. 2006. Cooperative Gestures: Multi-user Gestural Interactions for Co-located Groupware. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1201–1210. https://doi.org/10.1145/1124772.1124952

[41] Meredith Ringel Morris, Kathy Ryall, Chia Shen, Clifton Forlines, and Frederic Vernier. 2004. Beyond "Social Protocols": Multi-user Coordination Policies for Co-located Groupware. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 262–265. https://doi.org/10.1145/1031607.1031648

[42] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 1285–1294. https://doi.org/10.1145/1873951.1874203

[43] Bonnie A. Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, New York, NY, USA, 79–88. https://doi.org/10.1145/358916.358975

[44] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. 2008. It's Mine, Don'T Touch!: Interactions at a Large Multi-touch Display in a City Centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1285–1294. https://doi.org/10.1145/1357054.1357255

[45] Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. GazeArchers: Playing with Individual and Shared Attention in a Two-player Look&Shoot Tabletop Game. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 213–216. https://doi.org/10.1145/3012709.3012717

[46] Dominik Schmidt, Julian Seifert, Enrico Rukzio, and Hans Gellersen. 2012. A Cross-device Interaction Style for Mobiles and Surfaces. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 318–327. https://doi.org/10.1145/2317956.2318005

[47] Kjeld Schmidt. 2002. The Problem with 'Awareness': Introductory Remarks on 'Awareness in CSCW'. *Computer Supported Cooperative Work (CSCW): The Journal of Collaborative Computing* 11, 3 (2002), 285–298.

[48] Stacey D. Scott, M. Sheelagh T. Carpendale, and Kori M. Inkpen. 2004. Territoriality in Collaborative Tabletop Workspaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 294–303. https://doi.org/10.1145/1031607.1031655

[49] Baris Serim and Giulio Jacucci. 2016. Pointing While Looking Elsewhere: Designing for Varying Degrees of Visual Guidance During Manual Input. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5789–5800. https://doi.org/10.1145/2858036.2858480

[50] Baris Serim, Khalil Klouche, and Giulio Jacucci. 2017. Gaze-Adaptive Above and On-Surface Interaction. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 115–127. https://doi.org/10.1145/3064663.3064744

[51] Garth B. D. Shoemaker and Kori M. Inkpen. 2001. Single Display Privacyware: Augmenting Public Displays with Private Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 522–529. https://doi.org/10.1145/365024.365349

[52] Desney S. Tan and Mary Czerwinski. 2003. Information Voyeurism: Social Impact of Physically Large Displays on Information Privacy. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 748–749. https://doi.org/10.1145/765891.765967

[53] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. 2006. Collaborative Coupling over Tabletop Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1181–1190. https://doi.org/10.1145/1124772.1124950

[54] William Tolone, Gail-Joon Ahn, Tanusree Pai, and Seng-Phil Hong. 2005. Access Control in Collaborative Systems. *ACM Comput. Surv.* 37, 1 (March 2005), 29–41. https://doi.org/10.1145/1057977.1057979

[55] Edward Tse, Jonathan Histon, Stacey D. Scott, and Saul Greenberg. 2004. Avoiding Interference: How People Use Spatial Separation and Partitioning in SDG Workspaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 252–261. https://doi.org/10.1145/1031607.1031647

[56] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 137–146. https://doi.org/10.1145/1029632.1029656

[57] Lauren Westendorf, Orit Shaer, Petra Varsanyi, Hidde van der Meulen, and Andrew L. Kun. 2017. Understanding Collaborative Decision Making Around a Large-Scale Interactive Tabletop. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 110 (Dec. 2017), 21 pages. https://doi.org/10.1145/3134745

[58] Mike Wu and Ravin Balakrishnan. 2003. Multi-finger and Whole Hand Gestural Interaction Techniques for Multi-user Tabletop Displays. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, NY, USA, 193–202. https://doi.org/10.1145/964696.964718

[59] 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 (2017), 173–186. https://doi.org/10.1007/s00779-016-0969-x

Received February 2018; revised May 2018; accepted September 2018

Publication IV

Barış Serim, Giulio Jacucci. Explicating “Implicit Interaction”: An Examination of the Concept and Challenges for Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Paper No. 417, May 2019.

Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research

Barış Serim
Helsinki Institute for Information Technology HIIT,
Department of Computer Science, University of Helsinki
Department of Design, Aalto University, Finland
baris.serim@helsinki.fi

Giulio Jacucci
Helsinki Institute for Information Technology HIIT,
Department of Computer Science, University of Helsinki
giulio.jacucci@helsinki.fi

ABSTRACT

The term implicit interaction is often used to denote interactions that differ from traditional purposeful and attention demanding ways of interacting with computers. However, there is a lack of agreement about the term's precise meaning. This paper develops implicit interaction further as an analytic concept and identifies the methodological challenges related to HCI's particular design orientation. We first review meanings of implicit as unintentional, attentional background, unawareness, unconsciousness and implicature, and compare them in regards to the entity they qualify, the design motivation they emphasize and their constructive validity for what makes good interaction. We then demonstrate how the methodological challenges can be addressed with greater precision by using an updated, intentionality-based definition that specifies an *input–effect relationship* as the entity of implicit. We conclude by identifying a number of new considerations for design and evaluation, and by reflecting on the concepts of user and system agency in HCI.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; *User models.*

KEYWORDS

Implicit interaction; explicit interaction; framework; intentionality

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300647

ACM Reference Format:
Barış Serim and Giulio Jacucci. 2019. Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3290605.3300647

1 INTRODUCTION

The last two decades witnessed an increasing interest in supporting interactions that differ from traditional purposeful and attention demanding ways of engaging with computers. A term that is often used to denote these new types of user engagements is "implicit interaction". Implicit interactions are now being pursued in domains as diverse as ubiquitous interaction [83], information seeking [18, 54, 99], attentive interfaces [95, 106] and physiological computing [28]. A public display that shows content when it senses human presence or a recommendation engine that utilizes user actions for social recommendations are typical examples. As computers are getting increasingly capable of sensing the environment and making inferences about the situation, we can expect similar interactions to proliferate and partly replace what is called "explicit interactions".

At the same time, there is a lack of consensus regarding the precise meaning of the explicit–implicit distinction. Over time, the distinction has come to serve as a placeholder for different meanings such as intentionality, attentional focus and awareness. The terms are generally used as a quality of interaction (itself a concept with diverse interpretations [44]), but it is common to apply them to other entities such as the interface, the input or the sensing capability of the system. When the terms do qualify interaction, they are often described through built-in properties of the interface rather than the interaction as it unfolds. Adding to the confusion are other distinctions such as foreground–background [15], overt–covert, command–non-command [64] and active–passive [25] that have overlapping meanings and at times used interchangeably with explicit–implicit. The broad use of the terms led to the introduction of other, more strictly formulated distinctions such as intentional–incidental [21] and reactive–proactive [51].

The ambiguity is problematic, firstly because the term implicit interaction has become part of the HCI lexicon. Even though an ambiguous understanding of the term can function as a convenient shorthand, its effective use requires researchers to clarify what they mean by implicit interaction, validate whether an interface actually facilitates implicit interactions, and demonstrate that implicitness informs design or evaluation in a way other concepts do not. Otherwise, the term risks either being redundant or obscuring more than it explains. Perhaps more consequentially, however, we find the ambiguity exposed by the various uses of the term representative of broader methodological challenges concerning a very central phenomenon in HCI, namely the relationship between a user's mental attitude and what is considered appropriate system behavior.

Thus, we see value in developing "implicit interaction" further as an analytic concept. In the following sections, we first review and compare different meanings that the term implicit stands for in HCI, psychology and pragmatics, and identify the methodological challenges concerning HCI's particular design orientation. We then revisit these methodological challenges with an updated, intentionality-based definition of implicit interaction as *user's attitude towards an input–effect relationship* in which *the appropriateness of a system response to the user input (i.e., an effect) does not rely on the user having conducted the input to intentionally achieve it*, and show how this updated definition can be deployed to address various methodological challenges. We conclude by identifying a number of new considerations for the design and evaluation of interfaces that aim to facilitate implicit interactions, and by reflecting on the concepts of user and system agency in HCI.

2 THE MANY MEANINGS OF IMPLICIT

The HCI use of the term implicit can be traced back to as early as 1983, when Revesman and Greenstein noted that *"if the computer were able to predict accurately the actions of the human, information would be communicated with no overt communication on the part of the human; an implicit line of communication would exist"* [78, p. 107]. The term has since appeared in different domains of HCI such as mobile computing [37, 43] search interfaces [18] and ubiquitous computing [63, 84]. Various formulations of implicit interaction soon followed. Schmidt defines implicit interaction as *"an action performed by the user that is not aimed to interact with a computerised system but which such a system understands as input"* [83]. This is somehow echoed in Ju and Leifer's definition of implicit interaction as *"those that occur without the explicit behest or awareness of the user"* [52]. While these formulations provide good departure points, they also lead to many further questions. Does implicit stands for purposefulness, awareness or any other meaning? Is it a quality of "an action performed by the user" or something that occurs as a result of an action? A meta-analysis that compares different meanings and implications of "implicit" has so far been missing.

Equally problematic is the term "interaction". Perspectives on interaction, such as Norman's dialogue model [65], control theory [100] and activity theory [10], take user goals as a departure point for their analyses, which poses a challenge for the notion of interaction as something unintentional [21, 44]. Similarly, when defined through a complete lack of attentional focus or awareness, implicit interaction has been called *"not proper interaction in the sense that it engages us in addition... to what we otherwise are doing"* [49, p. 122] and *"hardly... an interaction at all, since there is no activity or awareness on our part"* [52, p. 77]. The problem partly arises from the diverse meanings of implicit but also from the ambiguity of the word interaction. In HCI, interaction has been conceptualized in terms of mechanical consequence by antecedent, but also as user experience, tool-use or control [44], all of which make user attention or intention a prerequisite for interaction. Additionally, HCI uses are not limited to interaction and one can encounter other phrases such as "implicit sensing" [38, 81], "implicit input" [53, 75, 93], "implicit interface" [87, 93] or a specific outcome of user input, as in "implicit authentication" [13]. The diversity of uses raises the question of whether (or what entity within) interaction provides a useful unit of analysis.

This section discusses various meanings and entities that the term implicit interaction can stand for. There are two reasons for our focus on the word implicit instead of a predefined meaning. First, we examine how the use of implicit as a placeholder for multiple meanings might not be accidental but grounded in the context of particular cases in which they overlap. Secondly, by comparing and contrasting various meanings, we sharpen our own definition of the term. Overall, our examination is oriented towards identifying the consequences of different meanings for research rather than determining their prevalence in the literature. We do not claim comprehensiveness of all use instances, but refer to HCI examples for illustration.

We started the examination by identifying a number of often cited, key publications that discuss the term implicit interaction in depth [21, 42, 51, 52, 83]. Even though not cited by key HCI publications, explicit–implicit distinction also features in other fields, notably psychology and pragmatics, and work in HCI occasionally reflect similar meanings. Thus, we found the reviews in these fields [16, 20, 29, 85] useful for comparison. As we have proceeded, we also expanded the scope of analysis to other terms that have overlapping meanings and are used interchangeably with explicit–implicit. For example, Buxton's [15] distinction of "foreground" *"which are in the fore of human consciousness – intentional"* and

Table 1: An overview of different meanings, their description, the entity they qualify and potential motivations for design.

Implicit as	Description	Implicit qualifies	Example motivation for design
Unintentional	The degree a particular effect is intended by the user	Input–effect relationship	Providing appropriate responses that go beyond what the user has intended
Attentional background	Attention reserved during the execution of an input or evaluation of a system response	Input or output	Freeing up attentional resources through external backgrounding
Unawareness	The degree of awareness of a particular effect caused by the user's input	Input–effect relationship	Providing appropriate responses that go beyond what the user is aware of
Unconscious	User's awareness of own mental process	Mental process	Reduced need for mental processing
Implicature	The degree an action represents an agent's intention	Intention–action relationship	Accomplishing an intended effect with less effort

"background" activities *"that takes place in the periphery"* is conceptually similar and has been used together with explicit–implicit distinction in later work [e.g., 42, 51]. "Incidental" interactions, similarly, describe instances when a system utilizes user actions that have been *"performed for some other purpose"* than the one executed by the system [21]. These formulations emphasize the purposefulness of user actions in achieving a specific outcome, a meaning that we will henceforth refer to as *intentionality.* However, various definitions point to other qualities, namely *attentional focus*, *awareness*, *consciousness* (psychology) and *implicature* (pragmatics).

Implicit as Unintentional

The formulations of explicit–implicit [51, 83] as well as foreground–background [15, 41] define them in terms of user's intentionality. Editing a document by typing on a keyboard is an explicit interaction insofar as the outcome, and the purpose, of the action is editing the document. In contrast, a smart room that activates the lights when a person walks in facilitates implicit interaction (unless the person walked into the room specifically to turn on the lights).

A defining feature of implicit interactions is their occurrence as a result of the user input. Perhaps due to this feature, implicit has occasionally been defined as a quality of user's "action" (as in Schmidt's definition [83]) or "input" [e.g., 1, 53, 75, 93]. Yet we argue that the suitable entity for intentionality is a specific outcome that results from user's input, namely an *input–effect relationship.* This is first of all due to our focus on human–computer interaction where intentionality is categorically about future states in the environment. Of course, the user input itself can be considered unintentional in certain cases (as in involuntary muscle movements), but this qualifies a mind–behavior relationship instead of a user–environment relationship. Apart from categorical precision, defining implicitness as an input–effect relationship allows for including situations in which the system executes both intended and unintended effects as a result of the user input. For example, a search system can harness a user's browsing or bookmarking actions that are intended for examining web pages as implicit relevance feedbacks to infer user interests and improve future search results [2, 76, 99].

Most accounts of HCI view user intentions central to what makes a good interaction, in other words, ppstulating the norm that an interaction is judged by [44]. Thus, the definition of implicit interactions as unintentional has been scrutinized. From a control theory perspective, Williamson defines background interactions as ideally dependent on *"known intentions and the inferred intentions which they act as proxy for"* [100, p. 23]. We argue that the objection is due to the multiple meanings of interaction, which can refer to a granular input–effect relationship but also to an activity conducted through an interactive system. In many examples described above, even if a specific effect of the user input is unintentional, implicit interaction can be motivated by a purposeful activity that encompasses the specific input–effect relationship. For example, implicit feedback in search systems can be motivated through a user's goal of seeking particular information. In contrast, when there are no commitments to inferred intentions, the interaction can be motivated by the assumed benefit to the user [21]. In other words, beneficial or any other quality can replace intentional as the norm of interaction.

We so far treated intentionality as self-evident. Yet it is worth noting that the HCI understandings of the concept range from a general "directedness of meaning" [24] to "the decision to act so as to achieve the goal" as in Norman's framework [65]. The difference in HCI understandings can be traced back to the philosophical conceptions of intentionality in mentalistic (as in Husserl and Searle) and non-mentalistic (as in Heidegger and Merleau-Ponty) terms (see Dreyfus

[26, 27] for a treatment of the subject). As a mentalistic concept, intentions correspond to well-formulated goals that are *held* by a knowing subject. In its non-mentalistic conception, intentions are instead *embodied* in our everyday habitual performance toward practical ends and characterize things that we do without self-reflection. This partly amounts to the expansion of intentionality from the mind and deliberate reasoning to the body, habits and *unconsciousness*—another meaning that we will unpack later in this section. However, unconsciousness tells only part of the story. An insight from phenomenology as well as from the situated action perspective [90] is that successful accomplishment of activities relies on background assumptions, namely external dependencies that are taken for granted until a breakdown. Implicit interactions in HCI approximate to such external dependencies when they assume embodied knowledge on users' side about what to take for granted.

Implicit as Attentional Background

Implicit [49, 51] or background [15, 41] interactions have also been defined in terms of attentional focus during interaction. HCI research is often informed by the limitations to the human processing capacity [e.g., 6, 46, 69], which, as a consequence, requires selectivity regarding what is being attended to. Researchers directed their design efforts to make interactive systems less attention demanding, for instance, by devising ambient, peripheral and low-bandwidth output [48, 74]. The same design strategy can be extended to the design of input techniques that require minimal user attention. Manual selection of small targets on a screen is explicit, while automatic activation of lights upon entering a room is implicit. Mobile interfaces can switch between landscape and portrait modes by sensing the device orientation instead of through attention-demanding GUI control [42, 84]. Systems can be designed to be operated without fine motor control to accommodate "casual" user inputs [73].

Buxton defined background interactions as both unintentional and in the attentional background [15], but the two qualities do not necessarily overlap. Ju et al. build their implicit interaction framework on Buxton's foreground–background distinction but supplement it with an additional reactive–proactive distinction to separate intentionality from attentional focus [51]. An interaction can be both intentional and in the attentional background when the system automates various sub-tasks triggered by a user input without providing attention-demanding feedback. Conversely, an interaction can be at the attentional foreground but executed without the user's initiative, such as when the system provides attention-demanding feedback for an unintended effect.

When defined through attentional focus, implicit interactions are motivated by freeing up users' cognitive resources for other activities. Level of attention also allows explicit–implicit or foreground–background distinctions to be defined across a continuum [15, 41] in which attention serves as a quantitative measure that can be operationalized through information throughput [15, 73].

Implicit as Unawareness

We use the term awareness to denote users' knowledge about how their input is utilized by the system. Traditional interface design has put heavy emphasis on the predictability of action outcomes through affordances and feedback/feedforward mechanisms that communicate system responses back to users [65]. In contrast, interfaces that target implicit interaction might avoid such features. Implicitness has occasionally been defined in terms of awareness [52, 56, 71]. Typing into a computer is explicit as long as users are aware of the action outcomes, while the utilization of users' gaze [71] or physiological signals [56] without their awareness is implicit. As with attention, awareness can be defined as a continuum. Users can be unaware that they are providing input to a system, they can have ambiguous awareness about how their input is utilized by the system or they can be fully informed [97]. For more granular analysis, users can be considered to have varying degrees of awareness of different effects that result from the same input.

Purposefulness of interaction requires a certain degree of awareness of action outcomes, making awareness a precondition for intentionality. Thus, the motivations that are described under unintentional also apply to unawareness. On the other hand, awareness does not always entail intentionality. Dix distinguishes between awareness and intentionality through the concept of "expected" interactions in which the user is aware of the effect of his or her action but has not performed the action with the intention to cause the particular effect [21]. Paraphrasing Dix's example, a person can expect the lights to switch on when entering a room, even though this is not the intention for entering. Dix's other category "incidental" refers to interactions where the outcome is neither intentional nor expected.

Implicit as Unconscious

Implicit–explicit distinction can also be defined through consciousness. This meaning is prevalent in psychology and related fields.[1] Implicit learning is a process by which knowledge is acquired "independently of conscious attempts to do so" [77, p. 219]. Implicit memory is the facilitation of

[1]The precise definition of implicit and how it relates to other concepts such as voluntariness, verbalization or intentionality has been extensively discussed in different domains of psychology [20, 29, 85]. Here, instead of an in-depth discussion of different psychological concepts, we will limit our scope to what we perceive to be a salient concept, consciousness, and outline how it differs from the HCI meanings discussed earlier.

performance "without conscious recollection" [82, p. 501]. The meaning of the term is similar in social psychology [29, 33, 102] where implicit social cognition is "unconscious" in the sense that it is "unavailable to self-report or introspection" [33, p. 5]. Researchers employed experimental techniques such as Implicit Association Test (IAT) [34] to measure "implicit attitudes" towards age, gender and other socially relevant attributes [67, 79]. The definition of the term consciousness is contested, but generally corresponds to "online phenomenological awareness" (i.e., awareness of one's own mental processes) [85, p. 138]. Unconscious processes are described as faster, more efficient but also harder to modify and verbalize than conscious processes [85].

Some work in HCI employs this meaning of the term but without targeting implicit interactions as a design goal. Poeller et al. study "implicit motives" ([61]) to predict video gamers' behavior and play experience [72]. Denning et al. utilize the "implicit memory" of users to aid password recovery [19]. Additionally, we observe that motivations for facilitating implicit interactions occasionally evoke the unconscious meaning of the term. Consider Schmidt's reference to nonverbal cues during face-to-face communication: *"In many cases the robustness of human-to-human communication is based on the implicitly introduced contextual information, such as gestures, body language, and voice"* [83, p. 91]. It can be argued that gestures are part of a person's purposeful social act, but their realization is more or less automatic or unreflective; they are intentional but not deliberate. The term "internally backgrounded" (as opposed to both "foregrounded" and "externally backgrounded" [62]) denotes a similar distinction. Execution of tasks without conscious processing has been related to skilled, well-practiced behavior [55, 62, 66].

Implicit as unconscious qualifies the mental process that leads to an action rather than the relationship between the user action and its effect. To illustrate the point, whether the user action was unconscious (e.g., walking into a room without thinking about it) or its outcome was unintended (e.g., turning on the lights when walking into the room) point to different aspects of an action. Implicit as such also hints at an internal division ("internally backgrounded" [62]) rather than to the division of labor between the user and the system ("externally backgrounded" [62]). For this reason, Ju et al. argued for its exclusion from the scope of implicit interaction by advocating a distinction between *"situations where users don't have to think and plan because the users have developed tacit knowledge of how to operate a task, and situations where the users don't have to think and plan because the system is acting proactively on their behalf."* [51, p. 20].

Implicit as Implicature

Finally, implicit can be understood as akin to "implicatures" [36] (or analogous "indirect speech acts" [86]) in pragmatics, the field that studies the contextual aspects of human communication. Definitions of implicit in pragmatics [e.g., 16, 57, 101] often trace back to Grice and his concepts of speaker's meaning and "conversational implicature" [35, 36] where the analytic focus lies on the relationship between what is said (i.e., the explicit meaning of an utterance) and what is meant (implied) by a speaker. Consider approaching to a group of friends before a trip and saying "I am ready". In addition to its explicit meaning of stating one's readiness, the utterance can be an implicit invitation for departure, which is a meaning that is intended (implied) by the speaker. While understood through speaker's intention, this use of implicit better corresponds to how *literal*, as opposed to how intentional, an utterance is. In fact, what underpins conversational implicatures is a clear demonstration by the speaker that his or her utterance is directed at the listener. This allows the listener to infer the speaker's intention by assuming the relevance of the utterance for the given context [101]. Implicit as such differs from the HCI meaning of implicit as unintentional and not directed at a computer [e.g., 15, 51, 83].

Some work in HCI conveys the implicature meaning of the term. Sun et al. distinguish between implicit and explicit behavior strategies a robot assistant can employ when interacting with a user [91]. The robot assistant can direct a user's attention to itself by asking "Hey, listen to me, it's important!" (explicit) or by implying that it has something to communicate by saying "No problem, I will wait for you". In this example, implicitness qualifies the relationship between the system goal (of capturing the user's attention) and the system output. Yet implicatures can also qualify the relationship between users' intentions and their actions. For example, to save from effort, a user can intentionally make an ambiguous query with the expectation that the search interface will successfully retrieve the intended search result.

Within pragmatics, speakers' motivations for employing implicatures have been formulated through various "conversational maxims" [36] or achieving a greater effect on the hearer with less effort [101]. Similar motivations can explain HCI equivalents of implicatures. For example, the user and the system can have a shared understanding that the user would aim to minimize her effort when communicating her intentions to the system. This in turn can allow the system to compensate for seemingly erroneous or ambiguous user inputs while at the same time affording the user to be less precise *on purpose*.

Note that, Gricean view of implicatures conceptualizes communication as a process of intention recognition. As we discussed earlier, the mentalism that accompanies this conception is not without controversy and has been scrutinized from the viewpoint of interactional pragmatics [5, 12, 39, 40]. Interactional pragmatics emphasizes that speakers' prior intentions are inherently vague and negotiable, and can be practically observed only through their uptake by the hearer. A conceptual consequence of this is the reframing of intentions as retrospective accounts that are *attributed* to the speaker rather than prior mental states that predetermine the communication [39]. A methodological consequence is the shift of the analytical focus from the relationship between speakers' utterances and their mental states to the sequential relationship between speakers' utterances (as in most conversation analysis). Parallel discussions played their part in HCI. Work informed by linguistic pragmatics identified intent recognition and categorization as primary goals for system design [4, 103]. Situated action perspective, on the other hand, emphasized the inherent vagueness of intentions for prescribing action [90].

Summary

Early in the section, we noted that intentionality is central to many definitions of implicit, but not always overlapping with other meanings (Table 1). To summarize, intentional actions can be conducted with different levels of attention, lack of awareness rules out intentionality, but awareness does not necessarily entail intentionality. In contrast, implicatures presume intentionality. Consciousness refers to internal self-awareness of mental processes, which differs from the awareness or intentionality of external action outcomes. Note that, all of the meanings deviate from classical explicit interactions that assume a straightforward coupling between users' mental states, their observable behavior and what is appropriate. However, they illustrate different ways in which this can be achieved. They thus relate differently to the ordinary language use of implicit as *"being understood from something else though unexpressed"*[2]. As implicature, implicitness qualifies how literally the observable behavior represents the intention of its agent. As intentionality, implicitness describes whether the appropriateness of an effect is grounded on user intentions (explicit) or additional assumptions (implicit). This reference to expected appropriateness is, in our view, what justifies the term "implicit" instead of describing the interaction simply as "unintentional", or any other word.

Having reviewed different meanings of implicit, we can also more clearly identify the roots of the confusion with implicit "interaction". Simply put, interaction lacks a unified meaning in the context of implicit; researchers can refer to different entities by "interaction" depending on the particular meaning of implicit they employ (Table 1). Consider intentionality; implicit interaction as unintentional can be problematic when interaction stands for a temporal window of user engagement (since users are assumed to be intentional at the activity level) but not when it stands for a granular input–effect relationship. For attentional focus, implicit can qualify both how users attend to their own input and to a particular system output. Consciousness qualifies the mental process that leads to an action rather than the relationship between the action and its outcome. Thus, it is not truly a quality of interaction. For implicatures, the main entity is the relationship between an action and its intended effect. A commonality between different meanings is their reference to the *user's mental attitude (or lack of it) toward a particular entity*. Thus, when dealing with implicitness we are dealing with phenomena that is not directly observable; claims about implicitness relies on inferences from other units of observation.

The lack of direct observation is a general challenge for research and the previous section already gave a glimpse of the methodological differences between psychology and different traditions in pragmatics regarding their approaches toward mentalism. Work in psychology aims to find empirical measures of mental processes for implicitness, which has led to various experimental techniques such as Implicit Association Test [34]. In linguistic pragmatics, implicitness is determined by comparing what is meant by the speaker against the verbal content of an utterance in a given context. Unlike psychology, this often relies on the formal reasoning by the analyst instead of *"getting down to the messy business of experimentation"* [101, p. 280]. Psychology and linguistic pragmatics both aim to reconstruct a first-person account of mental phenomena from outside. Conversation analytic and situated action perspectives, on the other hand, emphasize the inherent vagueness of intentions, leading to their methodological preference for the procedural analysis of action over mental modeling. In doing so, they embrace a third-person perspective of intentionality and study intentions to the extent social agents publicly attribute them to each other.

Various traditions in HCI inherited these different approaches toward mentalism. In contrast to psychology and pragmatics, however, HCI deals with interactions where one of the partners in interaction, the system, is the outcome of design. Although a user's mental state is not directly observable, the system behavior can be specified, enabling HCI to make claims about implicitness through design. Consequently, HCI research on implicit interaction emerged with a strong constructive orientation toward building interfaces that facilitate implicit interaction [51, 83]. This constructive orientation is an additional source of methodological

[2] https://www.merriam-webster.com/dictionary/implicit

challenges such as potential mismatches between the system design and user experience perspectives. The following section will examine these challenges in more detail.

3 METHODOLOGICAL CHALLENGES

This section elaborates on the consequences of different meanings for what we identify as the particular methodological challenges of HCI's constructive orientation, namely 1) determining implicitness by design, 2) establishing design and evaluation criteria and 3) scoping design problems.

Determining Implicitness by Design

A methodological challenge concerning design is understanding the extent users' mental attitudes can be determined through interface properties. Implicitness is occasionally treated as a quality that is predetermined by design, independent from the uncertainties of the actual user interaction. This is most obvious in the phrase "implicit interface", but also apparent in other terms such as "proactiveness" [51], "attentional demand" [51] or "predictability" [49] that respectively define user intentionality, attentional focus and awareness in terms of system properties. As such, implicitness of the actual use experience is treated as unproblematic, but the claim is ideally informed through a number of design decisions that favor unawareness or unintentionality. For example, it can be hard for a user to comprehend causal relationships between inputs and system effects, as in the case of complex personalization algorithms or when there is no salient feedback [56, 89]. Or, the input collected by the system might not have any immediate consequences for a user, such as when it is used for ranking search results [2] or monitoring the audience engagement [38].

At the same time, it might not be possible to know users' intentions or awareness in advance for a given situation, nor there is any guarantee that they will remain static over time. The uncertainties posed by situational factors pose a limitation to determining implicitness through design. Among different accounts of implicit interaction, Dix has acknowledged the "fluidity" between user attitudes for the same method and introduced vocabulary to express their transitions [21]. Interactions can transition from incidental to expected when users *comprehend* the causal relationships between their inputs and the system effects, and from expected to intended when users *co-opt* to intentionally trigger an effect, as in walking inside a motion-sensitive room to intentionally turn on the lights. One can call all of these interactions implicit but only in the loose sense of the word (alternatively meaning unaware, unintentional and implicature).

Dix uses constructed examples to describe incidental, expected and intended interactions, but implicit as a contingent quality suggests making the validation of implicitness part of the evaluation. For example, Kuikkaniemi et al. start with an a-priori distinction between implicit and explicit feedback mechanisms in a first-person shooter game, but interviews during the study show that some participants became aware of the effect [56]. Fisk et al. report the evaluation of an "implicit-only" interface, where two remote users could control a shared workspace only indirectly, by talking to each other [31], but observe that the proposed interaction method resulted in participants modifying the flow of their conversation to control the shared workspace. Verbal accounts and qualitative differences can be harder to observe for low-level interactions, meaning that researchers can expand their methodological toolbox into experimental techniques. For instance, Coyle et al. borrow the experimental procedure of "intentional binding" from cognitive neuroscience to assess the personal sense of agency based on participants' perceived time between input and effect for pointing tasks [17].

Establishing Design and Evaluation Criteria

The constructive orientation of HCI also means that implicit interactions are deliberately targeted by designers. Yet various meanings emphasize different motivations for design, such as providing appropriate responses that go beyond what the user has intended, freeing up attentional resources, decreased need for cognitive processing (consciousness) or accomplishing an intended effect with less effort (implicature). Different motivations are not necessarily mutually exclusive. Unintentional effects can also be at the attentional background. Yet some are obviously so. Implicatures presume intentionality and rule out any unintended benefits. With the exception of implicature, a common theme across different meanings is *their relaxed assumptions of mental representation as a condition for good interaction.* This position diverges from the traditional HCI focus on predictability, direct control and attentional bottlenecks.

Motivations emphasized by different meanings also differ regarding their value as evaluation criteria. For example, freeing up users' attention for other tasks can be a design goal in its own right, but this is less obvious when implicit stands for unintentional or unaware; an interaction is not better simply because an outcome is unintended or unexpected (users' ability to model the system behavior can even benefit the task performance [70]). In this case, implicitness seems more like a by-product rather than a design goal. This relates to previous calls to formulate implicit interactions *"less as a hammer, and more as a lens"* [52, p. 82], that is instead of striving for making interactions implicit, designers should consider their designs as facilitating implicit interactions.

A contentious issue across different meanings is the normative relevance of users' prior intentions (i.e., users' activity-level end goals) for evaluation. One possible position is to evaluate implicit interactions based on how effectively they realize users' inferred intentions [e.g., 30, 100]: Users' prior intentions remain the ultimate metric that the interactive process is evaluated against, but unlike explicit interactions, their successful accomplishment does not presuppose mental representation and intentional execution. One limitation we have noted with this position is that prior intentions can be vague and negotiable, thus, not necessarily good criteria for evaluation. A related methodological challenge is their elicitation as independent ground truths.

An alternative position that we find more widely applicable is to consider interface design as oriented towards providing system responses that are expected to be appropriate *in retrospect*, without any commitments to prior intentions attributed to the user. What is appropriate can be operationally defined in terms of user acceptance, performance gains or any other utility. Even so, different meanings point to different degrees of *constructive validity* regarding how implicitness translates into other utilities. For instance, whether an interface results in better task performance because it frees up users' attention through automation or because the action is done unconsciously through habituation points to different expectations about user skill. In some cases, what is appropriate can be defined in terms of benefits to other parties. It these cases, we find it necessary to establish whether the expected benefit *depends* on users' unawareness such as when the interaction involves a trade-off (assuming that the users will adjust their behavior once they become aware of the trade-off). Additionally, to ensure unawareness or unintentionality, researchers might want to avoid setting clear goals or disclosing their evaluation criteria to their participants. This differs from many controlled studies where experimental tasks act as proxies for user intentions.

Scoping Design Problems

We use the word "scope" to denote the design problem of deciding which interactions should be targeted as implicit. Previous work stressed the foundation of implicit interactions on existing use patterns [52, 83] or actions the user "would have had to perform anyway" for a primary task [41, p. 33]. For example, public displays can facilitate implicit interactions by sensing people's naturally occurring presence and orientation in physical space [7, 96]. Or, consider eye movements that have long been identified as an input for implicit interaction [60, 95, 106]. An argument for their use in many applications is their "naturalness" [106]; people need to monitor the environment to guide their actions. A subsequent design strategy has been employing eye movement data for implicit interactions (e.g., moving the mouse cursor [107] or personalizing search results [14]). Yet a designer can also approach the problem from a different perspective, by acknowledging that eye movements are largely shaped by existing design decisions such as attention-demanding GUIs. This can in turn lead to various alternative design strategies that decrease the need for visual attention. Thus, designers are confronted with a practical choice of whether they treat the existing use patterns as pregiven or modifiable.

Another issue related to the design scope is determining which particular interactions should *not* be targeted as implicit interactions. A design prediction from early research is that, for many applications, implicit interactions will occur alongside with explicit interactions [51, 83]. Previous theorizing provides at least two possible reasons for the continued existence of explicit interactions.

First reason is the aforementioned foundation of implicit interactions on a purposeful or attention-demanding primary activity [21, 41, 83]. In some cases, this activity can be external to a system, thus not part of the interaction. For example, navigation in physical space is often external to an interactive system, but can be utilized by a mobile service for providing contextual data. At the same time, as our interaction with the world is increasingly being mediated by computers, we can expect to encounter many cases where a primary activity is part of the interaction, such as in computer-mediated communication or many productivity tasks. For example, implicit feedback for search systems can rely on actions that users perform anyway to examine or share information [2, 76, 99]. In these cases, explicit interaction becomes a prerequisite for implicit interaction.

The second reason is the potential failure of design assumptions or system inferences. In HCI, it is generally acknowledged that system inference can fail or design decisions can prove inadequate during interaction [8, 9, 45]. The inference mechanisms and design assumptions that implicit interactions rely on are no exception. Consider auto-rotation of the screen content in mobile interfaces based on the device orientation. While helpful in many cases, this adaptation can be inappropriate when the user is lying down [41]. In information retrieval, a user's inferred interest does not always match with what the user finds relevant. Thus, researchers turned to quantify the prediction accuracies for different types of implicit feedback [e.g., 32, 50]. The challenge can to an extent be addressed by developing better models to decrease failures. When near-perfect accuracy is not possible, relying solely on implicit interactions can be impractical, leading to their combination with interaction methods that facilitate explicit interactions.

4 DEPLOYING AN UPDATED DEFINITION

We reviewed various meanings of implicit and identified their consequences in regards to the entity, motivation and the constructive validity of implicitness. In this section, we provide and deploy an updated definition based on the understanding of implicit interaction in terms of intentionality. We define implicit interactions as interactions in which *the* *appropriateness of a system response to the user input (i.e., an effect) does not rely on the user having conducted the input to intentionally achieve it.*

Implicit interactions thus differ from explicit interactions where the appropriateness relies on the assumption that the user has performed an action to intentionally achieve a particular effect or, conversely, has abstained from performing an action to avoid a particular effect. For implicit interactions, appropriateness of a particular effect is instead *understood from the user input, but does not rely on the user's intentionality.* An input refers to any kind of data that originates from the user that is available to the system. An effect refers to any outcome that is facilitated as a result of this user action or data, either with or without system mediation. For example, walking into a room facilitates navigation in space, but can also cause the lights to turn on in the presence of a motion sensor. The formulation of interaction in terms of causal input–effect relationships is not new (see [22, 23]), but has not been previously used to study implicit interactions. Below, we illustrate how specifying an input–effect relationship (or shortly an input–effect pair) as the entity of interaction can help us address the previously identified challenges.

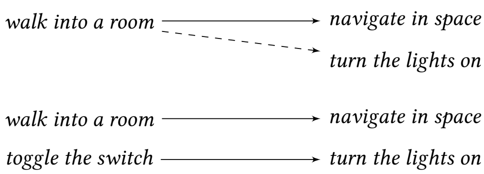


Figure 1: Depiction of implicit (above) and explicit (below) ways of turning on the lights. The dashed line shows an input–effect pair that is expected to be implicit. A common pattern in implicit interactions is the co-occurrence (or bundling) of multiple effects as a result of the same input, eliminating the need for additional actions.

Combining Input–Effect Relationships

The previous section, without going into the details, made the case that implicit interactions will exist alongside explicit interactions. Here, we elaborate on various possible combinations by specifying the entity of implicit as an input–effect pair. Let's start with our observation that implicit interactions often assume the presence of a primary and intentional activity. When viewed through the lens of input–effect pairs, this translates into situations in which an *input leads to multiple effects.* Some of these effects are intended by the user and can explain why the user has conducted the action in the first place. Other effects can be unintentional but still appropriate for a given situation. The expected benefits of implicit interactions can then be attributed to the decreased user effort that is achieved through this bundling (Figure 1) instead of an effect being unintentional.

What is appropriate for a given situation is not always known in advance, a consideration that calls for various additional interactions. First, reaching a particular effect can involve a complementary input–effect pair. For example, Buscher et al. utilize eye movements to model user interests when interacting with a search system [14], but their system retrieves the actual search results only after a typed query to manage the low accuracy of the user model that is inferred from eye movement data (Figure 2a). In another example, Buschek et al. reinforce "explicit" authentication methods (i.e., typing a password on a mobile device) with less accurate "implicit" authentication methods (i.e., using biometric data) as an additional layer of security [13]. In these examples, the ultimate execution of a particular effect (retrieving search results or logging into a mobile device) requires an additional interaction that is expected to be intentional.

Secondly, the system can allow falling back to a corrective input–effect pair when the design assumption or the system inference proves to be wrong. Unlike the previous case, this involves modifying an effect that is designed to be implicit *after* it has been executed. For instance, an intelligent whiteboard can organize a user's sketched notes as visual clusters, but the user can intentionally "override" these clusters if he or she disagrees with the system interpretation [51] (Figure 2b).

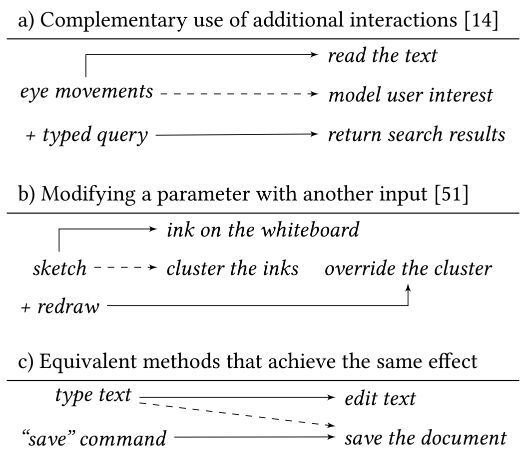


Figure 2: Diagrams showing an interaction that is designed to be implicit (dashed lines) within the context of other interactions.

Finally, there can be cases of false negatives, namely situations in which the system fails to provide an appropriate response when needed. For example, a text editor can utilize a user's editing actions as a trigger for auto-saving the document, but the user might also want to save the document without having to wait for the auto-save functionality (Figure 2c). These situations require the presence of an equivalent method in the interface that achieves the same effect.

Determining the Implicitness of an Effect

Specifying the entity of implicit also directs us to the more precise question of whether prior expectations of implicitness matches with users' reported mental attitudes toward an input–effect pair. Let's illustrate this by reviewing previous work through the lens of the updated definition. In their study of messaging applications, Hoyle et al. initially designate "read receipts" (i.e., visual notices that inform the sender that the receiver has opened the message) as "implicit" effects that occur as a by-product of viewing the message [47] (Figure 3a). Yet their interviews show that read receipts can be an explicit effect, such as when users abstain from opening messages to avoid informing the other party of their action.

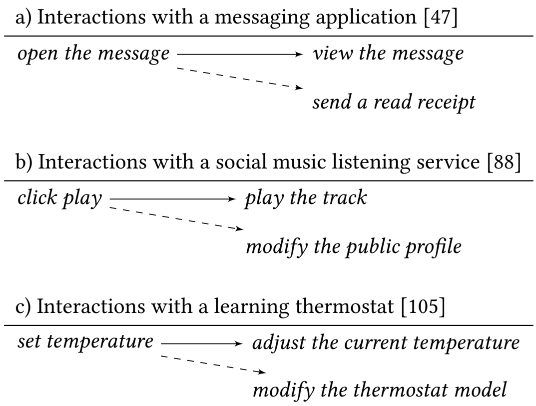


Figure 3: Depictions of input–effect pairs that are expected to be explicit and implicit (respectively shown as solid and dashed lines). The implicitness of a particular input–effect pair contradicts prior expectations in each example.

Outside implicit interaction literature, interviews have shown that users can put extra effort to achieve or modify seemingly implicit effects. Silfverberg et al. report a particular social media-linked music listening practice in which users hit the "play" button, but for presenting themselves as listening to a particular track in their social media profile instead of actually listening to it [88]. In this case, the input (hitting the play button) results in two effects, playing the music track and changes in the user's social media profile (Figure 3b). Contrary to the expectations, it is the latter effect that is intended by the user. Yang and Newman investigate the use practices that emerge around a learning thermostat [105]. When a user sets a new temperature, the thermostat facilitates both the instant effect of changing the temperature and the long-term effect of building a model of user preferences (Figure 3c). A finding from their interviews is that users can adapt their behavior of setting the temperature to prevent unwanted adaptations once they realize the limitations of the device's learning capability. Note that, in the above-mentioned cases, the explicitness of the interaction is not negative per se; the observations rather exemplify situations in which users adjust their mental attitude to compensate for the inappropriate system behavior.

Finally, we see a more precise definition helpful for reasoning about implicitness through interface properties, by expanding the scope of analysis from isolated "implicit interactions" to the *application context* an input–effect pair is situated in. For instance, among different effects that result from the same input method, some can be more visible, thus more likely to be noticed and intended by the user. Or, among equivalent input methods that lead to the same effect, some can require less interaction steps than others or require less skill, increasing the likelihood of that particular input for intentionally producing the effect.

The reasoning can be developed further by paying attention to the combination of different input–effect pairs. Consider our observation that implicit interactions are often enabled by other effects that are caused by the same input; bundling of multiple effects is a source of uncertainty for attributing user intentions to observable behavior. On the other hand, lack of an equivalent method for achieving a certain effect increases the likelihood that a particular input–effect pair becomes explicit. For example, a text editor can utilize users' editing actions as a trigger for auto-saving the document, but in the absence of a "save" command, users can resort to editing the document as a means to trigger an auto-save. Previous work documents such instances in which interactions that were designed to be "implicit-only" turned out to be intentional and at the focus of users' attention because they were the only means to reach a specific outcome [31].

5 DISCUSSION: ADDITIONAL CONSIDERATIONS FOR RESEARCH

HCI has long acknowledged that interactional achievements can depend on external conditions that have no mental counterpart in a user's mind. Designing for implicit interactions represent a constructive take on this insight by expanding the scope of design to unintentional and by potentially increasing the role of external dependencies even further. We identified the particular methodological challenges of this

Table 2: A list of additional considerations for research.

Determining implicitness by design: Do users' mental attitudes match with prior research or design expectations (if so, in which cases)? How do they adapt over time in relation to a design intervention?
Establishing criteria for design and evaluation: What are the assumptions about what is appropriate (from the perspective of the designers and the users)? To what extent do users' prior intentions represent what is appropriate? Does the expected benefit of a design intervention depends on the interaction being implicit (i.e., if a targeted implicit effect turns out to be explicit, is it still beneficial)?
Scoping design problems: Which effects are bundled together (co-occur as a result of the same input)? How does the bundling of effects make certain interaction outcomes harder? Are there any equivalent methods for achieving the same effect or correcting it? How does designing for implicit interaction can reinforce or modify existing use practices?

constructive orientation and showed how an updated definition can be deployed to address them with greater precision. While our analysis focused on the diverse uses of the term implicit interaction, we find the exposed challenges representative of more general concerns in HCI, namely the conceptualization of user and system agencies through design and the role of user intentions in research. Below, we discuss the broader implications of our analysis and identify additional considerations for design and evaluation. These are presented as a checklist of actionable research questions in Table 2.

Taking Implicit "Interaction" Seriously

First of all, framing implicitness as an interaction quality emphasizes the aspect of situational accomplishment; implicitness is not a design feature but an empirical claim about users' state in a given situation. Beyond terminological precision, what is the value of this insight for how we design and evaluate interfaces? If interfaces themselves are not implicit or explicit, what are they? And if implicitness is accomplished through interaction, how does it inform design?

For design, what is more consequential than the explicit–implicit dichotomy are the assumptions they rely on, namely designer's expectations about users' goals and what is appropriate for a given use case. These assumptions guide the practical design problems of bundling different effects or providing complementary input–effect relationships. Thus, even though interfaces themselves are not implicit, they can be expressed in the intention-agnostic language of different input–effect relationships. Distinguishing designers' assumptions from the actual interface properties is important as it exposes the role of interface in promoting or precluding different interaction outcomes. For example, an understanding of implicit as predetermined by design puts emphasis on the decreased user effort achieved through bundling multiple effects. Yet it can obscure how bundling multiple effects can make certain interaction outcomes harder, such as opening a message without sending a read receipt or adjusting the room temperature without changing the user model of the thermostat.

For evaluation, framing implicit as a quality of interaction calls for paying attention to users' actual mental attitudes and their long-term adaptations. Systems that are designed to facilitate implicit interaction often build on users' existing action routines to facilitate additional appropriate effects. The implicitness of an effect is then grounded on the expectation that user goals or actions will remain unchanged after a design intervention. What is missed in this conception is how design interventions can invalidate such expectations if users reformulate their goals in relation to these effects or avoid certain actions to prevent what they perceive as unwanted effects.

While our discussion focused on the term implicit, the relevance for similar concepts should be obvious. Distinctions between "user-controlled" versus "mixed-initiative"[3], "proactive" [59, 80, 104], "adaptive" or "automatic" often restate user intentionality in terms of system properties. Yet whether a system proactively adapts its behavior in response to the user input, or whether the users intentionally exploits the system adaptation cannot be stated independently from their knowledge and goals in a particular situation. Here, our emphasis parallels previous insights on foreground–background as situational properties [92], the accomplishment of meaning through interaction [24], and more recent criticisms against the use of qualifiers such as "natural" or "intuitive" [11, 68, 98] to describe interfaces. Thus, part of the problem can be attributed to a confusing mixture of experiential and system properties. That said, we distinguish this mixture from the more productive efforts of anticipating and designing for implicitness. And this is where we see the value of conceptual precision; a testable definition enables researchers to specify why implicitness is pursued and how prior design assumptions can be validated.

User Intentions from Normative to Descriptive

Finally, we frame designing for implicit interactions as part of a broader change in attitude in HCI that prioritizes appropriateness of action outcomes over their intentionality. Accompanying this shift is the changing role of user intentions from normative (i.e., as setting the evaluative standard of interaction [44]) to descriptive: Intentions can explain why the user has performed an action or found a system response appropriate, but as mental antecedents they provide incomplete resources for activity completion and can be negotiable. Thus, instead of taking the normativeness of user intentions for granted, *researchers should treat the extent user intentions become a measure of user acceptance as an empirical question* by taking situational factors such as users' capacity to model and control system behavior into account. Consequently, we argue for the need to expand the scope of evaluation from performance metrics and observable breakdowns to assessing how different interface configurations subtly influence user behavior and lead to qualitatively different outcomes.

Interface mediation in turn emphasizes the *normative aspect of designing* and the role of the designers in conceptualizing what is appropriate for a given situation. In HCI and design, similar thoughts led some to expand the notion of normative intentionality to artifacts under the phrases "material intentionality" [94] or "design with intent" [58]. Yet we observe that these discussions have been relatively absent in the context of implicit interactions. Normative aspect of designing becomes particularly relevant when user intentions are treated as incentives for end goals that are designed to benefit other parties. Potential trade-offs in these situations point to ethical questions related to users' informed consent and the distribution of burden and benefit.

6 CONCLUSION

In the wake of increased sensing capabilities and complex inference mechanisms, the concept of implicit interaction is ever more relevant. In this paper, we have drawn attention to the present ambiguity of the term and analyzed the consequences of different meanings for establishing why we facilitate implicit interactions and how we validate its benefits. In the most general level, our investigation is a call for researchers and designers to specify their particular use of implicit interaction instead of treating the term as self-evident. Our review provides a reference for future work by clarifying the differences between various meanings. We see the conceptual clarity brought by as a necessary step for larger, more systematic reviews.

In more particular, we call for paying closer attention to the implications of defining explicit–implicit distinction through intentionality and as a quality of an input–effect relationship in order to address the methodological challenges with greater precision. Importantly, we stress implicitness as something accomplished through interaction, which highlights the need to critically examine what prior distinctions of implicit and explicit presume about users' goals, their capacity to model the system behavior and what is appropriate for a given situation. We expect the additional considerations to be helpful as interfaces that aim to facilitate implicit interactions increasingly go beyond proof-of-concept and are deployed in more complex settings.

ACKNOWLEDGMENTS

We would like to thank Oswald Barral, Antti Salovaara, Luis A. Leiva and the anonymous reviewers for their feedback on the draft versions of this paper. This work has been supported by the European Commission (MindSee FP7-ICT; Grant Agreement #611570) and the Academy of Finland (Evidence, Grant Agreement #286440).

REFERENCES

[1] G. D. Abowd, E. D. Mynatt, and T. Rodden. 2002. The human experience [of ubiquitous computing]. *IEEE Pervasive Computing* 1, 1 (Jan 2002), 48–57. https://doi.org/10.1109/MPRV.2002.993144

[2] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06).* ACM, New York, NY, USA, 19–26. https://doi.org/10.1145/1148170.1148177

[3] James F Allen, Curry I Guinn, and Eric Horvitz. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14, 5 (1999), 14–23.

[4] James F Allen and C Raymond Perrault. 1980. Analyzing intention in utterances. *Artificial intelligence* 15, 3 (1980), 143–178.

[5] Robert B Arundale. 2008. Against (Gricean) intentions at the heart of human interaction. *Intercultural Pragmatics* 5, 2 (2008), 229–258.

[6] Saskia Bakker, Elise van den Hoven, and Berry Eggen. 2015. Peripheral interaction: characteristics and considerations. *Personal and Ubiquitous Computing* 19, 1 (01 Jan 2015), 239–254. https://doi.org/10.1007/s00779-014-0775-2

[7] Till Ballendat, Nicolai Marquardt, and Saul Greenberg. 2010. Proxemic Interaction: Designing for a Proximity and Orientation-aware Environment. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10).* ACM, New York, NY, USA, 121–130. https://doi.org/10.1145/1936652.1936676

[8] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making Sense of Sensing Systems: Five Questions for Designers and Researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02).* ACM, New York, NY, USA, 415–422. https://doi.org/10.1145/503376.503450

[9] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human-Computer Interaction* 16, 2-4 (2001), 193–212. https://doi.org/10.1207/S15327051HCI16234_05

[10] Susanne Bødker and Clemens Nylandsted Klokmose. 2011. The Human–Artifact Model: An Activity Theoretical Approach to Artifact Ecologies. *Human–Computer Interaction* 26, 4 (2011), 315–371. https://doi.org/10.1080/07370024.2011.626709 arXiv:http://dx.doi.org/10.1080/07370024.2011.626709

[11] Susanne Bødker and Clemens Nylandsted Klokmose. 2016. Dynamics, Multiplicity and Conceptual Blends in HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16).* ACM, New York, NY, USA, 2538–2548. https://doi.org/10.1145/2858036.2858530

[12] John Bowers and John Churcher. 1988. Local and Global Structuring of Computer Mediated Communication: Developing Linguistic Perspectives on CSCW in Cosmos. In *Proceedings of the 1988 ACM Conference*

on Computer-supported Cooperative Work (CSCW '88). ACM, New York, NY, USA, 125–139. https://doi.org/10.1145/62266.62277

[13] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1393–1402. https://doi.org/10.1145/2702123.2702252

[14] Georg Buscher, Andreas Dengel, and Ludger van Elst. 2008. Query Expansion Using Gaze-based Feedback on the Subdocument Level. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08)*. ACM, New York, NY, USA, 387–394. https://doi.org/10.1145/1390334.1390401

[15] William Buxton. 1995. Integrating the periphery and context: A new taxonomy of telematics. In *Proceedings of graphics interface*, Vol. 95. 239–246.

[16] Robyn Carston. 2009. The Explicit/Implicit Distinction in Pragmatics and the Limits of Explicit Communication. *International Review of Pragmatics* 1, 1 (2009), 35–62. https://doi.org/10.1163/187731009X455839

[17] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I Did That! Measuring Users' Experience of Agency in Their Own Actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2025–2034. https://doi.org/10.1145/2207676.2208350

[18] Mary Czerwinski, Susan Dumais, George Robertson, Susan Dziadosz, Scott Tiernan, and Maarten van Dantzich. 1999. Visualizing Implicit Queries for Information Management and Retrieval. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 560–567. https://doi.org/10.1145/302979.303158

[19] Tamara Denning, Kevin Bowers, Marten van Dijk, and Ari Juels. 2011. Exploring Implicit Memory for Painless Password Recovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2615–2618. https://doi.org/10.1145/1978942.1979323

[20] Zoltan Dienes and Josef Perner. 1999. A theory of implicit and explicit knowledge. *Behavioral and brain sciences* 22, 5 (1999), 735–808.

[21] Alan Dix. 2002. Beyond intention-pushing boundaries with incidental interaction. In *Proceedings of Building Bridges: Interdisciplinary Context-Sensitive Computing, Glasgow University*, Vol. 9.

[22] Alan John Dix. 1991. *Formal methods for interactive systems*. Vol. 16. Academic Press London.

[23] Alan J Dix and Colin Runciman. 1985. Abstract models of interactive systems. In *People and Computers: Designing the interface*. Cambridge University Press, 13–22.

[24] Paul Dourish. 2004. *Where the action is: the foundations of embodied interaction*. MIT press.

[25] Paul Dourish and Victoria Bellotti. 1992. Awareness and Coordination in Shared Workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work (CSCW '92)*. ACM, New York, NY, USA, 107–114. https://doi.org/10.1145/143457.143468

[26] Hubert L. Dreyfus. 1993. Heidegger's Critique of the Husserl/Searle Account of Intentionality. *Social Research* 60, 1 (1993), 17–38. http://www.jstor.org/stable/40970727

[27] Hubert L Dreyfus. 1996. The current relevance of Merleau-Ponty's phenomenology of embodiment. *The Electronic Journal of Analytic Philosophy* 4 (1996).

[28] Stephen H. Fairclough. 2009. Fundamentals of physiological computing. *Interacting with Computers* 21, 1–2 (2009), 133 – 145. https://doi.org/10.1016/j.intcom.2008.10.011

[29] Russell H. Fazio and Michael A. Olson. 2003. Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology* 54, 1 (2003), 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

[30] Gerhard Fischer. 2001. User Modeling in Human–Computer Interaction. *User Modeling and User-Adapted Interaction* 11, 1 (01 Mar 2001), 65–86. https://doi.org/10.1023/A:1011145532042

[31] Lyndsey Fisk, Marcus Carter, Behnaz Rostami Yeganeh, Frank Vetere, and Bernd Ploderer. 2014. Implicit and Explicit Interactions in Video Mediated Collaboration. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 250–259. https://doi.org/10.1145/2686612.2686650

[32] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. 2005. Evaluating Implicit Measures to Improve Web Search. *ACM Trans. Inf. Syst.* 23, 2 (April 2005), 147–168. https://doi.org/10.1145/1059981.1059982

[33] Anthony G Greenwald and Mahzarin R Banaji. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review* 102, 1 (1995), 4.

[34] Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology* 74, 6 (1998), 1464.

[35] H Paul Grice. 1957. Meaning. *The philosophical review* (1957), 377–388.

[36] H Paul Grice. 1989. Logic and conversation. In *Studies in the Way of Words*. Harvard University Press.

[37] Beverly L. Harrison, Kenneth P. Fishkin, Anuj Gujar, Carlos Mochon, and Roy Want. 1998. Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 17–24. https://doi.org/10.1145/274644.274647

[38] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A System for Implicit Audience Engagement Sensing Using Electroencephalography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5114–5119. https://doi.org/10.1145/3025453.3025669

[39] Michael Haugh. 2008. Intention in pragmatics. *Intercultural Pragmatics* 5, 2 (2008), 99–110.

[40] John Heritage. 1990. Intention, meaning and strategy: Observations on constraints on interaction analysis. *Research on Language & Social Interaction* 24, 1-4 (1990), 311–332.

[41] Ken Hinckley, Jeff Pierce, Eric Horvitz, and Mike Sinclair. 2005. Foreground and Background Interaction with Sensor-enhanced Mobile Devices. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (March 2005), 31–52. https://doi.org/10.1145/1057237.1057240

[42] Ken Hinckley, Jeff Pierce, Mike Sinclair, and Eric Horvitz. 2000. Sensing Techniques for Mobile Interaction. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*. ACM, New York, NY, USA, 91–100. https://doi.org/10.1145/354401.354417

[43] Ken Hinckley and Mike Sinclair. 1999. Touch-sensing Input Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 223–230. https://doi.org/10.1145/302979.303045

[44] Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5040–5052. https://doi.org/10.1145/3025453.3025765

[45] Eric Horvitz, Jack Breese, David Heckerman, David Hovel, and Koos Rommelse. 1998. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 256–265.

[46] Eric Horvitz, Carl Kadie, Tim Paek, and David Hovel. 2003. Models of Attention in Computing and Communication: From Principles to Applications. *Commun. ACM* 46, 3 (March 2003), 52–59. https://doi.org/10.1145/636772.636798

[47] Roberto Hoyle, Srijita Das, Apu Kapadia, Adam J. Lee, and Kami Vaniea. 2017. Was My Message Read?: Privacy and Signaling on Facebook Messenger. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3838–3842. https://doi.org/10.1145/3025453.3025925

[48] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM, New York, NY, USA, 234–241. https://doi.org/10.1145/258549.258715

[49] Lars-Erik Janlert and Erik Stolterman. 2017. The Meaning of Interactivity: Some Proposals for Definitions and Measures. *Human-Computer Interaction* 32, 3 (2017), 103–138. https://doi.org/10.1080/07370024.2016.1226139

[50] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. 2007. Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. *ACM Trans. Inf. Syst.* 25, 2, Article 7 (April 2007). https://doi.org/10.1145/1229179.1229181

[51] Wendy Ju, Brian A. Lee, and Scott R. Klemmer. 2008. Range: Exploring Implicit Interaction Through Electronic Whiteboard Design. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 17–26. https://doi.org/10.1145/1460563.1460569

[52] Wendy Ju and Larry Leifer. 2008. The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious. *Design Issues* 24, 3 (2008), 72–84. https://doi.org/10.1162/desi.2008.24.3.72 arXiv:http://dx.doi.org/10.1162/desi.2008.24.3.72

[53] Jakob Karolus, Paweł W. Wozniak, Lewis L. Chuang, and Albrecht Schmidt. 2017. Robust Gaze Features for Enabling Language Proficiency Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2998–3010. https://doi.org/10.1145/3025453.3025601

[54] Diane Kelly and Jaime Teevan. 2003. Implicit Feedback for Inferring User Preference: A Bibliography. *SIGIR Forum* 37, 2 (Sept. 2003), 18–28. https://doi.org/10.1145/959258.959260

[55] David Kirsh. 1995. The Intelligent Use of Space. *Artificial Intelligence* 73, 1-2 (Feb. 1995), 31–68. https://doi.org/10.1016/0004-3702(94)00017-U

[56] Kai Kuikkaniemi, Toni Laitinen, Marko Turpeinen, Timo Saari, Ilkka Kosunen, and Niklas Ravaja. 2010. The Influence of Implicit and Explicit Biofeedback in First-person Shooter Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 859–868. https://doi.org/10.1145/1753326.1753453

[57] Stephen C Levinson. 1983. *Pragmatics*. Cambridge University Press.

[58] Dan Lockton, David Harrison, and Neville A. Stanton. 2010. The Design with Intent Method: A design tool for influencing user behaviour. *Applied Ergonomics* 41, 3 (2010), 382–392. https://doi.org/10.1016/j.apergo.2009.09.001 Special Section: Recycling centres and waste handling âĂŞ a workplace for employees and users.

[59] Pattie Maes. 1994. Agents That Reduce Work and Information Overload. *Commun. ACM* 37, 7 (July 1994), 30–40. https://doi.org/10.1145/176789.176792

[60] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human–Computer Interaction*. Springer London, London, 39–65. https://doi.org/10.1007/978-1-4471-6392-3_3

[61] David C McClelland, Richard Koestner, and Joel Weinberger. 1989. How do self-attributed and implicit motives differ? *Psychological review* 96, 4 (1989), 690–702.

[62] Yoshiro Miyata and Donald A Norman. 1986. Psychological issues in support of multiple activities. In *User centered system design: New perspectives on human-computer interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 265–284.

[63] Elizabeth D. Mynatt, Takeo Igarashi, W. Keith Edwards, and Anthony LaMarca. 1999. Flatland: New Dimensions in Office Whiteboards. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 346–353. https://doi.org/10.1145/302979.303108

[64] Jakob Nielsen. 1993. Noncommand User Interfaces. *Commun. ACM* 36, 4 (April 1993), 83–99. https://doi.org/10.1145/255950.153582

[65] Donald A. Norman. 1986. Cognitive engineering. In *User centered system design: New perspectives on human-computer interaction*, Donald A. Norman and Stephen W. Draper (Eds.). L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 31–61.

[66] Donald A Norman and Tim Shallice. 1980. *Attention to action: Willed and automatic control of behavior*. Technical Report. California University San Diego La Jolla Center for Human Information Processing.

[67] Brian A Nosek, Mahzarin R Banaji, and Anthony G Greenwald. 2002. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice* 6, 1 (2002), 101. https://doi.org/10.1037//1089-2699.6.1.101

[68] Kenton O'Hara, Richard Harper, Helena Mentis, Abigail Sellen, and Alex Taylor. 2013. On the Naturalness of Touchless: Putting the "Interaction" Back into NUI. *ACM Trans. Comput.-Hum. Interact.* 20, 1, Article 5 (April 2013), 25 pages. https://doi.org/10.1145/2442106.2442111

[69] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. 2005. Interaction in 4-second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 919–928. https://doi.org/10.1145/1054972.1055101

[70] Tomi Peltola, Mustafa Mert Celikok, Pedram Daee, and Samuel Kaski. 2018. Modelling User's Theory of AI's Mind in Interactive Intelligent Systems. arXiv:arXiv:1809.02869

[71] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 261–270. https://doi.org/10.1145/2501988.2501998

[72] Susanne Poeller, Max V. Birk, Nicola Baumann, and Regan L. Mandryk. 2018. Let Me Be Implicit: Using Motive Disposition Theory to Predict and Explain Behaviour in Digital Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 190, 15 pages. https://doi.org/10.1145/3173574.3173764

[73] Henning Pohl and Roderick Murray-Smith. 2013. Focused and Casual Interactions: Allowing Users to Vary Their Level of Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2223–2232. https://doi.org/10.1145/2470654.2481307

[74] Zachary Pousman and John Stasko. 2006. A Taxonomy of Ambient Information Systems: Four Patterns of Design. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '06)*. ACM,

New York, NY, USA, 67–74. https://doi.org/10.1145/1133265.1133277

[75] Kathrin Probst, David Lindlbauer, Michael Haller, Bernhard Schwartz, and Andreas Schrempf. 2014. A Chair As Ubiquitous Input Device: Exploring Semaphoric Chair Gestures for Focused and Peripheral Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4097–4106. https://doi.org/10.1145/2556288.2557051

[76] Filip Radlinski and Thorsten Joachims. 2005. Query Chains: Learning to Rank from Implicit Feedback. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD '05)*. ACM, New York, NY, USA, 239–248. https://doi.org/10.1145/1081870.1081899

[77] Arthur S Reber. 1989. Implicit learning and tacit knowledge. *Journal of experimental psychology: General* 118, 3 (1989), 219.

[78] Mark E. Revesman and Joel S. Greenstein. 1983. Application of a Model of Human Decision Making for Human/Computer Communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '83)*. ACM, New York, NY, USA, 107–111. https://doi.org/10.1145/800045.801591

[79] Laurie A. Rudman, Anthony G. Greenwald, Deborah S. Mellott, and Jordan L. K. Schwartz. 1999. Measuring the Automatic Components of Prejudice: Flexibility and Generality of the Implicit Association Test. *Social Cognition* 17, 4 (01 Dec 1999), 437–465. https://doi.org/10.1521/soco.1999.17.4.437

[80] Antti Salovaara and Antti Oulasvirta. 2004. Six Modes of Proactive Resource Management: A User-centric Typology for Proactive Behaviors. In *Proceedings of the Third Nordic Conference on Human-computer Interaction (NordiCHI '04)*. ACM, New York, NY, USA, 57–60. https://doi.org/10.1145/1028014.1028022

[81] Munehiko Sato, Rohan S. Puri, Alex Olwal, Yosuke Ushigome, Lukas Franciszkiewicz, Deepak Chandra, Ivan Poupyrev, and Ramesh Raskar. 2017. Zensei: Embedded, Multi-electrode Bioimpedance Sensing for Implicit, Ubiquitous User Recognition. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3972–3985. https://doi.org/10.1145/3025453.3025536

[82] Daniel L Schacter. 1987. Implicit memory: History and current status. *Journal of experimental psychology: learning, memory, and cognition* 13, 3 (1987), 501–518.

[83] Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal Technologies* 4, 2 (2000), 191–199. https://doi.org/10.1007/BF01324126

[84] Albrecht Schmidt, Michael Beigl, and Hans Gellersen. 1999. There is more to context than location. *Computers & Graphics* 23, 6 (1999), 893 – 901. https://doi.org/10.1016/S0097-8493(99)00120-X

[85] Richard W Schmidt. 1990. The role of consciousness in second language learning. *Applied linguistics* 11, 2 (1990), 129–158.

[86] John R Searle. 1975. Indirect speech acts. In *Syntax and Semantics Volume 3: Speech Acts*. Academic Press, 59–82.

[87] Tomoki Shibata, Evan M. Peck, Daniel Afergan, Samuel W. Hincks, Beste F. Yuksel, and Robert J.K. Jacob. 2014. Building Implicit Interfaces for Wearable Computers with Physiological Inputs: Zero Shutter Camera and Phylter. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST'14 Adjunct)*. ACM, New York, NY, USA, 89–90. https://doi.org/10.1145/2658779.2658790

[88] Suvi Silfverberg, Lassi A. Liikkanen, and Airi Lampinen. 2011. "I'll Press Play, but I won't Listen": Profile Work in a Music-focused Social Network Service. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/1958824.1958855

[89] Xiaojun Su, Oscar Kin-Chung Au, and Rynson W.H. Lau. 2014. The Implicit Fan Cursor: A Velocity Dependent Area Cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 753–762. https://doi.org/10.1145/2556288.2557095

[90] Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge university press.

[91] Mingfei Sun, Zhenjie Zhao, and Xiaojuan Ma. 2017. Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 556–567. https://doi.org/10.1145/3025453.3025469

[92] Dag Svanaes. 2001. Context-aware Technology: A Phenomenological Perspective. *Hum.-Comput. Interact.* 16, 2 (Dec. 2001), 379–400. https://doi.org/10.1207/S15327051HCI16234_17

[93] Erin Treacy Solovey, Daniel Afergan, Evan M. Peck, Samuel W. Hincks, and Robert J. K. Jacob. 2015. Designing Implicit Interfaces for Physiological Computing: Guidelines and Lessons Learned Using fNIRS. *ACM Trans. Comput.-Hum. Interact.* 21, 6, Article 35 (Jan. 2015), 27 pages. https://doi.org/10.1145/2687926

[94] Peter-Paul Verbeek. 2009. Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology. *NanoEthics* 3, 3 (03 Dec 2009), 231–242. https://doi.org/10.1007/s11569-009-0077-8

[95] Roel Vertegaal. 2002. Designing Attentive Interfaces. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications (ETRA '02)*. ACM, New York, NY, USA, 23–30. https://doi.org/10.1145/507072.507077

[96] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 137–146. https://doi.org/10.1145/1029632.1029656

[97] Annika Waern. 2016. The Ethics of Unaware Participation in Public Interventions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 803–814. https://doi.org/10.1145/2858036.2858188

[98] Alex Wennberg, Henrik Åhman, and Anders Hedman. 2018. The Intuitive in HCI: A Critical Discourse Analysis. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction (NordiCHI '18)*. ACM, New York, NY, USA, 505–514. https://doi.org/10.1145/3240167.3240202

[99] Ryen W. White, Ian Ruthven, and Joemon M. Jose. 2005. A Study of Factors Affecting the Utility of Implicit Relevance Feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*. ACM, New York, NY, USA, 35–42. https://doi.org/10.1145/1076034.1076044

[100] John Williamson. 2006. *Continuous uncertain interaction.* Ph.D. Dissertation. University of Glasgow.

[101] Deirdre Wilson and Dan Sperber. 2002. Relevance Theory. In *Handbook of Pragmatics*. Blackwell.

[102] Timothy D Wilson, Samuel Lindsey, and Tonya Y Schooler. 2000. A model of dual attitudes. *Psychological review* 107, 1 (2000), 101. https://doi.org/10.1037/0033-295X.107.1.101

[103] Terry Winograd and Fernando Flores. 1986. *Understanding computers and cognition.* Addison-Wesley Publishing Company.

[104] Michael Wooldridge and Nicholas R Jennings. 1995. Intelligent agents: Theory and practice. *The knowledge engineering review* 10, 2 (1995), 115–152. https://doi.org/10.1017/S0269888900008122

[105] Rayoung Yang and Mark W. Newman. 2013. Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive*

and Ubiquitous Computing (UbiComp '13). ACM, New York, NY, USA, 93–102. https://doi.org/10.1145/2493432.2493489

[106] Shumin Zhai. 2003. What's in the Eyes for Attentive Input. *Commun. ACM* 46, 3 (March 2003), 34–39. https://doi.org/10.1145/636772.636795

[107] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

Aalto-DD 72/2020

ISBN 978-952-60-3870-4 (printed)
ISBN 978-952-60-3873-5 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Arts, Design and Architecture
Department of Design
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS