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?

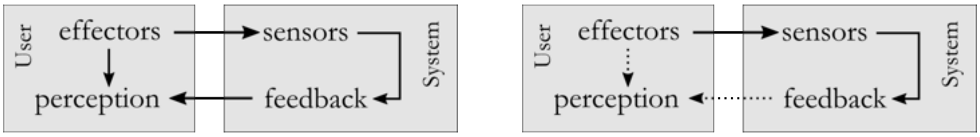

Figure 3.1. The unit of visual monitoring for the publications I and II was the individual human. The interaction with interactive systems often requires users to visually monitor their own actions as well as the system feedback (left). The research program focused on supporting input methods in which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

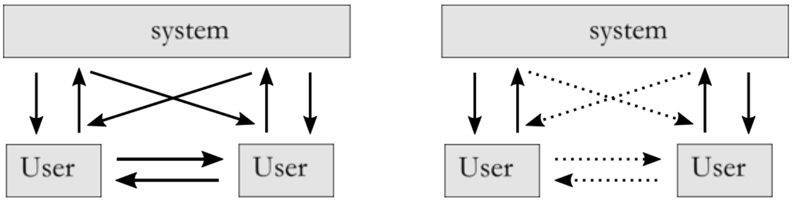

Figure 3.2. The unit of visual monitoring for Publication III was the group level. The publication proposed adapting the system response by distinguishing situations in which the users visually monitor each others' actions (left) or not (right).

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.

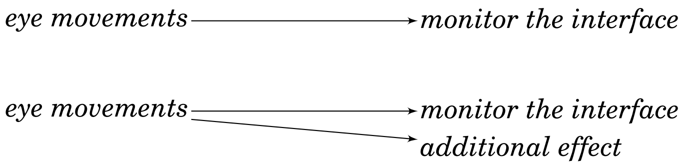

Figure 3.3. The comparison of two cases in which 1) eye movements only facilitate visual monitoring of the interface (above) or their communicative use results in multiple effects (below).

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]:

$$\begin{tikzcd}
\textit{eye movements} \arrow[r] \arrow[dr] & \textit{monitor the interface} \\
& \textit{trigger a command}
\end{tikzcd}$$

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:

$$\begin{tikzcd}
\textit{eye movements} \arrow[r] & \textit{monitor the interface} \\
\textit{eye movements + manual confirmation} \arrow[r] \arrow[ur] & \textit{trigger a command}
\end{tikzcd}$$

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:

$$\begin{tikzcd}
\textit{eye / head movements} \arrow[r] & \textit{monitor the interface} \\
\textit{eye / head movements + manual action} \arrow[r] \arrow[ur] & \textit{adaptive input handling}
\end{tikzcd}$$

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).

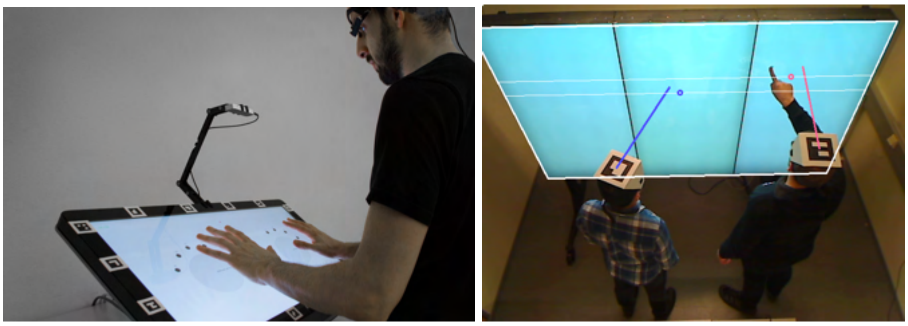

Figure 3.4. Different operationalizations of visual attention, using eye tracking for single users on a 27” touch screen (left, Publication II) and head tracking for multiple users on a wall-sized vertical display (right, Publication III).

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

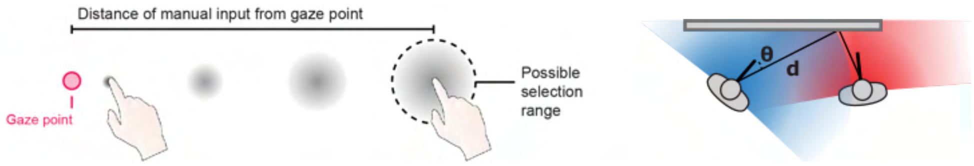

Figure 3.5. For single user touch screen interactions visual monitoring has been operationalized in terms of the distance of the gaze point to the manual input (left, Publication I and II). For users interacting with a wall-size display it has been operationalized in terms of the angle (θ) and distance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

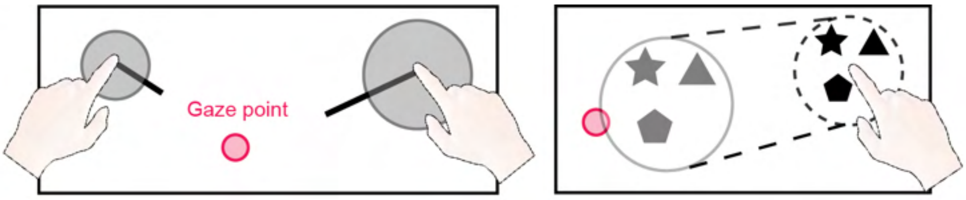

Figure 3.6. Providing peripheral awareness (left) and warping information content around manual input position to gaze point (middle) are two possible visual feedback techniques to communicate system interpretation of user input back to user (Publication I).

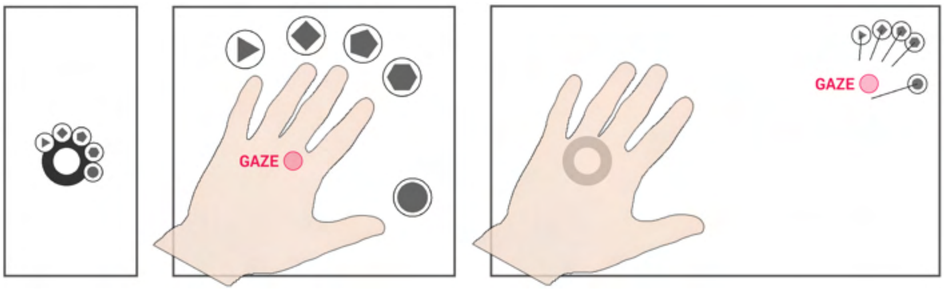

Figure 3.7. An example of warping information: the system adapts where to show visual feedback about a widget based on user's gaze direction (Publication II).

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).

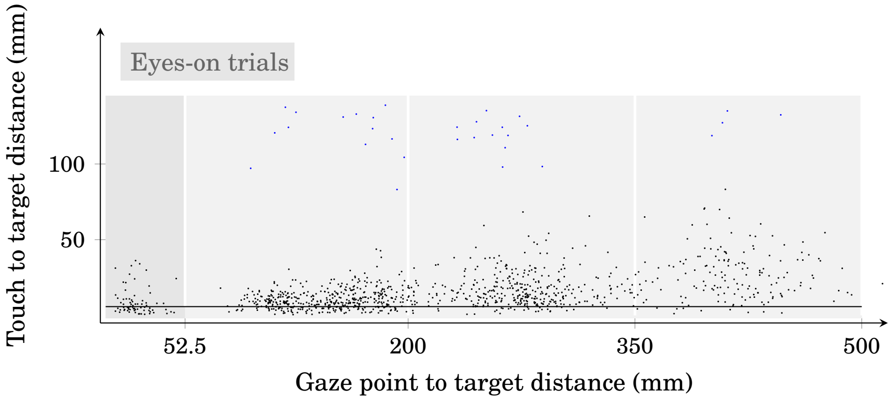

Figure 4.1. Scatter plot of peripheral target acquisition tasks across all participants. The horizontal line indicates the visual boundary of the circle target (rad=5.8mm). The darker background indicates the eyes-on in which the target was within the boundary of the circle the participants had to keep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

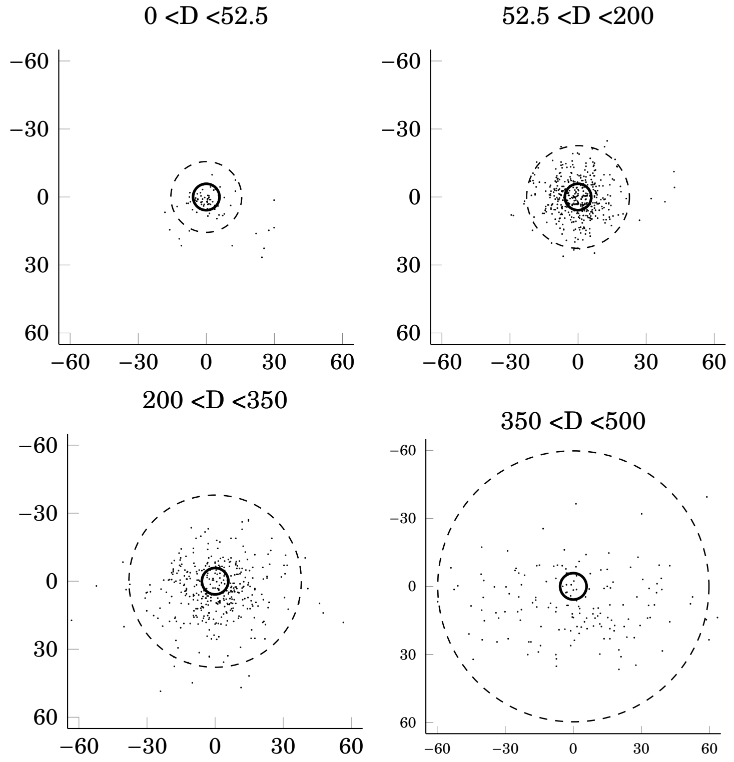

Figure 4.2. The distribution of touch points relative to the target across all users for chosen ranges of distance between gaze point and target. The dashed circles are the 95 % confidence circles. The inner solid circles show the target visual boundary. All units in mm.

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.