Integrative visual augmentation content and its optimization based on human visual processing

DISSERTATION

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard-Karls-Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

OLGA LUKASHOVA-SANZ

aus Kostroma, Russland

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	16.12.2022
Dekan der Math.-Nat. Fakultät:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Siegfried Wahl
2. Berichterstatter:	Prof. Dr. Martin Butz

Contents

1 Introduction **11**

1.1 Motivation . . . 11

1.2 Augmented reality (AR) . . . 12

1.2.1 Mixed reality taxonomy . . . 12

1.2.2 Level of realism and simulation of AR in VR . . . 13

1.2.3 Ecological approach to AR . . . 14

1.3 Human “optimal” performance . . . 15

1.4 Design principles for augmentation based on human visual processing in challenging tasks . . . 16

1.4.1 Posner attentional paradigm: attentional benefits and costs . . . 17

1.4.2 Subtle peripheral augmentation in a dual task interleaving scenario 19

1.4.3 Saliency-based augmentation in a visual search task . . . 20

2 Objectives **23**

3 The impact of shape-based cue discriminability on attentional performance **25**

3.1 Abstract . . . 25

3.2 Introduction . . . 25

3.3 Methods . . . 27

3.3.1 Subjects . . . 27

3.3.2 Apparatus and stimuli . . . 28

3.3.2.1 Apparatus . . . 28

3.3.2.2 Shape-based spatial cue . . . 28

3.3.2.3 Stimuli . . . 30

3.3.3 Experimental procedure . . . 30

3.3.3.1 General procedure . . . 30

3.3.3.2 Cue sensitivity test . . . 31

3.3.3.3 Training session . . . 31

3.3.3.4 Main experiment . . . 32

3.3.4 Analysis . . . 33

3.3.4.1 Cue sensitivity test . . . 33

3.3.4.2 Main experiment . . . 34

3.3.4.3 Cue processing temporal cost . . . 34

3.4 Results . . . 35

3.4.1 Cue sensitivity 35
3.4.2 Main Experiment 36
3.4.3 Cue processing temporal costs 38
3.5 Discussion and Conclusion 40
3.6 Data Availability Statement 44
3.7 Additional information 44
3.7.1 Author contributions 44
3.7.2 Funding 44
3.7.3 Conflict of interest 44

4 Augmentation impacts strategy and gaze distribution in a dual-task interleaving scenario 45
4.1 Abstract 45
4.2 Introduction 46
4.3 Methods 48
4.3.1 Participants 48
4.3.2 Apparatus and Stimuli 49
4.3.2.1 Apparatus 49
4.3.3 Experimental Procedure 49
4.3.3.1 Dual-task-interleaving Setting: Configuration 49
4.3.3.2 Visual Augmentation: a Head-contingent Peripheral Motion Onset Stimulus 52
4.3.3.3 Unequal Importance of the Tasks Implementation: Unequal Penalty 54
4.3.3.4 Estimation of the Suggested Switch Moment 54
4.3.3.5 Augmented Task-interleaving: Experimental Conditions 54
4.3.3.6 Response Time Test 55
4.3.3.7 Main Experiment: General Procedure 56
4.3.3.8 Gaze Direction Data 56
4.3.4 Analysis 57
4.3.4.1 Response Time Test 57
4.3.4.2 Main Experiment: Behavioral Data 57
4.3.4.3 Gaze Direction Data 58
4.4 Results 58
4.4.1 The Mean Duration on the Blue Task and the Total Score . . . 58
4.4.2 Gaze Direction Distribution 60
4.5 Discussion 61
4.5.1 General Discussion 61

4.5.2 Comparison of the Continuous and the Response-time-based Augmentation (Condition 1 vs Condition 4) 63
4.5.3 Comparison of the Individual and the Fixed Time-lead Based on Response Time (Condition 3 vs Condition 4) 64
4.5.4 Condition 2 64
4.5.5 Conclusion and Outlook 65
4.6 Acknowledgments 66
4.7 Data Availability Statement 66
4.8 Authors' Biography 66

5 Saliency-aware subtle augmentation improves human visual search performance in VR 69
5.1 Abstract 69
5.2 Introduction 69
5.3 Materials and Methods 72
5.3.1 Participants 72
5.3.2 Experimental setup 72
5.3.2.1 Hardware specifications 72
5.3.2.2 Software specifications 72
5.3.3 Virtual environment and stimuli 73
5.3.3.1 Real-world scenes 73
5.3.3.2 Saliency maps 73
5.3.3.3 Blurred images 75
5.3.3.4 Search target 75
5.3.3.5 Search target locations 76
5.3.4 Experimental procedure 77
5.3.4.1 General procedure 77
5.3.4.2 Training phase 78
5.3.4.3 Main experiment 78
5.3.5 Analysis 79
5.3.5.1 Behavioral performance metrics 79
5.3.5.2 Eye movement metrics 79
5.3.5.3 Eye movement raw data pre-processing 80
5.3.5.4 Fixation detection algorithm — I-VT 82
5.4 Results 83
5.4.1 Behavioral data 83
5.4.2 Eye movement data 84
5.5 Discussion 85

5.6 Conclusions 91
5.7 Data availability statement 91
5.8 Additional information 91
5.8.1 Author contributions 91
5.8.2 Funding 92
5.8.3 Acknowledgements 92
5.8.4 Conflict of interest 92

6 Summary **93**

7 Zusammenfassung **97**

List of Figures **101**

List of Tables **107**

Bibliography **109**

Conference contributions and talks related to this work **125**
Peer reviewed conference contributions 125
Non peer reviewed contributions 125

Statement of own contribution **127**
Publication 1 — The impact of shape-based cue discriminability on attentional performance 127
Publication 2 — Augmentation impacts strategy and gaze distribution in a dual-task interleaving scenario 127
Publication 3 — Saliency-aware subtle augmentation improves human visual search performance in VR 128

Acknowledgements **129**

Acronyms list

AR	Augmented reality
VR	Virtual reality
2AFC	Two-alternative-forced choice
SOA	Stimulus onset asynchrony
ISI	Interstimulus interval
OCT	Optical coherence tomography
AOI	Area of interest

1. Introduction

1.1 Motivation

Searching for a friend in a busy room, driving home through a crowded neighbourhood, and looking for a proper door to attend a doctor's appointment are everyday examples of when we are confronted with a large and overwhelming stream of visual information. And yet we appear to comprehend the scenes around us without much effort. It requires the skill to separate the information relevant to the task at hand efficiently. Our attention is the fundamental natural solution to this challenge — it is a selective mechanism that facilitates information processing and allows us to prioritize and select important visual content out of a large cascade of visual input [Carrasco, 2011].

Although, in many situations, the human brain is remarkably good at prioritizing visual information [Eckstein, 2011], it is undoubtedly not always capable of performing optimally, and all the more so in the ever-evolving demanding world. Supplementary visual guidance could enrich our lives from many perspectives on the individual and population scales. To name a few examples, the additional visual information could be used when an untrained employee is practising a new complex task following a set of visual instructions or when a surgeon is examining multiple X-ray scans of patients for potential pathology. In daily life, more often than not, a diversity of visual cues guides our attention, regardless of whether we are conscious of it. Consequently, in many scenarios, visual cues can be a potent tool to influence human behavior through guiding attention [Carrasco, 2011, 2018].

With promptly advancing technology such as augmented reality (AR), new opportunities emerged for the visual cues to guide attention and boost user's performance purposefully [e.g., Booth et al., 2013; Coughlan and Miele, 2017; Dey et al., 2018; Zarraonandia et al., 2014; Cipresso et al., 2018]. While original AR solutions focused on the marketing sector, contemporary AR technology expands beyond its marketing applications. Among others, today's augmented reality solutions made their way into education, industrial design, automotive, and health care sectors [Cipresso et al., 2018; Kim and Choi, 2021]. However, along with offering practical applications, new technology also introduces new challenges.

It is generally agreed that human processing capability to prioritize visual information is limited [Marois and Ivanoff, 2005]. This constraint manifests in various phenomena, such as, for example, change blindness [Simons and Rensink, 2005], or attentional blink [Dux and Rentmarois, 2009]. Furthermore, the phenomenon of visual stimuli competing for the limited cognitive resources is supported by solid evidence from numerous electrophysiological, behavioral, and neuroimaging studies [for review see Beck and Kastner, 2005; Carrasco, 2011]. When it comes to augmentation, the complementary visual information must be processed by the observer along with the natural environment, leading to potential challenges. With rising complexity, augmenting visual input with supplementary content initiates a trade-off between the possible performance gain and the captured attentional resources. Fundamentally, it becomes a blockage for the visual augmentation design [Akoumianakis and Stephanidis, 2005; Favela et al., 2010; Raja and Calvo, 2017]. Such, overcrowding and occlusion of the critical information by digital elements in the neurosurgical operating room workflow is an apparent barrier to acceptance of AR technology by surgeons [Nguyen et al., 2020]. Similarly, the adoption of AR solutions by students during the learning process appears to be firmly bound to the amount of augmenting visual content and its complexity, the adoption being weaker for more complex augmentation [Wu et al., 2013; Wang, 2017]. Neither the amount of information nor its intensity seems to be the answer to the performance improvement. It is, thus, crucial to explore the "sweet spots" for visual processing when considering augmentation design. Focusing on qualitative rather than quantitative aspects of augmentation, various domains of the visual content attributes should be explored, including its spatial and temporal properties. One direction is to embrace the possibility of efficiently processing the enhanced visual input and integrating it into behavior [Raja and Calvo, 2017]. In doing so, the subtle nature of augmentation content, which considers human visual processing factors, becomes an important milestone toward developing adaptive and not overloading AR systems. The latter refers to the solutions that will ultimately support the strategy chosen by the user in contrast to alternating it in an overruling manner.

To generate guidelines for perceptually optimal augmentation, which yet significantly improves users' performance in everyday life, it is necessary to investigate how manipulation of spatial and temporal properties of visual cues affects human performance. Exploring this question is the central goal of the present dissertation.

1.2 Augmented reality (AR)

1.2.1 Mixed reality taxonomy

The versatility of AR applications gives rise to a diverse set of definitions used across the literature. Originally the taxonomy of augmented reality can be traced back to the "virtuality continuum" introduced in [Milgram and Kishino, 1994]. Among others, alternatively to "augmented reality", such terms as "enhanced reality", "enriched real-

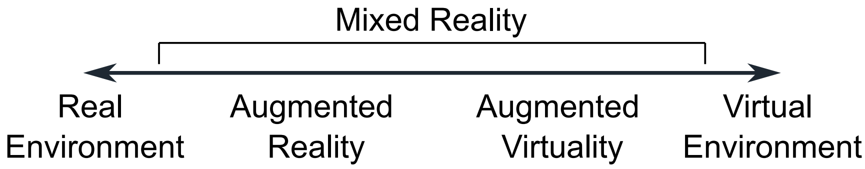

Figure 1.1: Representation of "virtuality continuum" (adapted and adjusted from [Milgram and Kishino, 1994]).

ity", and "mixed reality" are interchangeably used [Iatsyshyn et al., 2020]. Authors define different types of environments based on their level of realism ranging from the utterly virtual environment to the real world (Figure 1.1).

Over the past few decades, the available technology possibilities advanced tremendously, enabling virtual reality (VR) solutions to offer users more immersive experiences. The "virtual environment" is fully synthetic computer-generated content that can be presented on a screen-based setup as well as using a head-mounted display [Cipresso et al., 2018]. One can further increase the level of realism by incorporating some real elements into wholly artificial environments, which can be helpful to simulate interactive scenarios, for example, in a navigational task [Neugebauer et al., 2020]. This approach refers to "augmented virtuality" [Milgram and Kishino, 1994]. The next step towards a more naturalistic setting is "augmented reality, essentially an overlay of computer-generated content on the real surroundings. Similar to VR, AR can be implemented using various technologies ranging from screen-based systems such as the see-through design on a smartphone to the head-mounted see-through displays, all the way to the heads-up displays [Aydındoğan et al., 2021]. The choice of the technology is then defined by a specific application, necessary level of realism, and available resources. The final stage of the "virtuality continuum" in [Milgram and Kishino, 1994] is the "real environment", which naturally refers to the real world surrounding us.

1.2.2 Level of realism and simulation of AR in VR

One of the determining factors for a high level of realism of the user's experience is the set of characteristics of real-world human vision attributed to the technology. These characteristics include the large field of view, stereo vision allowing seeing in-depth, and, importantly, the possibility to freely move the head and gaze, enabling natural exploration of the environment. Advanced development of AR technology offers a high level of realism [Cipresso et al., 2018]. However, it also brings in a diversity of variables when blending the real and digital worlds, be it lighting setting, objects recognition, or position detection. It, in turn, introduces challenges to the researchers where complete control over the experiment is often preferred.

On the other hand, traditional screen-based setups offer a possibility of a great extent of control over the experiment but are short on the level of realism. To research different aspects of AR, one way to overcome these challenges and bridge both solutions is to simulate AR in VR. By doing so, the researcher obtains a nearly complete overview of various parameters of the simulated experimental conditions while keeping the level of vividness strong. Furthermore, a fully digital VR simulation opens possibilities to reproduce many naturalistic scenarios, which can sometimes be challenging to implement in real-world settings due to space constraints or limited resources.

An essential aspect of the realistic and vivid experience is immersion, a quality of technology enabling the user to feel and perceive as if they are present in the virtual world [Georgiou and Kyza, 2017]. An immersive experience is constituted by (1) sensory-motoric immersion, allowing moving within the environment and multisensory feedback, (2) spatial immersion, enabling free exploration of the environment via unconstrained head, body, and gaze movement, (3) emotional immersion, affecting the user and engaging into the story through emotions, and (4) cognitive immersion, triggering the observer to think and act within the virtual environment [Georgiou and Kyza, 2017]. Combining all four components of immersion, VR technology provides an emergent perceptual feeling of interacting, being engaged, and being present in the virtual world. Moreover, as was discussed earlier, an excessive amount of visual information can sometimes be cognitively overloading, leading to a low level of immersion. Thus, the opportunity to simulate subtle visual augmentation in VR where the user can focus on integrative perceptual experience and not be disturbed by the additional content is an inspiring path towards an overall immersive experience.

In the present dissertation, as the initial step, the first study was implemented in a screen-based setting, whereas the remaining two studies were extended to a virtual reality setup.

1.2.3 Ecological approach to AR

Contemporary AR solutions offer an immense variety of applications that can be incorporated into human activities. Besides the consumer entertainment market, AR made its way into the higher complexity sectors spanning from education to engineering and piloting coaching to surgeon training [Carmigniani and Furht, 2011; Mena Vargas et al., 2019]. However, as was mentioned in previous sections, the rapid advancement of technology introduces a problem of visual information excess displayed to the user, which can lead to an overload of human limited processing resources [Raja and Calvo, 2017; Carrasco, 2011]. In [Raja and Calvo, 2017] authors propose a possible solution to this problem: an ecological design of augmentation in AR. In contrast to simply enhancing the natural environment, this approach aims to augment the meaningful features of reality, where no active integration of the digital and actual visual content would be necessary.

The concept of ecological AR (E-AR) is tightly bound to a notion of affordances in ecological psychology introduced by J. Gibson back in 1978 [Gibson, 1978]. According to this notion, we perceive the world as an opportunity to interact with the surrounding environment. Concerning AR, the main idea of the ecological approach is introducing new affordances to the user, which can be perceived only as a blend of reality and augmentation, and not in the virtual or real-world alone. Building upon the concept of E-AR, the idea in the present dissertation is to explore the attributes of augmentation, blended not only in the physical domain but also in perceptual space. Specifically, the work aims to address the perceptual properties of visual augmentation.

1.3 Human "optimal" performance

One of augmented reality solutions' primary functions is to improve user performance. Broadly, it cannot be automatically assumed that every user would initially seek a performance improvement. For example, not all students are equally eager to learn new material in a studying course. Yet, bringing in adaptive supporting solutions such as AR targeting optimizing performance can positively impact the user experience, e.g., on different aspects of the learning process [Wu et al., 2013; Wang, 2017].

The optimality of human performance is, however, a versatile term. In many situations, it is convenient to evaluate how optimal the person's performance is in a specific task by examining objective metrics such as response times, duration of the task completion, accuracy, etc. [e.g., Wolfe, 2020; Farmer et al., 2018]. However, a particular adopted strategy in a given task can be influenced by various factors ranging from the nature and complexity of the task to the personal motivation, value, and experience of the person with the task [Janssen and Brumby, 2015; Mone and Shalley, 1995; Janssen and Brumby, 2010; Howes et al., 2009]. Depending on the research question, it is thus, important to clearly define what is considered an optimal performance in a particular study. Nonetheless, to approximate performance, it is often reasonable to use indicators such as how long it took the person to complete the task, how many fixations were made, or how accurately the task was executed. These and other objective metrics were used in the present work to evaluate possibilities to develop a subtle but efficient augmentation content which would optimize human performance to its higher potential (Figure 1.2).

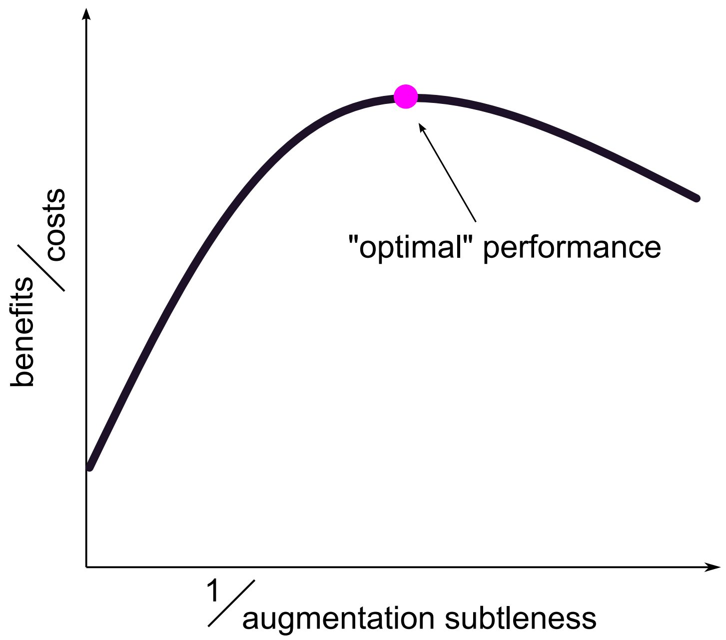

Figure 1.2: The illustration of the concept of a subtle supporting visual augmentation where the human performance is defined by the ratio of the benefits of visual augmentation to its costs. The more prominent (less subtle) the visual augmentation becomes, the higher the benefit-to-cost ratio. However, a very prominent visual augmentation can become overloading for the user leading to a decrease in the benefits-to-costs ratio. An “optimal” performance is attributed to a “sweet spot” for the augmentation, which is supportive for the performance but does not significantly disturb the user perceptually.

1.4 Design principles for augmentation based on human visual processing in challenging tasks

In the present dissertation, various everyday challenging settings were considered where people could benefit from additional support in the form of visual augmentation. While focusing on the perceptual domain, the following questions were addressed: First, can human behavior scale gradually in space and time by systematically varying the properties of central visual cue? Moreover, can we target individual visual cue processing differences by cue properties modulation? Second, can the periphery be exploited by the augmentation without significantly disturbing central visual processing, and if so, how should the augmentation stimulus look like? Furthermore, can peripheral augmentation design be individually tailored to the user, and if yes, what are the limitations of this approach? Finally, can subtle filtering out visual information from the scene be helpful for the user's performance? The following sections discuss the approaches to answering these questions and describe specific examined scenarios.

1.4.1 Posner attentional paradigm: attentional benefits and costs

When developing subtle augmentation which improves users' performance in daily tasks, visual attention is a concept that naturally comes into consideration. To investigate the impact of diverse visual augmenting cues on human performance, it is important to understand how our attention is allocated in different scenarios.

Attention essentially introduces performance differences in the attended vs unattended locations [Carrasco, 2011, 2018; Eckstein, 2011]. In behavioral science, a fundamental way to study attention is the Posner cueing paradigm [Posner, 1980]. In the original Posner paradigm, the reaction time is used as a measure of attention [Posner, 1980]. In the original procedure, participants are seated in front of a screen and are instructed to fixate in the center of the screen (Figure 1.3). Two boxes are positioned at the same eccentricity: one on the left and one on the right from the fixation target. For a short period, a cue is presented on the screen, cueing the observer's attention to one of the two locations. After the cue is removed, a stimulus appears in either of the boxes, and the participant is requested to respond to it as soon as possible. The period of time between the cue and the stimulus onsets is referred to as stimulus onset asynchrony (SOA). An alternative metric to describe the stimulus onset relative to the cue is the inter-stimulus interval (ISI) which is the time between the cue offset and the stimulus onset. The experimental condition when the stimulus appears in the cued location is referred to as the valid condition. In contrast, the invalid condition is when the stimulus is presented in the uncued location. The neutral condition is typically a control condition where none of the two locations was cued.

Multiple paradigm variants emerged since its original version [Carrasco, 2011, 2018]. The cue location can be central or peripheral, targeting endogenous or exogenous attentional shift, respectively [Carrasco, 2018]. Furthermore, the number of possible stimulus locations and their positions can vary and not necessarily be constrained to the left and right side relative to the fixation point [Carrasco, 2011, 2018]. Next, instead of only one stimulus, multiple stimuli can be simultaneously presented in several locations, where the response has to be given only about one of them. Finally, the performance measure can be represented not only by the reaction time but also by other metrics such as, for example, feature discrimination (orientation, contrast) [Carrasco, 2011, 2018; Hayward and Ristic, 2013].

The attentional paradigm and its variations enable comparing performance in the attended and unattended locations. The performance facilitation (e.g., shorter reaction times, higher discrimination level) is typically associated with the ability of the cue to engage attention [Hayward and Ristic, 2013]. Thus, the Posner paradigm and its

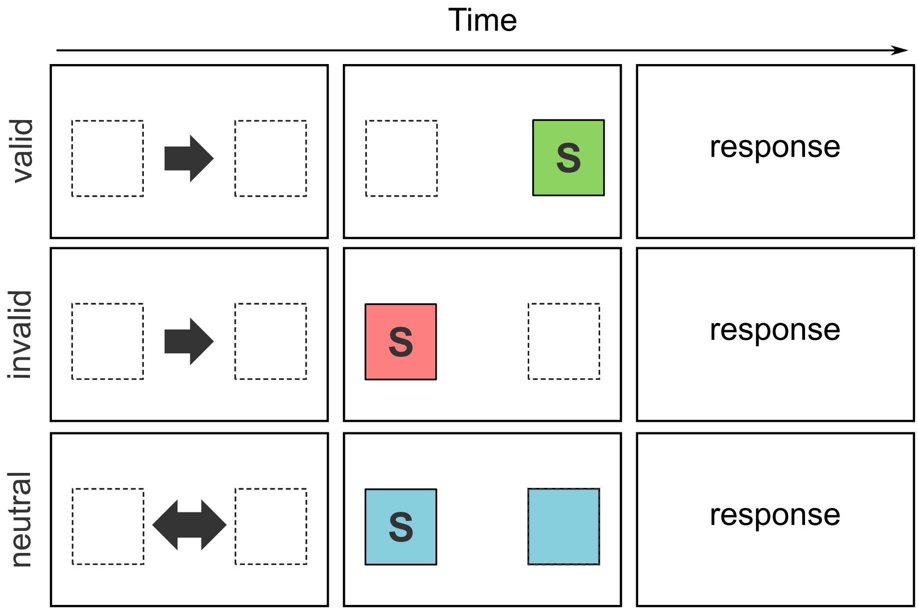

Figure 1.3: Representation of basic Posner attentional paradigm [Posner, 1980]. Here, as an example, a symbolic central cue is shown. During the experiment, first, a cue is displayed, cueing the observer's attention to a specific location. Thereafter, the stimulus appears either in the cued location (valid cueing condition) or in the uncued location (invalid cueing condition). Finally, the observer is requested to respond to the stimulus as soon as possible. The visual performance in the attended location or unattended location can be then compared. A neutral cue is typically exploited for a control condition, where the cue does not guide the observer's attention to any particular location.

variants serve as a potent tool to study the effect of different types of cues on human visual performance and, therefore, evaluate the efficiency of visual cues.

From extensive attentional literature, the evidence shows that when we attend to one locus, our performance mainly improves in the respective site, but at the same time deteriorates in the remaining not attended locations due to limited resources [Posner, 1980; Carrasco, 2011; Eckstein, 2011]. The former case is referred to as the attentional benefit, whereas the latter — as the attentional cost. These attentional costs in unattended locations become a bottleneck when it comes to augmentation, aiming to improve human performance deliberately. Thus, to advance towards a more adaptive and supportive visual augmentation, it is essential to investigate how the modulation of properties of visual cues affects our performance. Can the user's attentional benefits and costs, and thus, behavior, scale gradually with systematically varying cue properties? Moreover, can we customize the cue and address individual cue processing differences through temporal modulation of the cue? The first study of the present dissertation targeted these questions.

1.4.2 Subtle peripheral augmentation in a dual task interleaving scenario

In day-to-day life, humans are naturally confronted with attention distribution between several locations while performing two or more tasks concurrently [Janssen et al., 2015], for instance, when a researcher is working on a manuscript while cooking lunch at the same time. The scenario where, given a limited amount of time, one must attend to several tasks while performing only one task simultaneously is referred to as task interleaving. It has a high prevalence in our society and, consequently, has been researched across various fields [see Janssen et al., 2015]. Essentially, to interleave two tasks, one must distribute the allocated time among tasks at hand, deciding how long to stay on one task before switching to the other and when to return to the first task.

Various areas of people's professional and personal lives are affected by the selected task interleaving strategy, such as, for example, productivity [Janssen et al., 2019; Farmer et al., 2018] and safety [Janssen et al., 2012; Brumby et al., 2009]. Previous research demonstrated that in plenty of dual-task scenarios, people could adjust their task interleaving strategy, resulting in a more efficient performance [Farmer et al., 2018; Janssen and Brumby, 2015; Janssen et al., 2015]. Yet, there are some demanding settings where human performance falls short of its optimal level [Janssen et al., 2019]. A specific adopted strategy of task interleaving can depend on various components, among others are the person's inner motivation, the complexity of the tasks, and the person's experience with the task [Janssen and Brumby, 2015; Mone and Shalley, 1995; Janssen and Brumby, 2010; Howes et al., 2009]. One important constituent impacting humans' decision on how to allocate the time to the given tasks is the payoff function [Janssen et al., 2019; Farmer et al., 2018; Janssen and Gray, 2012]. This function describes how crucial one task is for the person relative to the other task in terms of gain-to-loss ratio. Previously, it has been demonstrated that although humans show some flexibility in their task interleaving strategy, in the case when one of the tasks is significantly more important, people tend to switch to the more crucial task earlier than they could have, demonstrating risk-aversion to fail the more critical task [Farmer et al., 2018]. Naturally, it leads to a performance decrease (Figure 1.4).

In the second study of the present work, we investigated whether the augmentation can employ the periphery to subtly support the user's task interleaving strategy. We examined a broad peripherally processable stimulus with the idea of preventing decreasing performance in the central part of the visual field. We further addressed individual customizing of the augmentation through modulating its temporal properties.

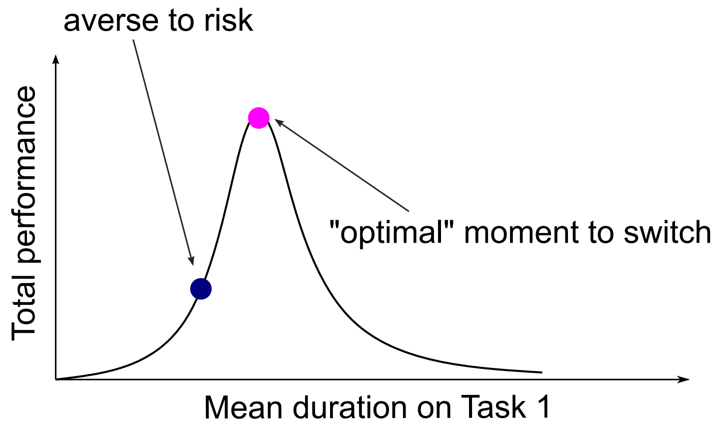

Figure 1.4: Illustration of lower performance due to aversion to risk in task interleaving scenario with unequal payoff function for each task. For simplification, only two parallel tasks are considered: Task 1 and Task 2. Here, Task 2 is considered to be significantly more critical to the user.

1.4.3 Saliency-based augmentation in a visual search task

Another prevalent task in everyday life is visual search. A substantial number of studies addressed people's strategy to search for a target surrounded by distractors, implementing numerous paradigms for the visual search task [for reviews, see Chan and Hayward, 2013; Wolfe, 2020; Verghese, 2001].

Various factors affect the visual search difficulty ranging from the target-background similarity to the scene complexity to whether the scene has been observed by the searcher before. [Chan and Hayward, 2013; Wolfe, 2020; Verghese, 2001]. In visual search research, the complexity of the search is traditionally determined by the evolution of search time as a function of the number of distractors. In easy searches, an increase in the distractors does not significantly impact the search time, indicating a pop-out effect. In contrast, in complex searches, the search time steadily increases the more distractors are present in the search set [Wolfe, 2020]. Particularly in complex searches, due to humans' limited capacity for visual information processing, it is critical to prioritize visual content to complete the search task.

Multiple elements impact visual search strategy: the top-down feature guidance, the object value, the history of the search, the scene guidance, and finally, the bottom-up saliency [Wolfe, 2020]. The top-down control is determined by the internal state of the observer, whereas the bottom-up mechanism is guided by salient external stimuli [Bertleff et al., 2017].

Across the literature, one can meet numerous definitions of saliency. From a broader perspective, a salient region is a part of the scene which highly contrasts with the surrounding areas across one or multiple domains, be it spatial frequency, color or the area's contextual meaning [Schütt et al., 2019a]. Although saliency is often attributed to the bottom-up contrast of features [Itti and Koch, 2000; Bahmani and Wahl, 2016], the two characteristics are tightly connected but not interchangeable. Thus, a location is generally considered salient if it is more probable to attract fixations which is typically associated with attentional shift towards that region [Schütt et al., 2019a].

Numerous studies demonstrated that the observer's attention could be attracted by salient distractors, even though they are uninformative from a goal-oriented perspective [e.g., Barras and Kerzel, 2017]. The extent to which saliency affects human visual performance has been long debated [Schütt et al., 2019a]. In visual search literature, some studies showed that the search strategy is primarily driven by the top-down mechanisms, where initial few fixations can mostly explain the fixations distribution [Chen and Zelinsky, 2006; Henderson et al., 2007; Rothkegel et al., 2019]. Other researchers, however, showed the possibility of salient goal-irrelevant distractors to attract the observer's attention, leading to a slower search [Jung et al., 2019; Bertleff et al., 2017; Foulsham and Underwood, 2011; Theeuwes, 2004]. The substantial variability of the reported results endorses the idea that attentional guidance in visual search is affected by a combination of top-down and bottom-up factors [Wolfe and Horowitz, 2017].

Concerning augmentation design supporting visual search strategy, it is essential to consider how the additional visual information impacts gaze distribution in the scene. One can generally think of a top-down augmentation content based on the observer's internal state, and a bottom-up augmentation built in a stimulus-driven manner based on the visual scene features [Awh et al., 2012]. The bottom-up approach was taken in the third study of the present dissertation. Specifically, we investigated whether a subtle saliency-aware visual augmentation can be applied to guide the gaze and facilitate the visual search while minimally altering the scene information. A filtering-out method was applied, examining the possibility of subtly reducing the saliency of some irrelevant salient regions and, by doing so, helping the user find the target more efficiently.

2. Objectives

Through rapid technological advancements such as VR and AR systems, diverse visual cues demonstrate a powerful potential to deliberately guide attention and improve users' performance in daily tasks. Currently, existing solutions are confronting the challenge of overloading and overruling the natural strategy of the user with excessive visual information once digital content is superimposed on the real-world environment. The subtle nature of augmentation content, which considers human visual processing factors, is an essential milestone towards developing adaptive, supportive, and not overwhelming AR systems. The focus of the present dissertation was, thus, to investigate how manipulation of spatial and temporal properties of visual cues affects human performance.

The first study investigated the impact of spatial cue complexity on attentional benefits and costs. To do so, in a screen-based setting, an adapted attentional paradigm was implemented where the complexity of a fully discriminable visual cue was manipulated. The discriminability of the shape of the cue was used as a proxy for the cue complexity, where a set of cues with systematically manipulated complexity was generated. The obtained results made it possible to assess the gradual modulation of benefits and costs in the attended and unattended locations, respectively. Furthermore, the effect of individual cue processing temporal differences was evaluated.

The second study examined another challenging scenario, where a division of attention to multiple locations is necessary — task interleaving. A more realistic scenario was simulated where two parallel tasks were implemented in a dynamic interactive 3D environment. The impact of a subtle peripheral progress indicator on the user's performance was evaluated. The study's results allowed us to probe the usability of a broad peripheral time-bound augmentation in a challenging task-interleaving scenario when accurate time estimation is necessary. Furthermore, the limitations of generalization of the temporal characteristic of the augmentation, namely, time-lead before the optimal task-switch moment with respect to the individual response times were assessed.

Finally, the focus of the third project was to evaluate the potential of a saliency-aware subtle augmentation in a visual search task. Specifically, the possibility of gaze guidance via subtle blurring of salient regions to support the observer's more efficient search strategy was assessed. The paradigm was implemented using real-world scenes in a 3D virtual environment. The study's results provided an insight into potential visual augmentation design aiming to improve users' performance in everyday visual search tasks.

3. The impact of shape-based cue discriminability on attentional performance

Lukashova-Sanz, O., S. A. Wallis, T., Wahl, S., and Rifai, K. (2021). The impact of shape-based cue discriminability on attentional performance. *Vision*, 5(2):18

3.1 Abstract

With rapidly developing technology, visual cues became a powerful tool for deliberate guiding of attention and affecting human performance. Using cues to manipulate attention introduces a trade-off between increased performance in cued, and decreased in not cued locations. For higher efficacy of visual cues designed to purposely direct user's attention, it is important to know how manipulation of cue properties affects attention. In this verification study, we addressed how varying cue complexity impacts the allocation of spatial endogenous covert attention in space and time. To gradually vary cue complexity, the discriminability of the cue was systematically modulated using a shape-based design. Performance was compared in attended and unattended locations in an orientation-discrimination task. We evaluated additional temporal costs due to processing of a more complex cue by comparing performance at two different inter-stimulus-intervals. From preliminary data, attention scaled with cue discriminability, even for supra-threshold cue discriminability. Furthermore, individual cue processing times partly impacted performance for the most complex, but not simpler cues. We conclude that, first, cue complexity expressed by discriminability modulates endogenous covert attention at supra-threshold cue discriminability levels, with increasing benefits and decreasing costs; second, it is important to consider the temporal processing costs of complex visual cues.

3.2 Introduction

On a regular day, our attention is guided by multiple visual cues, whether we are aware of them or not. In many contexts purposely designed visual cues can influence human behavior through attentional guidance. To name a few, it can be when browsing on a web page of an online shop with a user-friendly design or driving from work to home using navigational assistance. With rapidly developing technology, visual cues become a powerful tool to deliberately guide human's attention [e.g. Madsen et al., 2013; Awan, 2021]. A systematic design of visual cue properties is then necessary for intended guidance. Thus, it is of importance to know how manipulation of cue properties affects attention. From extensive research, it is well known that when we attend to one location, our visual performance generally enhances in that region, however, it decreases in other not attended locations due to limited processing capacity [Carrasco, 2011; Eckstein, 2011; Carrasco et al., 2002, 2006; Pestilli and Carrasco, 2005; Montagna et al., 2009; Abrams et al., 2010; Liu et al., 2009]. The attentional costs in the unattended locations become essentially a bottleneck for the design of visual cues aiming to deliberately improve human's performance [Lu et al., 2012; Orlosky et al., 2019]. Considering such applications as signage design or augmented reality, a potential benefit of rather subtle visual aids is emerging in contrast to strong conventional symbolic cue stimuli, to improve performance where it is intended to be improved, but not significantly decrease the performance in the remaining tasks [Raja and Calvo, 2017]. A gradual manipulation of cue impact would allow a very controlled attentional guidance through visual spatial cueing. Previously, visual cues have been manipulated parametrically where some studies reported gradual change in visual performance when varying cue properties such as its size or contrast [e.g. Müller et al., 2003; Fuller et al., 2009], whereas others argued for lack of flexibility of attentional allocation [e.g. Yeshurun and Carrasco, 2008]. In the present verification study, we intended to systematically modulate attentional benefits and costs via the spatial design of the cue, namely, its shape. But, increasing the cue design complexity aiming to improve the benefits and costs relation, generally might increase the time to process the cue. In the present study, we examine if there are additional temporal costs in performance due to the processing of more complex cues.

One way to facilitate an attentional shift in a classical attentional paradigm is to precue a target location with a spatial cue preceding the target stimulus onset [Posner, 1980]. Numerous studies investigated the modulation of attention deployment when varying the target, distractors, or the task in an array of attentional paradigms [see Wolfe and Horowitz, 2004; Carrasco, 2018]. Several studies have examined the modulation of the properties of the cue itself [e.g., Folk et al., 1992; Bertleff et al., 2017; Goldsmith and Yeari, 2012; Fuller et al., 2009]. Regarding a gradual modulation of visual cue properties, previously it was shown that attentional benefits and costs scaled with the contrast of an exogenous spatial cue indicating a gradual prioritizing of attentional resources allocation for the exogenous attention [Fuller et al., 2009]. In many circumstances, our attention is guided by endogenous visual cues, which in turn, can vary in complexity [Brignani et al., 2009; Takahashi and Watanabe, 2013; Trachel et al., 2015].

In the present verification study, we address how modulation of cue complexity affects attentional performance. In the context of this work, the cue discriminability was used as a proxy of the cue complexity. The more difficult it is to discriminate the cueing direction, the more complex the cue becomes. It is investigated, how manipulation of cue discriminability of an endogenous spatial cue, even above discrimination threshold, modulates endogenous covert attention. Specifically, in an orientation-discrimination two-alternative-forced choice (2AFC) task it is studied how the discriminability of an endogenous cue impacts attentional performance in the cued and not cued locations. A shape-based spatial cue was designed where through a systematic manipulation of the shape of the cue a set of cues of various cue discriminability levels was generated, including supra-threshold levels of cue discriminability. Performance in the valid, invalid, and neutral conditions was assessed. By doing so, the benefits and costs of the allocated attention were quantified. We hypothesize that if the cue discriminability of an endogenous cue is increased, the benefits and costs of the cue will also grow in magnitude across the cue discriminability even when the cued direction is fully discriminated.

Less discriminable cues, even at discriminability level over the sensitivity threshold, generally might require more time to process. This can influence the time course of an attentional effect [Brignani et al., 2009]. From an extensive body of literature it is known that endogenous attention has a specific time span approaching its maximum at ~300 ms after the cue onset [e.g., John Jonides, 1980; Müller and Rabbitt, 1989; Cheal and Lyon, 1991; Remington and Pierce, 1984; Liu et al., 2007]. In the present verification study, we intended to evaluate temporal delays of attentional shift through cue processing. Specifically, to decode a less discriminable cue, above the discrimination threshold, one would need a longer time, thus, attentional performance can benefit from a longer period between the cue offset and stimuli onset, i.e. inter-stimulus interval. The inter-stimulus interval is linked to another common temporal measure in attentional paradigms, stimulus-onset asynchrony (SOA) where the latter is the time between the cue and stimulus onsets. In the present study, to address the temporal effect of the cue discriminability on attentional performance, conditions with different inter-stimulus intervals were compared.

3.3 Methods

3.3.1 Subjects

In this verification study, 11 naive participants with an average age of 24.4±3.7 years were tested. Eight participants were female, and three participants were male. All procedures conformed to Standard 8 of the American Psychological Association's "Ethical Principles of Psychologists and Code of Conduct (2010)". The study was approved by the ethics committee of the Faculty of Medicine of the University of Tübingen. Signed informed consent was obtained from each participant prior to the measurements.

3.3.2 Apparatus and stimuli

3.3.2.1 Apparatus

The visual stimuli were presented on a VIEWPixx monitor (VPixx Technologies Inc., Montreal, Canada) with a refresh rate of 60 Hz, resolution of 1920×1080 pixels, at a distance of 50 cm from the monitor. The paradigm was generated using PsychoPy library [Peirce et al., 2019] and an additional Python package for integrating eye-tracking, PyGaze [Dalmaijer et al., 2014], running on a Windows 10 PC. All experiments were performed in a dark room. To ensure covert attentional shift and exclude the possibility of saccades towards the target stimuli, all participants were instructed to maintain their fixation at the central location throughout the experiment. The fixation was monitored using EyeLink 1000 Plus (SR Research, Ottawa, Canada) eye tracker. In case the participant looked away from a defined area of 2.5° of diameter during the stimuli display, the respective trial was withdrawn, and a new trial started. The head was stabilized with a chin rest. Responses of participants were registered with an external keyboard.

3.3.2.2 Shape-based spatial cue

An intuitive choice of an endogenous cue is an arrow pointing to a certain direction. Overlearning of this stimulus, however, limits the possibility to gradually modulate the cue property [e.g. Brignani et al., 2009; Guzzon et al., 2010]. To avoid a binary attention allocation, we propose a novel cue design where a parametric shape manipulation enables progressive modulation of cue discriminability.

As an endogenous shape-based direction-indicating cue a filled irregular radial frequency pattern was used. Sinusoidal modulation of the radius R at polar angle θ in a radial frequency pattern is generally described by Eq. 3.1 [Wilkinson et al., 1998]:

$$R(\theta) = R_0(1 + A \cdot \sin(\omega\theta + \phi)), \tag{3.1}$$

where R_0 is the mean radius, A is the radial modulation amplitude, ω is the radial frequency and ϕ is the angular phase of the pattern. In an irregular radial frequency pattern, the radial frequency varies between different lobes. In this study, filled irregular radial frequency patterns with six cycles of sinusoidal modulation were used. For details on generating the cue please refer to the Supplementary Information Section S1 and Figure S1. To indicate cueing direction the radial frequency of one of the lobes of the pattern was modulated in order to make it larger compared to the rest of the five lobes. In particular, to identify the cued direction participants were to compare areas covered by the right and left lobes: the largest of two indicated the cueing direction, left or right. The ratio between the modulation factors for the radial frequency of two relevant lobes determined cue discriminability level: the smaller is the radial frequency at a given amplitude and mean radius of the cueing lobe, the larger is the cueing lobe compared to the opposite lobe, thus, the more discriminable is the cue. When systematically modulating the radial frequency of one lobe, the radial frequency of the opposite lobe was kept constant and the remaining four lobes were adjusted such that the total filled area of the cue stayed unchanged. The ratio of the modulation factor of the cueing lobe and the opposite lobe was systematically varied from 0.98 down to 0.68 with a step of 0.02, corresponding to the least and the most discriminable cues, respectively. The angular phase was adjusted for each cue such, that the cueing and the opposite lobes are oriented horizontally. As a neutral cue, a two-lobes regular radial frequency pattern was used, lobes' amplitude of which was adjusted to keep the filled area equal to that of the irregular patterns. In Fig. 3.1 the examples of cues used in the experiments are demonstrated. Essentially the arbitrary units of cue discriminability are determined by the modulating factor of the radial frequency for the cueing lobe. The values of 0.96, 0.84, and 0.68 were selected as the modulating factor ratios for the cues used in the main experiment. These values were selected based on a cue sensitivity test conducted before the main experiment (see details in Section 3.3.3.2 and Section 3.3.4.1). In the context of this study, the discriminability levels of these cues are respectively denoted by CD1, CD2, and CD3, in arbitrary units.

The type of the cue is defined to be endogenous following the definitions of voluntary attentional shift induced by a central cue [see Carrasco, 2011]. Furthermore, to some extent the designed shape-based cue shares some cueing properties with a conventional arrow cue such as pointing towards a certain direction. Although previously it was demonstrated, that arrows induce a fast attentional shift which can be debated as automatic stimulus-driven, the neurophysiological evidence suggests that this fast shift is due to an overlearned association mechanism rather than exogenous attentional process [e.g. Brignani et al., 2009; Guzzon et al., 2010]. Due to a novel cue design in the present study, an overlearning is not to be expected. Thus, the cue in the present study is considered to be endogenous.

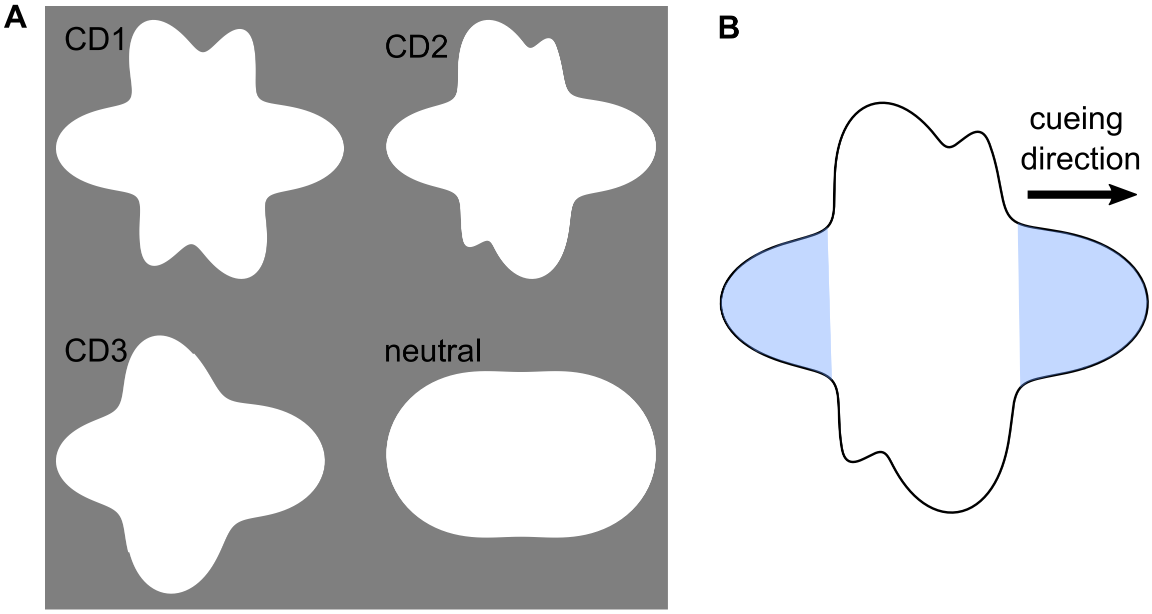

Figure 3.1: The shape-based cue construction (**A**) at different cue discriminability levels: low (CD1), medium (CD2) and high (CD3), together with the neutral cue; (**B**) Schematic representation of cue direction identification. The "largest" lobe of the left and right lobes, estimated by the area of the lobe, indicates cueing direction. In this figure, as an example, all presented shape-based cues indicate the direction to the right. In the experiment, the right and the left directions were cued equally often.

3.3.2.3 Stimuli

As stimuli two tilted Gabor patches were used of spatial frequency 2 cpd, visual angle 2° and Michelson contrast 50 % located at 4° eccentricity left and right from the central fixation cross. The magnitude of the tilt angle of the Gabor patches tested in the experiment was fixed to a set of five different values: 0.5°, 1°, 3°, 5° and 12° relative to the vertical. The magnitude of the tilt angle and the tilt direction of two stimuli simultaneously presented in each trial were independent of each other.

3.3.3 Experimental procedure

3.3.3.1 General procedure

The study consisted of three sessions, each performed on different days. The first session consisted of three experiments: a cue sensitivity test with the shape-based cue, a reference cue sensitivity test with a line cue, and a practice session for the main experiment. During the second and the third sessions on two different days participants performed the main experiment with two different inter-stimulus-intervals, respectively. The inter-stimulus-interval was fixed for a given experimental session. Each experimental session lasted approximately one hour where participants performed the experimental task for a maximum of 15 min in a row. In particular, the second and third sessions were split into three sub-15 min sessions with a short break in between to let the participants rest. During breaks, participants remained in the same room under unchanged lighting conditions. For each participant, the number of days between the experimental sessions varied from three to seven days.

3.3.3.2 Cue sensitivity test

Before the beginning of the first session cues of various cue discriminability were demonstrated and explained to the participants. The participants were allowed to familiarize themselves with the cues. In the cue sensitivity test participants were presented with a spatial cue for 150 ms indicating one of the two locations, left or right of the central fixation cross. After the cue display participants were instructed to report the perceived cue direction with a keypress as fast as possible upon the cue offset. Therewith the response times for different cues for each participant were registered. When a neutral cue was presented, participants were asked to press either of the two response keys. The neutral cue differed from the informative cues by its symmetric design to avoid its confusion with a very subtle informative cue. In the cue sensitivity test with shape-based spatial cues, participants were presented with cues of 16 different cue discriminability levels. Each cue discriminability level was presented 20 times randomly alternating between different cue discriminability levels and the neutral cue, as well as in random order of left and right cueing directions. In addition to the cue sensitivity test with a shape-based cue, each participant performed a separate line cue sensitivity test where the spatial cue was a black line of size 0.7° and the neutral cue was a black dot of radius 0.2°. The line cue spatially cued the participant either towards the left or the right location, depending on which side from the center it appeared. The line cue test was performed directly after the shape-based cue sensitivity test with a short 3-5 minutes break. During the break the participant remained in the same room and same lighting conditions.

In total each participant performed (16 cue discriminability levels +1 neutral cue) $\times$ 20 trials resulting in 340 trials during the cue sensitivity test with shaped-based cue, and (1 line cue +1 neutral cue) $\times$ 20 trials resulting in 40 trials during the line cue sensitivity test.

3.3.3.3 Training session

The purpose of the practice session was for the participants to familiarize themselves with the procedure of the main experiment. The data obtained during the practice session was not used for further analysis. The experimental flow of the practice session was identical to the main experiment, only, instead of varying target stimuli tilt angles, a fixed large angle of 20° for training purpose was used. The practice session consisted of 48 trials.

3.3.3.4 Main experiment

In the main experiment participants performed a direction-discrimination 2-alternative-forced-choice (2AFC) task in a modified Posner attentional paradigm (Fig. 3.2). After a variable period of fixation between 350 ms and 450 ms, a shape-based direction-indicating cue was presented for 150 ms. Thereafter, participants shifted their attention to the cued location while fixating in the center. Next, after an inter-stimulus interval, two stimuli were presented simultaneously for 40 ms: one on the left and one on the right side of the center (see the detailed description of the stimuli in Section 3.3.2.3). Directly after the offset of the stimuli, a white line of 1.5 degree visual angle (the response-cue) was presented at fixation position indicating the final target stimulus location. The approach to use a response-cue to collect the responses in attended and unattended locations was previously used [e.g. Pestilli and Carrasco, 2005]. At the end of each trial, participants had to report the orientation of the target stimulus. The response cue disappeared upon the participant's response. The participant reported the tilt orientation of the target, clockwise (CW) or counterclockwise (CCW), with a keypress right or left, respectively. Two different inter-stimulus-intervals were used: 150 ms and 250 ms. In the context of this study, as the cue display duration was fixed, the two experimental conditions are defined by the inter-stimulus-interval instead of stimulus onset asynchrony. The smaller ISI 150 ms was selected based on the time period ~300 ms after the onset of the cue when endogenous attention typically approaches its maximum, as also described in the Introduction [e.g., John Jonides, 1980; Müller and Rabbitt, 1989; Cheal and Lyon, 1991; Remington and Pierce, 1984; Liu et al., 2007]. In the present study the idea was to test how benefits and costs evolve with systematic cue modulation, but at the same time monitor the potential temporal processing disadvantage of the cues. Considering the novel design of the cue, in addition to the conventional ISI, a second longer ISI 250 ms was selected. The time interval between the stimuli offset and the response-cue onset was set randomly between 700 ms and 800 ms. To ensure that participants shifted attention to the cued location, the number of valid trials at a given cue discriminability level was set twice as large as the number of invalid trials. In addition, a quarter of the total amount of trials at a given cue discriminability level was neutral trials. At a given cue discriminability level and target stimulus tilt angle, there were collected 6 measurement points for the invalid, 12 for the valid, and 6 for the neutral conditions. This limited number of measurements per data point was chosen due to lengthy and focus-demanding exper-

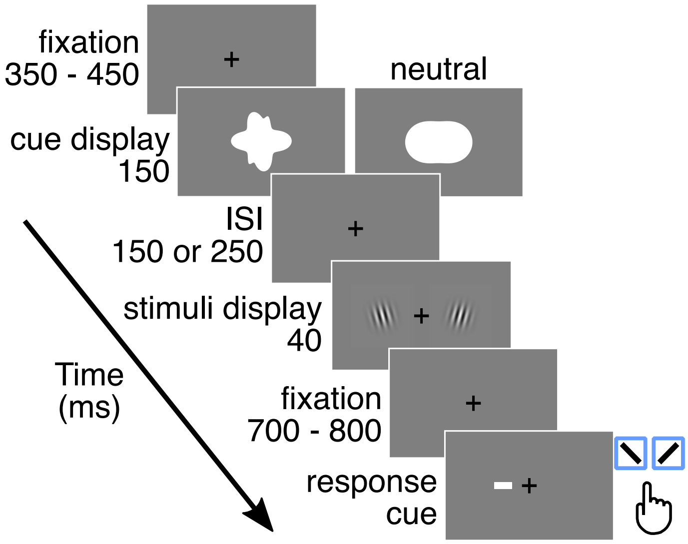

Figure 3.2: Experimental flow of the main experiment.

imental sessions. The validity of the cue was, thus, 67%, which have been previously used [e.g. Giordano et al., 2009]. In each of two sessions during the main experiment, each participant performed 720 trials resulting in 1440 trials in total.

3.3.4 Analysis

3.3.4.1 Cue sensitivity test

From the cue sensitivity test with a shape-based spatial cue for each participant, the proportion of correct responses as a function of cue discriminability was obtained. The data were fitted with the Weibull psychometric function with a chance level of 50% resulting in a cue sensitivity curve for each participant. The proportion correct values at a given cue discriminability were averaged over all participants. The present study was aimed to test a possible scaling of attentional performance with cue discriminability. Given that the task was rather challenging for the participants, the possible duration of the experimental session was constrained by the fatigue of the participants. Therefore, a set of three cue discriminability values was selected for the main experiment as a minimum set size to evaluate the gradual effect. Furthermore, although the novel cue design enabled an incrementing cue sensitivity supported by the cue sensitivity test, it also had some constraints in terms of the maximum cue discriminability possible. For the main experiment, the cue with CD3 – the maximum allowed by the cue design – was selected. The second value of cue discriminability CD2 was selected approaching the saturation point of the cue sensitivity curve. Both CD2 and CD3 correspond to a high cue sensitivity represented by proportion correct ∼95%, nonetheless, they differ based on the relative horizontal lobes size. Finally, the third value of CD1 was selected to complete a set of three cue discriminability levels. The value was chosen to be different from CD2 and CD3 but still to be significantly over the chance level with respect to proportion correct in the cue sensitivity test.

3.3.4.2 Main experiment

In the orientation-discrimination task for each participant, a ratio of CW responses to the total amount of responses was determined at a given target stimulus tilt angle. Value 1 of this ratio corresponds to all responses at a given target stimulus tilt angle being CW, and value 0 — to all responses at a given target stimulus tilt angle being CCW. The CW tilt angles relative to the vertical are denoted with positive values, whereas the CCW — with negative values. Stimulus response curves were analyzed separately for a given inter-stimulus interval, cue discriminability, and validity. The proportion of CW responses as a function of the target stimulus tilt angle was fitted with a psychometric function fixing asymptotes to 0 and 1. The slope of the psychometric function at the point of subjective equality (PSE) was used as a measurement of performance in attended and unattended locations, corresponding to the valid and invalid conditions, respectively. To evaluate attentional benefits (valid) and costs (invalid) of the cues, the slopes in the valid and invalid conditions were normalized to the slope in the neutral condition and referred further in the manuscript as just slopes. An impact of cue discriminability on attentional performance was estimated from the difference in slope in the valid and invalid conditions. The main effects as well as the interaction between two factors, cue discriminability, and inter-stimulus interval, were then statistically evaluated by performing a two-way repeated-measures analysis of variance (ANOVA) with valid-invalid slope difference as the dependent variable, followed by a post hoc analysis.

3.3.4.3 Cue processing temporal cost

To estimate an effect of cue processing on attentional performance, the difference in performance at two different inter-stimulus intervals was statistically evaluated within the repeated-measures ANOVA. Furthermore, the correlation between individual cue response times and performance in the valid/invalid conditions at two different inter-stimulus intervals was assessed. We assumed the total response time to be a sum of a "baseline" response time and a specific cue processing time. The "baseline" response time is considered here to be the response time to a line cue, a widely used simple endogenous visual cue [Carrasco, 2011]. Therefore, as an approximation of the specific shape-based cue processing time, the response time of a line cue for each participant was subtracted from the total response time of the corresponding shape-based cue.

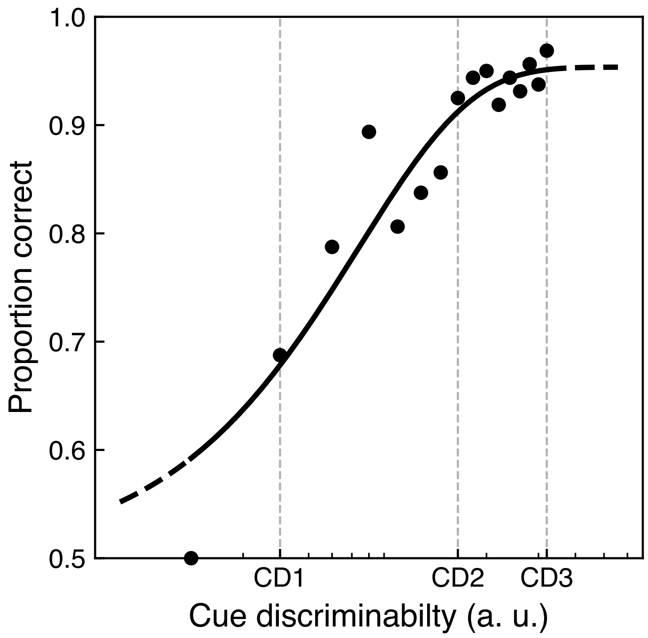

Figure 3.3: Results of cue sensitivity test with a shape-based cue: mean proportion correct averaged over all subjects as a function of the cue discriminability. The grid lines correspond to three cue discriminability levels in arbitrary units selected for the main experiment.

The resulting processing time was then correlated with the performance difference in the sessions with different inter-stimulus intervals. The performance difference was estimated by a difference in slope at ISI 250 ms and ISI 150 ms. A positive correlation between reaction time of slope difference would denote that participants with longer cue processing performed better in the session with a longer inter-stimulus interval. Pearson correlations were computed for the three cue discriminability levels separately at a given validity.

3.4 Results

3.4.1 Cue sensitivity

Fig. 3.3 shows the mean proportion correct values averaged over all participants as a function of cue discriminability level. The proportion correct increases with stimulus discriminability level. Based on these data three cue discriminability levels, CD1, CD2, and CD3 were selected. The individual response times as a function of the cue discriminability levels are shown in Fig. 3.6 in Section 3.4.3 where it is more relevant.

3.4.2 Main Experiment

The individual fits for all measured participants can be found in Supplementary Information Figure S3–S24. Out of 11 participants, three participants were excluded (see Supplementary Information Figure S9–S10 and Figure S21–S24). For two of the excluded participants, the task appeared to be too difficult which was indicated by the feedback from participants as well as by consistently high noise level of the fitted psychometric functions for all experimental conditions. Another participant explicitly reported by himself that he did not follow instructions of the task as suggested, specifically, he did not use the cue. The data sets collected from 8 participants were used for the final analysis. The goodness of the fit can be described by deviance extracted from the fitting models (see Supplementary Information Figure S2). The distribution of the deviance was originally planned to be used as an objective exclusion criterion. It largely varies across the validity which is expected as the invalid condition is particularly challenging. It also has fewer measured data points than valid or neutral conditions to keep the valid-invalid trials ratio sufficient for the cue to remain informative, which led to higher noise levels in the invalid condition. Given the difficulty of the experimental design, a big number of fits per participant, and the limitation in the duration of the experiment to avoid participants' fatigue, it was rather challenging to strictly apply the objective exclusion criterion. For future studies, a larger sample size, as well as more measurement points per participant, could improve the goodness of the fits and enable following objective guidelines to exclude participants. For the purpose of this verification study to obtain preliminary data on cue discriminability effect on attentional performance, the above-described data set was used, where a rather subjective exclusion criterion was applied [1].

Figure 3.4 shows response curves of one example participant, sorted in the valid, invalid and neutral conditions. Three cue discriminability levels are shown in three subgraphs. Different colors of the curves correspond to different validity conditions. In this example, as well as for most of the participants, for the CD2 and CD3 cue discriminability levels, the invalid curve is more shallow compared to the valid and neutral curves, whereas the neutral curve lies in between the valid and invalid. Thus, slopes tend to be steeper in the valid condition, compared to the invalid condition. This trend becomes more pronounced with increasing cue discriminability level which serves as a hint that the cue discriminability affects the performance.

[1] It is worth mentioning that when the full participants set counting also excluded participants was tested, statistical outcome did not change notably. In articular, there was a significant effect of cue discriminability between slopes in the valid and invalid conditions ($F(2, 20) = 8.53, p < 0.01, \eta_p^2 = 0.46$). The effect of inter-stimulus interval was not significant ($F(1, 10) = 3.93, p = 0.08, \eta_p^2 = 0.28$). The interaction between cue discriminability and ISI was not significant ($F(2, 20) = 1.48, p = 0.25, \eta_p^2 = 0.13$).

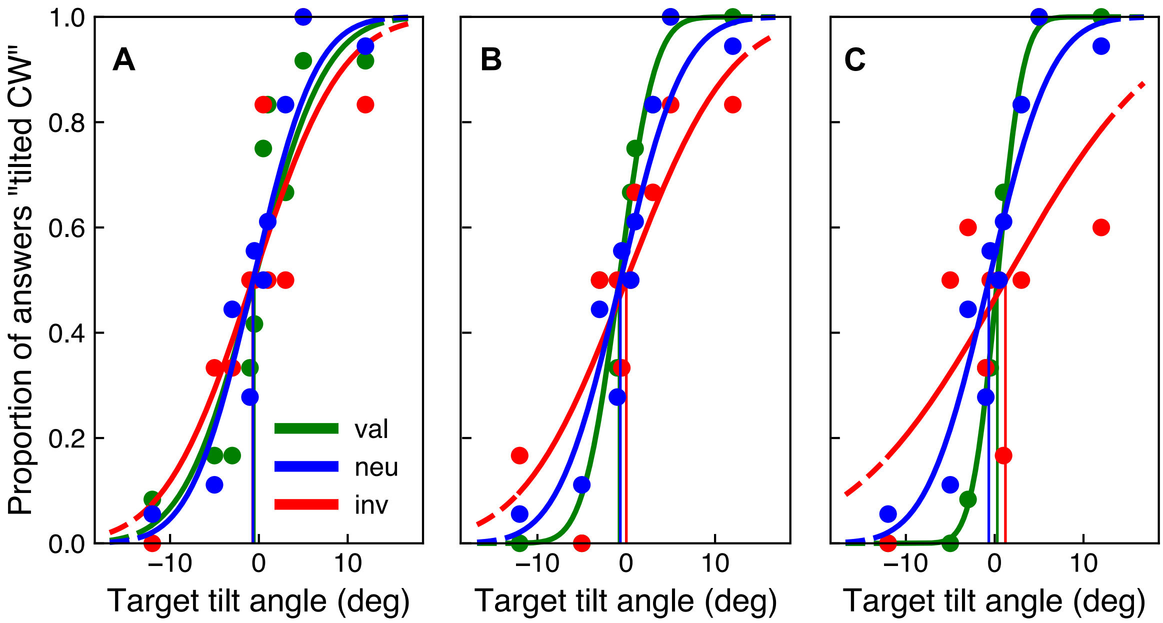

Figure 3.4: An example of a family of curves obtained from one session (ISI 150 ms) for one participant. The proportion of answers "tilted CW" as a function of the target stimulus tilt angle is fitted with a psychometric function. Green, blue and red curves correspond to the valid, neutral, and invalid conditions, respectively. The data sets correspond to cue discriminability levels (**A**) CD1, (**B**) CD2, and (**C**) CD3. Note that data points for some tilt angles are superimposed - all tilt angles were equally shown to the participants.

In Figure 3.5 the mean results overlaid with the individual data are depicted. The slopes were normalized by subtracting the slope in the neutral condition. Thus, the neutral level depicted in Figure 3.5 is independent of the cue discriminability level and lies at zero. The group data with unnormalized data in the neutral, valid and invalid conditions can be found in the Supplementary Information.

From the two-way repeated-measures ANOVA, first, we found a significant effect of cue discriminability on the difference between slopes in the valid and invalid conditions ($F(2, 14) = 7.62, p < 0.01, \eta_p^2 = 0.52$). To evaluate which subgroups of measurements are different we performed a post hoc analysis which revealed for ISI 150 ms a significant difference between the slope difference at cue discriminability levels CD1 and CD2. ($p < 0.05$), and for ISI 250 ms a significant difference between slope difference at CD1 and CD3. ($p < 0.05$). The difference between the slope difference at CD2 and CD3 levels for ISI 250 ms was close to but not significant ($p = 0.14$). The effect of cue discriminability suggests that an easier more discriminable cue has a stronger attentional effect than a more complex less discriminable cue. This result suggests the impact of a shape-based cue discriminability on attentional performance.

The effect of inter-stimulus interval is not significant ($F(1, 7) = 4.81, p = 0.06, \eta_p^2 = 0.41$). The interaction between two factors, the cue discriminability and inter-stimulus

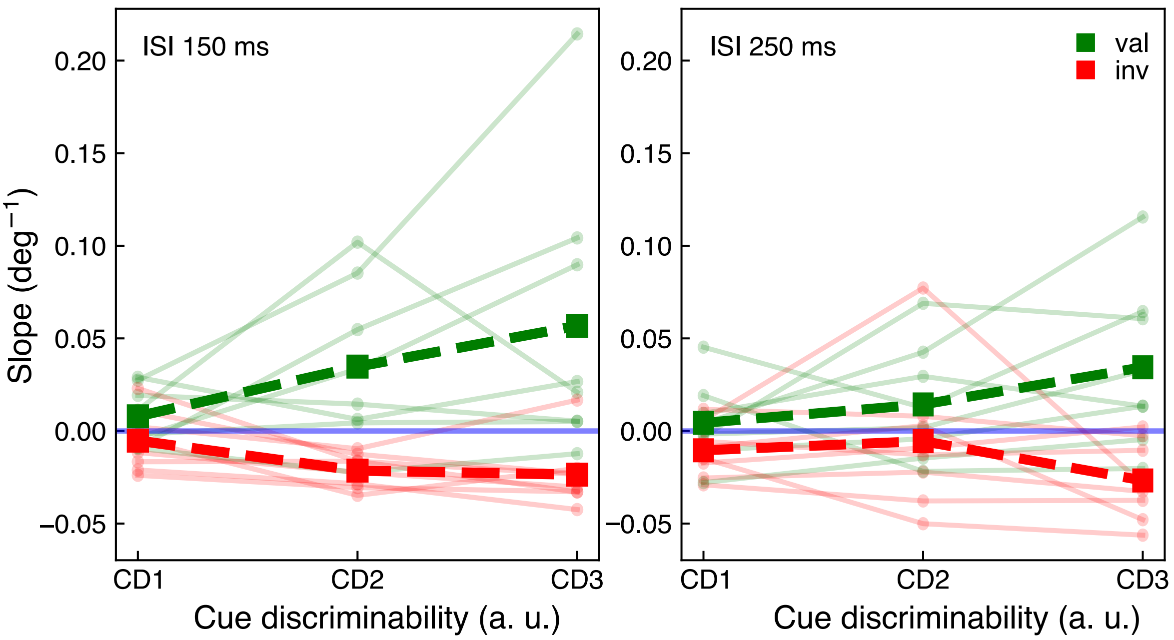

Figure 3.5: The average slope of response curves over all participants (dashed lines) for valid (green), invalid (red), and neutral (blue) conditions at different cue discriminability levels, together with single-participant data (continuous semi-transparent lines).

interval, is also not significant ($F(2, 14) = 1.01, p = 0.39, \eta_p^2 = 0.13$). The absence of a significant difference in performance at different inter-stimulus intervals might originate from an individual variation in cue processing times obtained from the cue sensitivity test. The difference in performance at distinct inter-stimulus intervals might be influenced by individual variances between participants. In Section 3.4.3 the correlation of the individual performance at different inter-stimulus intervals with the cue processing time was evaluated.

The effect size of the cue discriminability represented by partial eta squared is large according to the Cohen's rule of thumb [Pierce et al., 2004]. Note, however, that the large effect size is also influenced by a small sample size tested in the present study. Beyond the scope of this verification study, future work with extended number of participants is necessary to target more representative effect size.

3.4.3 Cue processing temporal costs

In Figure 3.6 individual response times as a function of the cue discriminability levels are shown fitted with a linear function. The response times for the neutral and the line cues for each participant are also depicted in Figure 3.6. As expected, the general trend for all the participants is a decreasing response time with increasing cue discriminability. The individual response times range from approximately 300 ms to 700 ms,

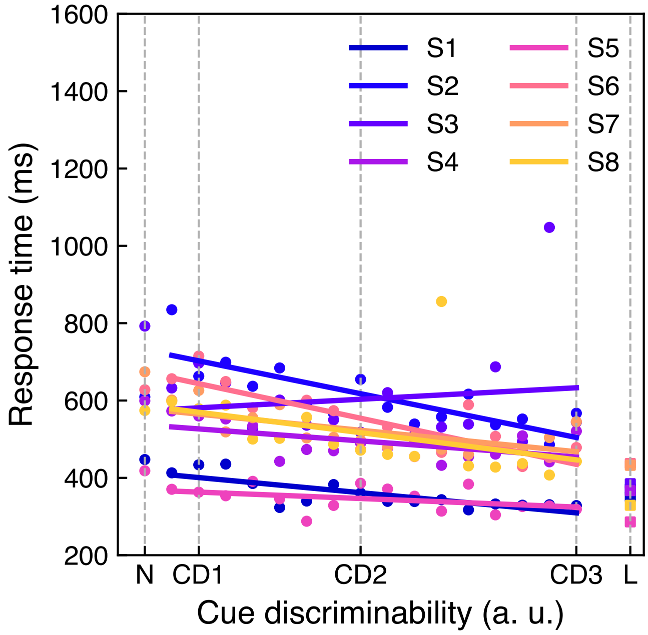

Figure 3.6: The individual response time recorded in cue sensitivity test as a function of cue discriminability. The N tick label corresponds to the neutral cue. On the right, the line cue response times are depicted, denoted by a tick label L. The straight solid lines are the linear fits of the measurement points for each participant. Different colors correspond to different participants.

and the line cue response times lie between 250 ms and 400 ms. By subtracting the line cue response time as a baseline from the response time at a given cue discriminability, the corresponding shape-based cue processing time was determined. Analyzing the wide range of cue processing times varying among participants we suggest that some participants were "faster" and some were "slower" in cue processing. In particular, from the analysis of individual slopes at different inter-stimulus intervals, for some participants performance was better (the slope was larger) at ISI 250 ms than at ISI 150 ms at all tested cue discriminability levels, whereas for some other participants it was vice versa. This suggests that some participants benefited from a longer interstimulus interval, whereas the others did not. The higher reaction times measured for the neutral cue are suggested to be caused by a decision process, as the participant knew that the cue was not cueing any specific direction, but the response still had to be made.

Figure 3.7 shows the difference between individual slopes at ISI 250 ms and ISI 150 ms as a function of cue processing time at a given cue discriminability and validity. Smaller abscissa values correspond to "faster" participants, that is, participants with shorter cue processing times and larger abscissa values correspond to "slower" participants with longer cue processing times. Negative values on the y-axis represent a

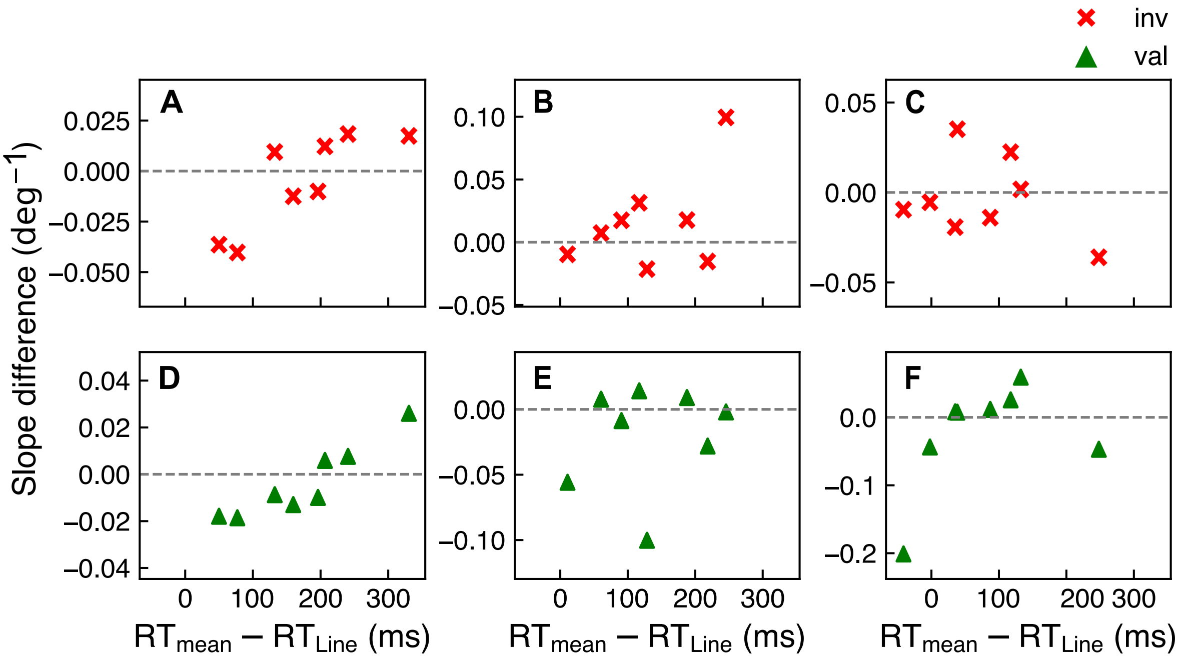

Figure 3.7: The difference between individual slopes at ISI 250 ms and ISI 150 ms as a function of the cue response time normalized to the line cue response time at a given cue discriminability and validity. Red crosses and green triangles represent invalid and valid conditions, respectively. (**A**) and (**D**) correspond to CD1 cue discriminability; (**B**) and (**E**) CD2; (**C**) and (**F**) CD3. A positive Pearson correlation for the valid and invalid conditions at CD1 was obtained with Pearson coefficients 0.83 and 0.93, respectively.

larger slope, i.e. a better performance, at ISI 150 ms compared to ISI 250 ms at a given cue discriminability and validity condition for respective participants.

From the Pearson correlation test, it was found that there is a positive correlation for the least discriminable cue CD1 for both valid and invalid conditions with Pearson coefficients 0.93 ($p < 0.05$) and 0.83 ($p < 0.001$), respectively. Even though correlations based on a small sample can be noisy, this result suggests that for CD1 cue discriminability level, the "slower" participants indeed benefit from additional 100 ms between cue offset and stimuli onset, compared to the "faster" participants with larger attentional effect at a shorter interstimulus interval.

3.5 Discussion and Conclusion

In a verification study, the impact of a shape-based endogenous cue discriminability on attentional performance was evaluated in a 2AFC orientation-discrimination task. Applying a gradual systematic shape-based design, a set of cues with different cue discriminability levels was generated which were then tested in a cue sensitivity test (Figure 3.3). For the main experiment three different levels of cue discriminability represented by the relative size of two horizontal lobes of the cue, CD1, CD2, and CD3 were selected. Attentional benefits and costs of the cue were evaluated comparing performance in the valid and invalid conditions at a given cue discriminability. Repeated-measures ANOVA results revealed a significant impact of cue discriminability on attentional performance. Furthermore, the temporal costs due to cue processing were addressed. In particular, the difference in the attentional performance at two tested interstimulus intervals positively correlated with the individual cue processing times for the low cue discriminability CD1.

Previously some studies demonstrated a gradual change in visual performance when varying cue properties [e.g. Müller et al., 2003; Fuller et al., 2009], whereas others argued for lack of flexibility of attentional allocation [e.g. Yeshurun and Carrasco, 2008]. Interestingly, Fuller et al. [2009] showed that manipulation of an exogenous cue modulates appearance even at supra-threshold contrast levels of the cue, indicating the flexibility of attention. The results of the present verification study preliminarily suggest an increasing trend of the effect of cue properties modulation on endogenous covert attention. These findings support the hypothesis that through a careful systematic design of a visual cue, in this case, based on its shape, a gradual distribution of attention appears to be possible. This, in turn, gives an insight into potential visual cues aiming to subtly and deliberately guide person's attention.

More complex cues, even at a discriminability level above the threshold, generally can take more time to process. Such, in [Brignani et al., 2009], in a detection task, authors found a stronger attentional effect at longer SOA when comparing eye-gaze and arrow cue with a more complex texture cue. In the present study, it was investigated whether additional processing time of less discriminable cues affects the temporal course of attentional effect. We addressed the hypothesis, that modulation of cue design with a purpose to improve performance, can come at a cost of additional cue processing. Based on performance differences at two different ISI and its correlation with individual cue processing times, it was found that "slower" participants who presumably needed more time to process less discriminable cue performed better at a longer ISI, in contrast to "faster" participants who needed less time for cue processing. This result can be explained if to assume a subsequent attentional mechanism on a time scale where the participant, first, decodes the cue and thereafter shifts attention to the cued location approaching attentional maximum at approximately 300 ms after the cue onset. With this assumption, at a shorter inter-stimulus interval, attention is already close to its maximum for the "faster" participants. In contrast, for the "slower" participants at shorter interstimulus interval attention does not yet reach its maximum. At a longer interstimulus interval, however, the "faster" participants are already past the maximal attentional performance, whereas the "slower" participants approach the maximum.

Considering the temporal range of the present experiment, these individual differences in the performance at different ISI could also be linked to the inhibition of return phenomena, when the attentional benefit in the cued location is observed only within a limited time period [Klein, 2000; Pratt and Fischer, 2002].

A possible question is, why a positive correlation between the attentional performance at two interstimulus intervals with the individual cue processing times was found only for the least discriminable cue CD1 and not for the other cues. This could happen due to unequal learning of cues of different discriminability levels after participants get more familiar with the cue. The individual response times depicted in Figure 3.6 and Figure 3.7 were recorded during the cue sensitivity test when participants were presented with the cue for the first time. We suggest that with practice participants get more familiar with the cues which led to an overall shorter processing time of the cues. It is proposed that during the main experiment the processing of the more difficult cue, CD1, remained temporally demanding, whereas, for the simpler and more discriminable cues, CD2 and CD3, participants required shorter processing times than in the cue sensitivity test. Therefore, the correlation between the individual cue processing times and performance at different interstimulus intervals was found only for the most complex and least discriminable cue but not for the more discriminable ones. In [Brignani et al., 2009; Guzzon et al., 2010] authors also did not find significant differences in the attentional effect for the simplest cues such as eye-gaze and arrows at various SOA. Furthermore, from [Brignani et al., 2009] no significant difference emerged for a more complex texture cue compared to the simpler ones, although a trend of a larger validity effect for a more complex cue was observed for larger SOA. The results of our study conform to previous findings showing no significant difference in attentional performance at different ISI for less complex cues. Nonetheless, a positive trend of correlation of performance at two different ISI with the individual cue processing time for the most complex cue, indicates that the individual cue processing time can affect the temporal course of attentional effect for the most complex cue in the time range used in the present study. Testing more interstimulus intervals as well as different cue display times is of interest for further investigation of cueing effect on the temporal course of attention [Gibson and Bryant, 2005].

It is important to comment on the limitations of the present study. The results in Figure 3.5 suggest a gradual effect of cue discriminability on the attentional performance, however, the difference for the slopes at the supra-threshold cue discriminability levels CD2 and CD3 was close to but not significant. This could be contributed by the ceiling effect caused by a challenging paradigm design with parametrically modulated endogenous cue discriminability. The difficulty of the task and individual variability, as well as a small sample size of this verification study, could also partially offset the low significance level. Furthermore, it is possible that performance differences for corresponding cue discriminability levels partly originate from incorrect decoding of the cue direction, particularly for the most complex cue. Specifically, to be sure about actual perceived cue direction in the main experiment, in each trial one would have to collect an additional response from participants. It would, however, yet increase the load of the task for the participants and make it more tiring. This verification study intended to verify a gradual effect of cue complexity on attentional benefits and costs. Small sample size was, thus, tested for this preliminary feedback purpose. Nonetheless, the results of this verification study hint towards the scaling of attentional benefits and costs with cue discriminability which can be useful when a flexible allocation of attentional resources is facilitating performance. In future studies, a larger number of participants, as well as a bigger set of cue discriminability levels, could be used to get more insight into attentional scaling.

Generally, when discussing potential visual cue design with the purpose to improve person's performance, it is rather counter-intuitive to develop a more complex or less discriminable cue. The rationale in this study, however, was to show that the cue properties can be tailored to vary the relation between attentional benefits and costs induced by the cue, but at the same time, one has to keep in mind that this tuning is limited by the emerging additional cue processing temporal costs. Considering the increasing tendency towards rather subtle visual cues in contrast to strong symbolic cues in diverse applications, in future validation studies, it would be interesting to test whether it is possible to find a "sweet spot" of cue complexity for attention. Namely, whether the cue can be discriminable just enough to bring the performance represented by attentional benefits and costs to its maximum, but does not impose significant cue processing costs.

To conclude, first, attentional performance seems to resemble a scaling behavior with varying properties of a spatial endogenous cue, specifically, its discriminability, even when cue discriminability is above the threshold level. Second, individual temporal differences in the processing of complex cues can impact attentional performance, particularly, the cue is complex. Thus, when designing endogenous cues and tuning their properties to purposely guide user's attention, it is important to consider the temporal processing costs of the cue, specifically, in case of a high cue complexity, but not for simple cues. The current verification study serves as an insight into the potential design of visual cues aiming to improve people's performance. The main implication of this study is that to improve the total performance of the user over numerous locations using a flexible visual cue design, it is important to consider the trade-off between the cue complexity and its benefits vs costs effect, as well as the individual processing time of the cue.

3.6 Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: `https://osf.io/yfbqt/`.

3.7 Additional information

3.7.1 Author contributions

Conceptualization, O.L-S., S.W. and K.R.; methodology, O.L-S., S.W., T.S.A.W. and K.R.; software, O.L-S.; validation, O.L-S., S.W. and K.R.; formal analysis, O.L-S.; investigation, O.L-S., S.W. and K.R.; resources, S.W.; data curation O.L-S. and K.R.; writing–original draft preparation, O.L-S.; writing–review and editing, S.W., T.S.A.W. and K.R.; visualization, O.L-S. and K.R.; supervision, S.W. and K.R.; project administration, S.W. and K.R.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

3.7.2 Funding

This study was conducted with support of the Integrative Augmented Reality (I-AR) intramural funding of the University of Tübingen, Germany.

3.7.3 Conflict of interest

We declare that Siegfried Wahl and Katharina Rifai are scientists at the University of Tübingen and employees of Carl Zeiss Vision International GmbH, as detailed in the affiliations. Intellectual contributions of Thomas S. A. Wallis to the publication were made prior to commencing his employment at Amazon. There is no conflict of interests regarding the current study.

4. Augmentation impacts strategy and gaze distribution in a dual-task interleaving scenario

Lukashova-Sanz, O., Wahl, S., and Rifai, K. (2022). Augmentation impacts strategy and gaze distribution in a dual-task interleaving scenario. *International Journal of Human-Computer Interaction*, 38(4):383-394

4.1 Abstract

When interleaving multiple tasks, people are confronted with a decision of how to distribute a finite amount of time between several tasks, which defines the task-interleaving strategy. In some challenging task interleaving scenarios where accurate timing is essential, people perform worse than they could have. With the growing advancement of technology, such as augmented reality, it became possible to impact people's strategy and improve their performance. However, when augmenting visual input with additional visual content, the augmentation not only introduces the possible benefit but can also capture attentional resources. It is, thus, important to investigate how visual augmentation affects people's performance in cases when otherwise people underscore in their performance. In the current study, using a psychophysics approach, it was investigated how visual augmentation impacts the task-interleaving strategy and, thus, performance in a dual-task setting with unequal task importance. In a simple dynamic 3D environment, four visual augmentations were generated aiming to prompt the user when it is more beneficial score-wise to switch from one task to another. The mean duration on one task before the task switch, as well as the resulting total performance, were evaluated in combination with the gaze direction distribution. In terms of the strategy and the total performance, all augmentations showed an advantage compared to when augmentation was not present. Furthermore, an abrupt augmentation onset based on the individual response time of the participant was more beneficial score-wise for the strategy compared to a constantly present visual augmentation. However, it affected the natural gaze direction distribution indicating the allocation of attentional resources to the augmentation. The results of this study provide an insight into potential visual augmentation designs aiming to improve user's performance in a challenging dual-task interleaving setting.

4.2 Introduction

In daily life, people are inevitably faced with performing multiple tasks at the same time. One important type of multitasking is task interleaving. Given a finite amount of time, a person performs one task at a time, alternating between attending to several concurrent tasks. Task interleaving is very prominent in our society ranging from office workers [Dabbish et al., 2011] to air traffic controllers [Lee and Taatgen, 2002], and therefore, has been extensively studied in various disciplines [see Janssen et al., 2015]. When interleaving between multiple tasks, people are essentially confronted with a decision of how to distribute a fixed amount of time between the given tasks, or how much time to spend on one task before switching to the other, and at which moment in time to return to the first task. This scheduling decision is referred to as the task-interleaving strategy. The way how people distribute their time when interleaving between tasks is essential for different aspects of people's personal and professional lives such as safety [Janssen et al., 2012; Brumby et al., 2009] as well as their productivity [Janssen et al., 2019; Farmer et al., 2018]. Numerous studies investigated how efficiently people interleave between tasks [see Janssen et al., 2015]. There is much evidence that in many dual-task scenarios people are capable to adjust and optimize their strategy to make their performance more efficient [Farmer et al., 2018; Janssen and Brumby, 2015]. However, there are some challenging dual-task settings where people appear to underscore in their performance [Janssen et al., 2019]. From multiple studies it is apparent that a particular adopted task-interleaving strategy can depend on numerous factors including complexity and characteristics of the tasks, personal motivation and experience with the tasks, and many others [Janssen and Brumby, 2015; Mone and Shalley, 1995; Janssen and Brumby, 2010; Howes et al., 2009]. Among the factors affecting people's decision on when to switch from one concurrent task to another while task interleaving, is the payoff function of the tasks [Janssen et al., 2019; Farmer et al., 2018; Janssen and Gray, 2012]. The payoff function reflects the relative importance of each task and describes a reward level for each of the tasks in a task interleaving scenario. In one recent study it was shown that even though generally people appear to be flexible and adjust their strategy towards the higher performance during task interleaving, in a scenario when one task is significantly more important relative to the other, people tend to averse to the risk of failing the more important task and switch to it from the less important one more often than they could have [Farmer et al., 2018]. Specifically, in a dual-task scenario with the unequal importance of two tasks, given a final amount of time, people tend to fall short on the favorable performance-wise mean duration on the less important task, leading to a decrease of the total performance. These findings well relate to everyday life observations when the accurate estimation of time is necessary for more efficient performance. One plain example is when a researcher has to interleave between, first, correcting a closed-answer test of his students, and, second, monitoring his or her manual coffeemaker while preparing the morning coffee in the adjacent kitchen corner. Even though the researcher might be familiar with how long it takes for coffee to be ready, in the absence of additional information, he or she is likely to interrupt the test correction and go and check the coffeemaker more often than necessary to averse to the risk of missing the critical point and switch it off in time. The total performance in such a scenario, however, would be lower than it could have been as the time spent on the test correction is shorter than it could have been, given a finite total time for both tasks.

With rapidly developing technology such as augmented reality, additional visual cues became a potentially powerful tool to purposefully guide user's attention and improve user's performance [Dey et al., 2018; Coughlan and Miele, 2017; Booth et al., 2013; Zarraonandia et al., 2014; Lukashova-Sanz and Wahl, 2021]. On the other hand, augmenting visual input with additional visual content has also its drawbacks. In particular, it introduces a trade-off of the potential benefit for the performance and, at the same time, captures attentional resources, which essentially becomes a bottleneck for the design of visual augmentation [Raja and Calvo, 2017; Akoumianakis and Stephanidis, 2005; Favela et al., 2010]. It is, therefore, of practical importance to investigate how visual augmentation impacts people's performance on occasions when they perform inefficiently. In the case of a task-interleaving scenario with unequal payoff functions of each task, an augmentation can potentially support the user's task-interleaving strategy by prompting him or her, when it is more beneficial performance-wise to switch from one task to another. It is not clear, however, how much in advance the augmentation should become informative before the suggested switch moment in time.

In this study, using a psychophysics approach, it was investigated how visual augmentation affects the task-interleaving strategy and, thus, performance in a dual-task setting with unequal task importance. The experimental paradigm was implemented in a simple dynamic 3D virtual environment. A set of visual augmentations was generated aiming to prompt the user when it is potentially better score-wise to switch between two tasks. In particular, the time-lead of the augmentation onset before the predefined suggested switch moment was modulated, ranging from a constantly present visual augmentation during the task performance, to an abrupt onset of the visual augmentation based on the individual response time of the participant. As a measure of the task-interleaving strategy, the mean duration on one task before switching to the other, as well as the resulting total performance over both tasks, were evaluated. As a measure of efficiency, a more efficient strategy is expressed in a longer mean duration spent on one task before switching to the other, resulting in higher total performance. In the current study, even though the total performance can possibly be improved using the augmentation, additional visual information can also draw attentional resources, which can be an issue in some real-life scenarios [Wickens, 2021]. We addressed this potentially negative effect of visual augmentation by evaluating another performance parameter — the distribution of the gaze direction during the task performance. Specifically, the hypothesis was that visual augmentations, including the abrupt response-time-based and constantly present, would improve the participant's task-interleaving strategy, but to a different extent. We expected a higher performance expressed in the total score when the participant was prompted with an exact switch moment (response-time-based augmentation) than when the participant, even though having an additional piece of visual information as a reference, still had to decide when it is long enough before the switch (continuous augmentation). Furthermore, an impact of augmentation on the gaze distribution was expected with a bias of gaze direction towards the location of augmentation when augmentation was present.

In the present study, four experimental augmented conditions were tested, namely, four time-leads of a motion onset of a head-contingent visual peripheral augmentation stimulus are compared. In a dual-task interleaving scenario, first, the impact of the time-lead of the augmentation on the mean duration on one task before switching to the other was evaluated. Second, the total performance was evaluated across various time-leads. Finally, the gaze direction distribution for all experimental conditions was evaluated.

4.3 Methods

4.3.1 Participants

In this study, 14 participants with an average age of 24.4±3.7 years were tested. All procedures conformed to Standard 8 of the American Psychological Association's "Ethical Principles of Psychologists and Code of Conduct (2010)". The study was approved by the ethics committee of the Faculty of Medicine at the University of Tuebingen. Signed informed consent was obtained from each participant before the measurements. All participants except one had had previous VR experience, however, the prior experience with virtual reality was not a requirement to participate in the experiment.

4.3.2 Apparatus and Stimuli

4.3.2.1 Apparatus

A dual-task setting was implemented in a 3D virtual environment. The visual content was displayed to the participant using HTC Vive Pro Eye (HTC Corporation, Taoyuan, Taiwan) virtual reality headset at a refresh rate of 90 Hz running on a Windows 10 PC with NVIDIA GeForce GTX 1070 graphics card (NVIDIA Corporation, Santa Clara, California, USA). The horizontal and vertical fields of view of the headset reported by the manufacturer are 100° and 90°, respectively. The interaction with the environment was conducted by the participant via the HTC Vive controller. The eye-tracking data was collected using a built-in eye tracker at a rate of 90 Hz. The head rotation data were collected using tracking base stations 2.0 of HTC Vive Pro setup. The experimental paradigm was generated using the Unity Game engine version 2019.3.15.f1 [Unity Technologies, 2019]. The data analysis was performed using Python 3.6 packages NumPy [Van Der Walt et al., 2011] version 1.19.1, SciPy [Virtanen et al., 2020] version 1.5.2 and Pandas [Mckinney, 2010] version 1.1.3. The statistical analysis was conducted using R [R Core Team, 2020] version 3.6.1 The data visualization was performed using Python packages Matplotlib [Hunter, 2007] and Seaborn.

4.3.3 Experimental Procedure

4.3.3.1 Dual-task-interleaving Setting: Configuration

A dual-task-interleaving scenario was designed and implemented in a virtual 3D environment. In contrast to a 2D paradigm, a 3D virtual reality environment offers a possibility to rotate the head freely while tracking the eye movements. This, in turn, enables the recording of a more natural gaze behavior compared to a screen-based experiment. The experimental setting is depicted in Figure 4.1. The environment consisted of two virtual rooms. Three blue spheres were placed in one room, and a red sphere in the other room. During the experiment, all of the spheres were continuously expanding in size until a maximum of five times the original size. Once a sphere reached its set maximum size, it “exploded” and reset to its original size followed by an expansion again. At the beginning of each trial, the participant had a starting capital of points. Each time a sphere exploded, the participant received a penalty of a certain amount of points (see Section 4.3.3.3). The ultimate goal of the participant was to keep as many total points in each trial as possible. To prevent a sphere from an explosion, the participant could reduce its size by touching an expanding sphere with a virtual stick operated by a mobile manual controller. As long as the participant was touching the sphere with the stick, the sphere was reducing in size at a constant velocity. If the sphere reached its minimum size, the size reduction was terminated. Once the participant stopped touching the sphere, the sphere continued to expand again. The participant was able to instantly switch between two rooms by pressing a trigger button of the manual controller. The participant could perform a task only in one virtual room at a time. Thus, during each trial, the participant interleaved two tasks, the blue and the red, by switching between two rooms: with the blue and the red spheres. Upon each expansion of the red sphere, its expansion speed was slightly varying between $\pm 5\%$ of its mean value. Doing so we intended to prevent adaptation of the participant to an exact time duration. In case the participant switched to the red room too late and the red sphere had previously exploded, it was indicated by the red sphere changing its color to black, and then turning red again once the participant reduced the size of it via touching it with the virtual stick. The participant was additionally provided with audio feedback upon each sphere explosion with distinct sounds for the blue and red spheres, respectively. Also, haptic feedback was provided to the participant via a vibration impulse of the manual controller to indicate a collision of the virtual stick and a sphere. The height at which spheres were presented was adjusted before the start of the experiment according to the height of each participant. Namely, the vertical coordinate of the headset was used as a reference to set the vertical position of the spheres. The vertical coordinate of the blue spheres as well as of the red sphere was set to a value 0.3 Unity-meters lower than the headset recorded height. To prevent a patterned hand movement on the blue task, the vertical coordinate of each blue sphere was set to a random value between ± 0.2 Unity-meters of its set value after each sphere explosion as well as each new entry to the blue room. The horizontal coordinate of each of the blue spheres was set to -0.6, 0, and 0.6 Unity-meters, respectively, where 0 corresponds to the center of the room. The horizontal coordinate of the red sphere was set to a random value corresponding to $\pm 45°$ of visual angle relative to the headset to make the red task more challenging. The depth at which each sphere was located was set to 1 Unity-meter from the center of the room. Each time the participant switched to a room, the direction of the headset was set towards the middle of the corresponding room. The events of a collision of the virtual stick with any of the spheres, as well as the explosion of each sphere were constantly tracked and correlated with the timeline of the experiment. The spatial configuration of the experiment including augmentation is schematically shown in Figure 4.2. The details on the augmentation design are described in Section 4.3.3.2.

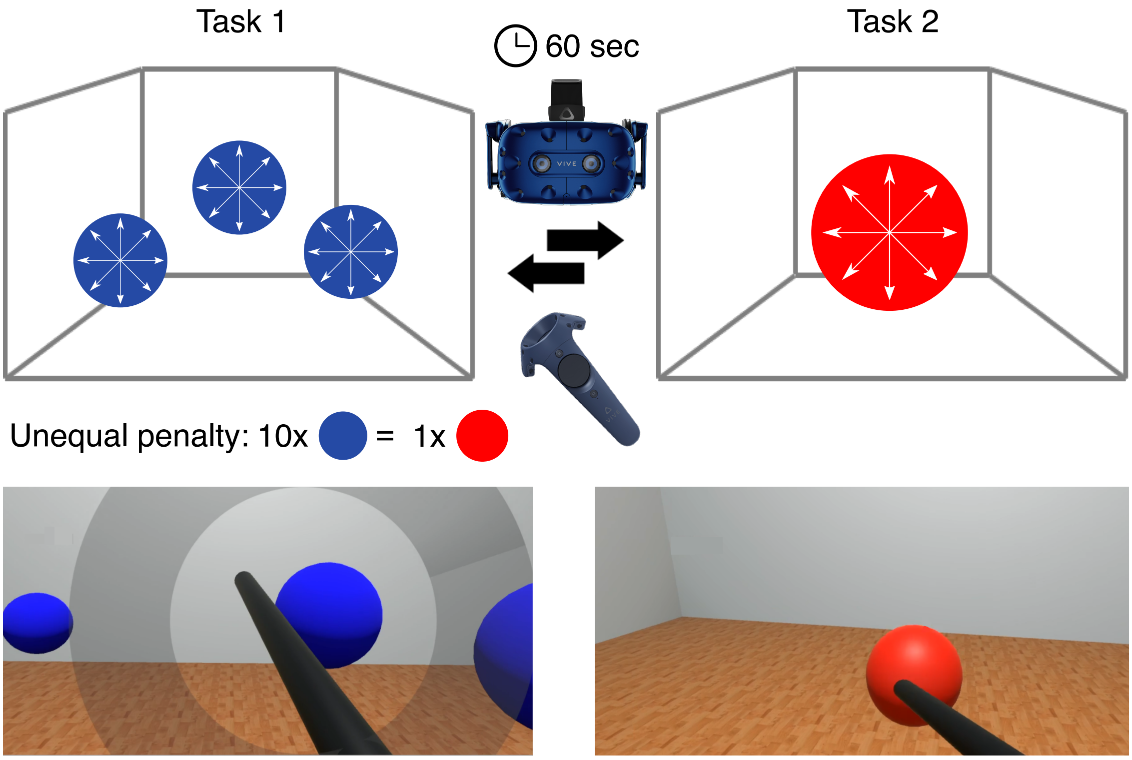

Figure 4.1: Experimental paradigm: a dual-task interleaving setting with the unequal importance of each sub-task. The unequal task importance was implemented through 1:10 penalty ratio for the explosion of each blue and red sphere, respectively. During each 60 sec-trial the participant interleaved between the blue and the red tasks performing one sub-task at a time. The participant switched between two tasks by pressing a button on the HTC Vive controller. The example scene views demonstrate the case when augmentation was present, thus, the head-contingent ring is overlaid with the visual scene of the blue task. In the control baseline condition, the ring was not present. Note, that the example scene views demonstrate a larger field of view than the actual one in the headset — when performing the experiment in the virtual environment with the headset, the outer edge of the augmentation ring was not visible. For more details on the experimental paradigm see the description in the text.

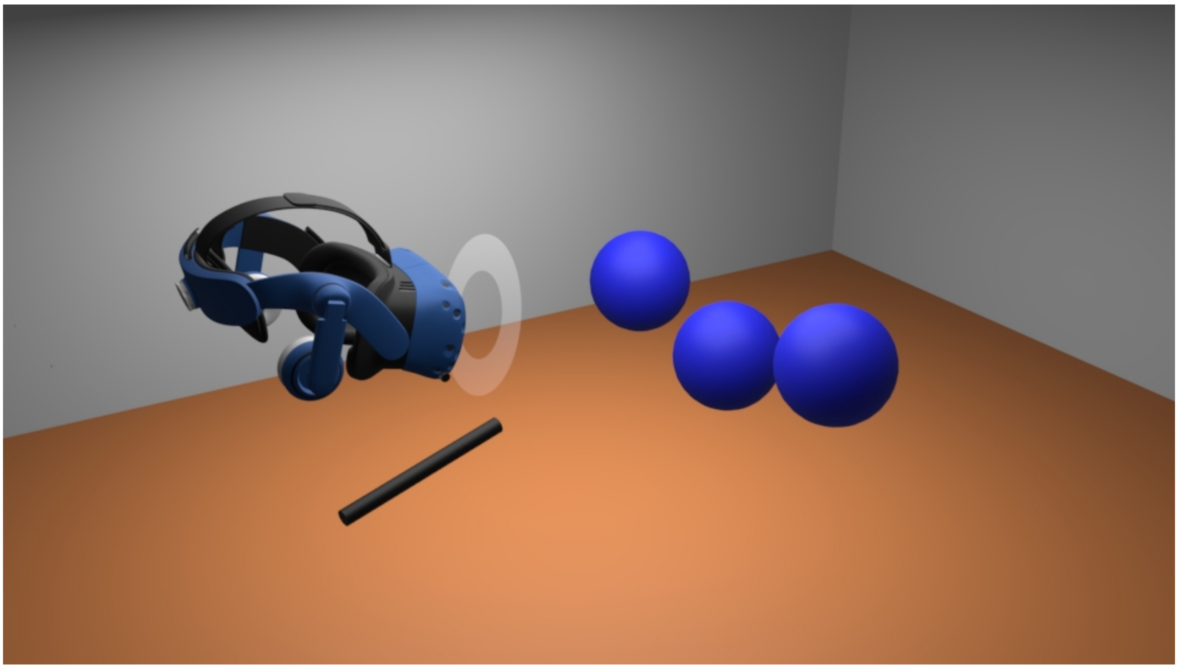

Figure 4.2: The spatial configuration of the experiment. As an example, the blue room configuration is shown. The head-mounted device represents the head position of the participant, the rest of the figure schematically represents the virtual environment designed for the experiment. The details about the augmentation stimulus are described in Section 4.3.3.2.

4.3.3.2 Visual Augmentation: a Head-contingent Peripheral Motion Onset Stimulus

As the visual augmentation, a subtle semi-transparent head-contingent ring was used. The term "head-contingent" indicates that the ring was always concentric with the forward direction of the VR headset, meaning, it was always in the same position within the visual field of the participant regardless of the head movement. The informative part of the augmentation was represented by a motion onset of an additional ring minor sector with a lower transparency level (Figure 4.3). The augmentation ring was designed as a head-contingent to be always fully visually available for the user irrespective of his or her heading direction. The peripheral location of the ring was selected to keep the augmentation unobtrusive for the tasks at hand. The inner and outer radii of the ring were arbitrarily set to 30° and 50° relative to the center of the visual field of view, respectively. Considering the field of view of the head-mounted device used to display the visual content, the outer border of the ring was not visible to the participants, whereas the inner border of the ring was laid in the peripheral part of the visual field of view. The size of the ring sector was arbitrarily fixed to 20% of the ring circumference resulting in 18° angle made by the arc of the sector at the center of the circle. To maintain the color scheme of the virtual environment when displaying

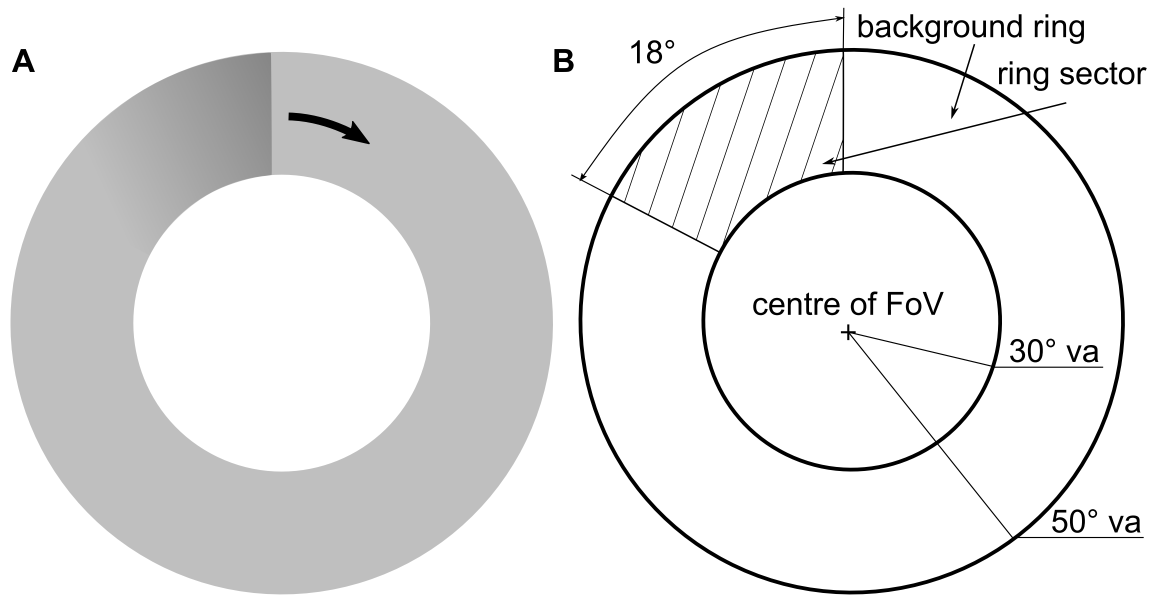

Figure 4.3: Visual augmentation: a head-contingent peripheral stimulus consisting of a background ring and an additional ring sector moving clockwise. The motion speed was set to a constant value of 68°/sec resulting in a full 360° turn in approximately 5 sec. The ring sector is indicated in (**B**) by a striped pattern. The size of the sector was arbitrarily set to 20% of the full ring circumference resulting in 18°. The ring was always concentric with the heading direction of the participant regardless of the head movements. The inner and outer radii of the ring were set to 30° and 50° in visual angle (va), respectively. The transparency of the ring sector was gradually decreasing such that one edge of the sector is maximally discriminable on the background ring, whereas the other edge is blended into the background ring. For more details, see the description in the text.

augmentation, the color of the background ring and moving ring sector was set to gray, namely, RGB values (127, 127, 127). The transparency for the background ring was set to a constant value of 100. To avoid the ambiguity of which part of the ring sector to attend to when it is moving, a gradient was applied to the transparency. Specifically, the transparency of the moving ring sector was set to a gradually decreasing value starting from a value of 100 to match the background ring, reaching a minimum value of 50 to be well discriminable on the background ring. Upon the motion onset, the ring sector was moving clockwise at a constant speed of 68°/sec resulting in a full 360° turn in approximately 5 sec. The starting position of the moving ring sector was set to match its most right and least transparent (most opaque) edge with the top middle position of the background ring (see Figure 4.3). The time-lead of the ring sector motion onset was varied depending on the experimental condition. More details on the informative content of the augmentation and different experimental conditions can be found in Section 4.3.3.5.

4.3.3.3 Unequal Importance of the Tasks Implementation: Unequal Penalty

The unequal importance of the tasks was introduced via the game score. For each trial, the participant had a starting capital of 200 points. Each time a sphere exploded, the participant was given a penalty of 1 point for a blue sphere and 10 points for a red sphere. At the end of each trial, the participant was provided with cumulative feedback with the total amount of points that were left from the starting capital, as well as how many points were lost on the blue and red spheres separately. Before each trial, the starting capital was reset to 200 points.

The penalty ratio is an important factor for the task-interleaving strategy [Farmer et al., 2018]. Here, the selected ratio was used to capture a challenging scenario where the performance could be improved. Beyond the scope of the present study, a wider range of penalty ratios could be evaluated.

4.3.3.4 Estimation of the Suggested Switch Moment

Before the main experiment, the best score-wise switch moment from the blue to the red task was estimated. The following independent parameters were considered: the penalty ratio for each respective blue and red sphere explosion, the trial duration, and the expansion speed of the blue and red spheres. The basic assumption was made that none of the spheres were saved from an explosion during the trial, which surely does not correspond to the reality, but is sufficient for the initial estimate. It was estimated that the minimum total penalty, or the maximum total score in a trial is achieved when the mean duration on the blue task is just below the time necessary for one red sphere to expand from its minimum to its maximum size. This estimate indicates that in the current paradigm the most efficient strategy in terms of performance for the participants was to stay on the blue task for such a period before switching to the red task just enough to keep the red sphere from an explosion. The individual parameters reflecting how well and how fast the participant performs the blue and the red task, namely, the rate of blue spheres' explosions when performing the blue task, and the mean duration on the red task per visit were not considered for the estimation. For the current experiment, it was sufficient to know that the most efficient strategy in terms of performance for the participants was to keep as many red spheres from an explosion as possible.

4.3.3.5 Augmented Task-interleaving: Experimental Conditions

The idea of the augmentation was to prompt the participant when it is better to switch from one task to another to achieve a higher score, here: from the blue to the red task. In the experimental conditions where augmentation was present, the participant was presented with the visual augmentation whenever performing the blue task (for details on the augmentation design see Section 4.3.3.2). The participant was suggested to use the augmentation which aimed to optimize the participant's mean duration of time spent on the blue task. Doing so, it was intended to increase the total score over each trial. The ring sector motion onset was based on how much time was left before the explosion of the red sphere. As described in Section 4.3.3.1, the expansion speed of the red sphere was slightly varied, and, thus, the duration of the sphere expansion from its minimum size to its maximum was also moderately varied. To account for this, and to ensure that augmentation was always presented on time, we considered the minimum duration of the red sphere expansion as a reference. In the present study, this minimum duration was set to 5 sec. The specific time-lead of the ring sector motion onset essentially defined the experimental conditions of the study. In total, five experimental conditions were tested, which are described in Table 4.1. In particular, Condition 0 was a control condition when no augmentation was presented, whereas in Conditions 1 to 4 the motion onset of the ring sector was initiated with respective time-leads before the red sphere explosion. In Condition 1, the motion onset was initiated instantaneously whenever the participant was performing the blue task, thus, the motion was continuous with no abrupt onset. In Condition 2, the motion started later during the performing of the blue task, however, with a sufficient time-lead for the participant to detect the motion and still have some time on the blue task before the switch. Specifically, the time-lead in Condition 2 was set to 2200 ms. In Condition 3, a short fixed time-lead of 1200 ms was used. The value was selected based on the results from the response time test (see Section 4.3.3.6), where the response time is defined as the time that the participant needed to detect the motion onset, switch to the red room, and touch the red sphere. The value of 1200 ms was selected as approximately the average response time among the participants. Finally, in Condition 4, individual response times were used as the time-leads, which were obtained from the response time test (see Section 4.3.3.6).

4.3.3.6 Response Time Test

In this study, the individual response time is referred to as the time necessary for the participant to notice the motion of the augmentation ring sector while performing the blue task, switch to the red task, and touch the red sphere with the virtual stick. To estimate individual response times, each participant performed a response time test. During this test the participant was instructed to follow the general procedure of the main experiment, except the score was not of importance in contrast to the main experiment, but rather the speed of response mattered. All of the experimental parameters were identical to the main experiment setting. During each trial of the response time test, the participant was, first, focusing on the blue task, then after a 2 sec period the motion of the ring sector was initiated. Before the test, the participant was instructed to switch to the red task via controller and touch the red sphere with the virtual stick as soon as possible after detecting the motion of the ring sector. Thereafter, the participant switched back to the blue task by pressing the trigger button and the cycle started again. Each participant continued this procedure for 5 trials of 60 sec, resulting in a total of approximately 75 response time estimates for each participant.

4.3.3.7 Main Experiment: General Procedure

Each participant performed three experimental sessions. In the first session, the participant performed two experiments: first, the baseline task interleaving where augmentation was not present (Condition 0), followed by a separate response time test. During the second session, another two experimental conditions were performed. Finally, during the third session, the remaining two experimental conditions were executed. The order of the experimental conditions in the second and third sessions was randomized. Each participant in each experimental condition performed 15 trials of duration 60 sec. The expansion speed of spheres was set such that if a participant would not touch any of the spheres during the whole trial, he or she would lose 30 blue spheres and 12 red spheres in each trial. The number of switches from the blue to the red task varied individually between participants and experimental conditions being approximately 15 switches to the red task per trial. Each condition was preceded by a practice session of five trials of 60 sec with the procedure identical to the main experiment. During the practice session all participants familiarized with the dynamics of the tasks.

4.3.3.8 Gaze Direction Data

To evaluate gaze behavior, the gaze direction data in the headset coordinates, as well as the world coordinates, were collected at a frequency of 90 Hz. Before each experimental session, for each participant, the built-in calibration procedure of the eye tracker was performed. The gaze position data was accessed using a customized written Unity script utilizing the HTC SRanipal SDK package functions. To prepare the data for further processing, first, similar to [Imaoka et al., 2020], the raw data were filtered based on the eye data validity bitmask value, which represents the bits containing all validity for the current frame. After the filtering, only the data where the eye data validity bit mask had value 31 for both eyes, were selected. Doing so, the data where the eye tracker partly or completely lost the pupil (including blinks) was filtered out. Next, the gaze position was calculated in spherical coordinates. In particular, the polar

ϕ and azimuthal θ angles were computed using Equation (4.1) and Equation (4.2). In Unity, the z-axis corresponds to the depth dimension.

$$\phi = \arctan \frac{x}{z}, \tag{4.1}$$

$$\theta = \text{arctan2}\,(y, \sqrt{x^2 + z^2}), \tag{4.2}$$

where (x, y, z) are coordinates of normalized gaze directional vector in headset coordinates. Note that SRanipal returns the gaze direction vector in the right-handed coordinate system. To convert the coordinates to the left-hand coordinate system (same as Unity world coordinate system), the x-coordinate was multiplied by -1. To compute the gaze position in Unity world coordinate system, the gaze position in headset coordinate system was multiplied by the head rotation quaternion.

To summarize, the gaze direction in visual angle was computed in visual angle relative to the heading direction (headset coordinates), as well as relative to the forward direction in the virtual environment (world coordinates).

4.3.4 Analysis

4.3.4.1 Response Time Test

To estimate the individual response time of each participant from the response time test, the time duration between the onset of the ring sector motion and the moment of collision of the virtual stick and the red sphere was computed.

4.3.4.2 Main Experiment: Behavioral Data

The variable of interest was the mean duration on the blue task which reflects the task-interleaving strategy of the participant. Another variable of interest was the total score in points over both tasks in each trial, which is suggested to be directly correlated to the mean duration on the blue task. Both variables were normalized to the corresponding baseline value in Condition 0. Specifically, the mean duration on the blue task in Condition 0 was subtracted from those in the remaining four conditions. Similarly, the total score in Condition 0 was also subtracted from the respective scores in the augmented conditions. The main effect of the augmentation time-lead was statistically evaluated by one-way repeated-measures analysis of variance (ANOVA), followed by post-hoc analysis with Tukey correction to reveal significant differences between the conditions. The ANOVA assumptions were checked using the Shapiro-Wilk test for the normality assumption and the Levene test for the homogeneity of variances assumption. In case the ANOVA assumptions were not met, an alternative Kruskal-Wallis test as an alternative to one-way ANOVA was performed, with a subsequent Dunn test for multicomparison. The mean duration on the blue task and the total score were two separate dependent variables, and the augmentation condition was the independent variable. Furthermore, the correlation between the total score and the previously measured response time was evaluated. Specifically, the difference in performance in Condition 3 and Condition 4 as a function of the individual response time was computed and evaluated using Pearson correlation test.

4.3.4.3 Gaze Direction Data

The gaze direction in the headset coordinates, as well as the world coordinates, was evaluated for every participant in each experimental condition. Specifically, the gaze direction during performing the blue task was of interest as augmentation was present only when the participant was performing the blue task. The gaze direction distribution was evaluated by plotting the variance of the horizontal visual angle distribution of the gaze direction. The variance of the gaze distribution for each experimental condition was normalized to the variance of the gaze distribution in Condition 0 by taking the ratio. The main effect of augmentation time-lead on the gaze distribution was evaluated by one-way ANOVA, followed by post-hoc analysis with Tukey correction. The ANOVA was performed separately for the gaze direction in the headset and world coordinates, respectively. In case the ANOVA assumptions were not met, an alternative Kruskal-Wallis test was performed with a subsequent Dunn test for multicomparison.

4.4 Results

4.4.1 The Mean Duration on the Blue Task and the Total Score

In Figure 4.4, the normalized mean duration on the blue task as well as the normalized total score, averaged over 14 participants are presented. The average between-the-participants' mean duration on the blue task in the baseline condition (Condition 0) was 3210 ms with a standard error of the mean of 112 ms. The corresponding total score value was 150 points with a standard error of the mean of 2 points. For the individual data see Supplementary information.

From Figure 4.4, the mean duration on the blue task as well as the total score in the baseline Condition 0 are the lowest compared to other conditions. This indicates that the experimental paradigm captured well the sub-optimal in terms of the total score performance of the participants, leaving some room for improvement for the augmented experimental conditions 1, 2, 3, and 4. The ANOVA assumptions for the mean

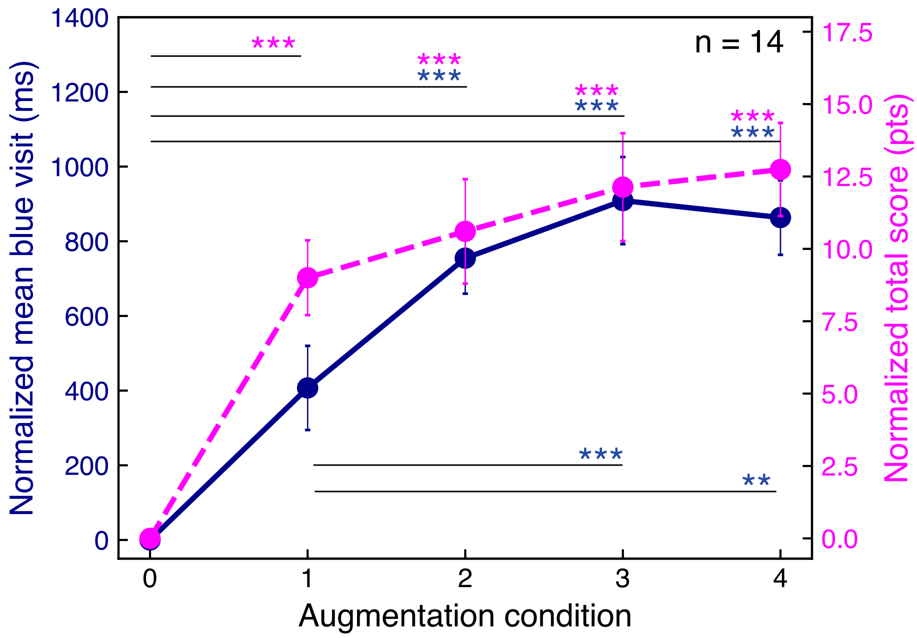

Figure 4.4: The normalized mean duration on the blue task (dark blue), and the normalized total score (magenta), averaged over 14 participants. The x-axis represents experimental augmentation conditions. The error bars represent the standard error of the mean. The level of significance represents the results of a multicomparison post-hoc test with a Tukey correction. The indicators of significant differences are *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

duration on the blue task were not met, thus, a Kruskal-Wallis test was performed as a non-parametric alternative to one-way ANOVA. The main effect of the augmentation condition on the mean duration on the blue task was significant with $p < 0.001$. The total score follows a similar trend as the mean duration on the blue task, being higher for conditions where augmentation was present. The ANOVA assumptions for the total score were met. From ANOVA with the total score as the dependent variable, the main effect of the augmentation time-lead was significant with $p < 0.001$. From the post-hoc analysis (Dunn test), the difference in mean duration on the blue task between Condition 0 and Condition 1 was not significant. Comparing Condition 0 with the remaining three conditions, the difference was found to be significant with $p < 0.001$ for all conditions (Condition 2, Condition 3, and Condition 4). Furthermore, there was a significant difference in the mean duration on the blue task between Condition 1 and Condition 3, as well as between Condition 1 and Condition 4 with $p < 0.001$ and $p < 0.01$, respectively. The multicomparison of the total score revealed significant differences between Condition 0 and all other conditions with $p < 0.001$ for all conditions. General learning of the task was evaluated on a trial-by-trial basis for each participant and was not significant (see Supplementary information).

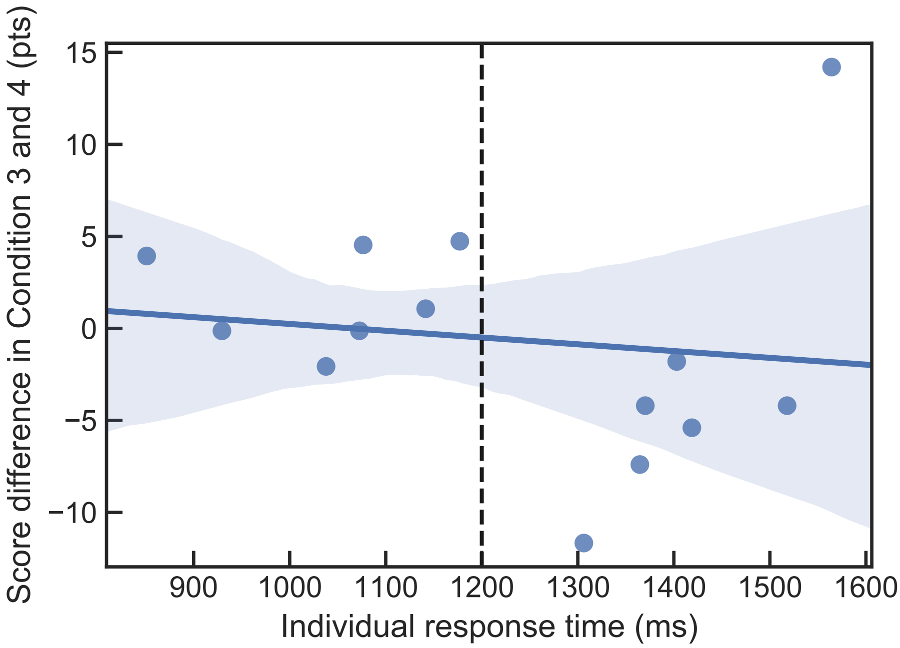

Figure 4.5: The difference in total score in Conditions 3 and 4 as a function of individual response time. Note, that the score in Condition 3 was subtracted from the score in Condition 4. The blue line corresponds to the best linear regression estimate with a confidence interval of 95%. The black dashed vertical line indicates the fixed value of the time-lead 1200 ms in Condition 3.

Furthermore, we evaluated the performance in the condition with short fixed time-lead (Condition 3), and the condition with individually set response-time based time-leads (Condition 4). In Figure 4.5, the difference in the total score between Condition 3 and 4 as a function of individual response time is shown. Note, that the score in Condition 3 was subtracted from the score in Condition 4. The dependency on Figure 4.5 shows a negative trend, however, there is no significant correlation found. The absence of significance is highly influenced by a data point for one participant with the largest response time. One possible explanation of this outlier is a general improvement in the task performance for this participant as the response time was measured during the first experimental session, Condition 4 was performed during the second session, and Condition 3 — during the third session.

4.4.2 Gaze Direction Distribution

In Figure 4.6 (A,B) an example of a gaze direction dataset for one participant in one experimental condition is shown. The plot represents only the data when the participant was in the blue room. The data is visualized through a joint histogram plot with hexagonal bins where the depth of the color represents the count of data points in the corresponding region. Figure 4.6 (A) demonstrates the gaze direction in degrees of visual angle in the headset coordinates. The gray ring represents the interval of visual angles in the headset coordinates corresponding to the augmentation ring in experimental conditions with augmentation present. In Figure 4.6 (B) the same data is combined with the head orientation and represents the gaze direction in the world coordinates, thus, no head-contingent augmentation ring is depicted (for details on computing the gaze gaze direction in the headset coordinates please refer to Section 4.3.3.8). The variance of the horizontal visual angle of the gaze direction distribution in all experimental conditions was compared. To illustrate the trend of the shape of the gaze distribution in different experimental conditions, in Figure 4.6 (C,D) the gaze direction distribution in horizontal visual angle is demonstrated, combined for all participants, and grouped by the augmentation condition. The ANOVA assumptions were not met for the gaze distribution in the headset coordinates as well as the world coordinates. Thus, a Kruskal-Wallis test was performed as a non-parametric alternative to one-way ANOVA. A significant main effect of the augmentation time-lead on the variance of the horizontal visual angle distribution of the gaze direction in the headset coordinates was found with $p < 0.01$. From the post-hoc analysis of the gaze direction horizontal visual angle variance in the headset coordinates, a significant difference was found between the baseline Condition 0 and Condition 3, as well as between Condition 0 and Condition 4 with $p < 0.05$ and $p < 0.001$, respectively. Statistical analysis of the variance in the world coordinates did not reveal any significant differences between experimental conditions. These results are further discussed in Section 4.5.

4.5 Discussion

4.5.1 General Discussion

In a dual-task interleaving scenario with unequal task importance using the psychophysics approach, the impact of visual augmentation on the strategy in combination with the gaze direction distribution was examined. It was investigated how the time-lead of the onset of the informative property of motion-based augmentation relative to the suggested switch moment impacts the user's performance.

In a 3D virtual environment, a set of visual augmentations was generated aiming to prompt the user when it is more efficient to switch from one task to another, given a finite time of each trial. As the informative property of augmentation, a motion of a head-contingent peripheral visual stimulus was used. The mean duration on one task before switching to the other task, and the total performance, as well as the gaze direction distribution, were evaluated.

Overall, the experimental paradigm well captured a sub-optimal performance in a

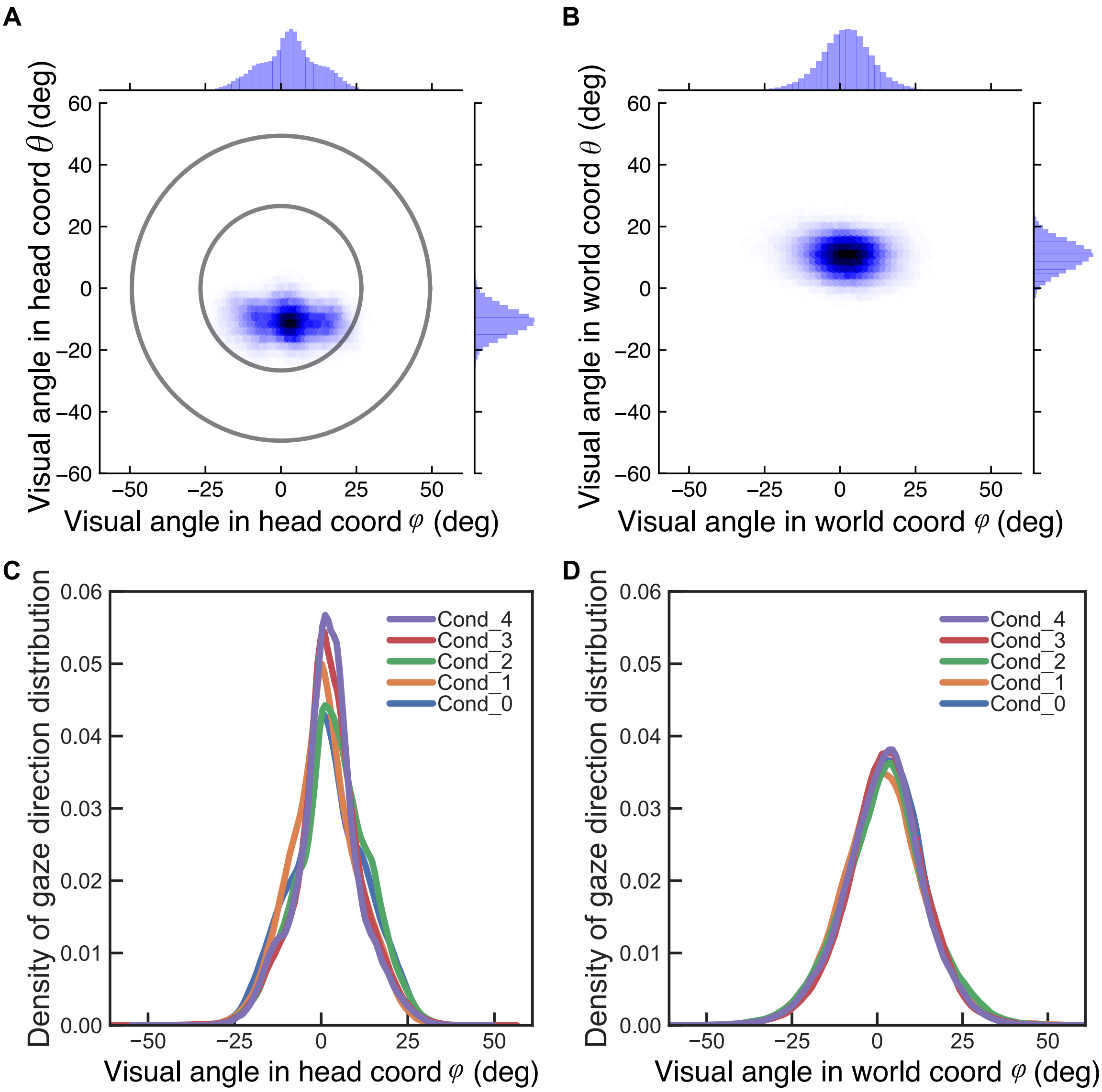

Figure 4.6: Gaze direction distribution: (**A**) distribution of the gaze direction in horizontal and vertical visual angle in head coordinates for one participant for one experimental condition (Condition 1). Data is visualized through a joint histogram plot with hexagonal bins where the depth of the color represents the count of data points in the corresponding region. The grey ring represents the interval of visual angles in head coordinates overlaid with the head-contingent augmentation ring in experimental conditions with augmentation present. (**B**) the same data set as in (**A**) but presented in world coordinates. (**C**) the density of the horizontal gaze direction distribution in head coordinates combined for participants. The curve describes the Gaussian kernel density estimate. Different colors represent different experimental conditions. (**D**) the density of the horizontal gaze direction distribution in world coordinates combined for participants.

dual-task interleaving scenario with unequal task importance. Similar to the results reported in [Farmer et al., 2018], in our experiment participants on average spent less time on the less important task before the switch which led to a lower total performance represented by the score. This is supported by significant differences between the baseline condition and the augmented conditions for both the mean duration on the blue task as well as the total score (see Figure 4.4).

4.5.2 Comparison of the Continuous and the Response-time-based Augmentation (Condition 1 vs Condition 4)

A significant difference in the mean duration on the blue task between Condition 1 and Condition 4 demonstrates that the response-time-based augmentation can be more beneficial performance-wise compared to the continuously present augmentation. This result was anticipated, as one would expect the participant to feel safe to stay longer on the blue task if it is known that he or she would be prompted with an exact optimal moment of the task switch (Condition 4). In Condition 1, even though a visual reference of the process was available to the participants, it was more difficult for the participant to stay the same amount of time on the blue task, when they still had to decide when it is long enough but not too lengthy (Condition 1). On the other hand, the analysis of the gaze direction distribution revealed differences between experimental conditions. We initially expected a bias of gaze direction towards the location of augmentation when augmentation was present, particularly, for the response-time-based augmentation due to its abrupt motion onset nature [Abrams and Christ, 2003]. In contrary to the original expectations, for Condition 1 and 4, no bias of the gaze towards the periphery was found, however, in Condition 4 the gaze direction was more confined towards the center of the visual field. The gaze direction distribution in the world coordinates was not found to be different in various conditions. This can be explained in the following way: when awaiting for an abrupt motion onset in the periphery in Condition 4, to gaze to a direction in the world coordinates further from the center of the room, participants preferred to rotate their head and keep the gaze closer to the center of the visual field instead of directly moving eyes and gazing further from the center. It is suggested, that as rapid detection of the augmentation motion onset was crucial in Condition 4, participants allocated more covert attention to the periphery of the visual field, thereby ensuring the fast motion detection by keeping the eccentricity of the visual stimulus stable. In contrast, in Condition 1, where instant motion onset detection was not crucial, participants pursued their natural gaze behavior similar to the baseline condition.

4.5.3 Comparison of the Individual and the Fixed Time-lead Based on Response Time (Condition 3 vs Condition 4)

The response-time-based augmentation was based on the individually measured response times for each participant, measured before the main experiment. However, keeping in mind a real-life application, a more universal augmentation design is more time-efficient avoiding excessive individual pre-measurements. Considering this, the third experimental condition was equivalent to the response-time-based condition where instead of using the individual response time as a reference for the motion onset of the ring sector, a fixed time-lead was used. In Condition 3, the mean duration on the blue task as well as the total performance was very similar to those in Condition 4. A negative trend of the correlation of the total performance in Condition 4 and 3, as a function of the individual response time, was found, however, it was not significant. Nonetheless, for the "slower" participants, who needed more time to detect the augmentation and switch to the red task, the performance was lower when the time-lead of augmentation was fixed to a value smaller than their response time, except one outlier with the largest response time (see Figure 4.5). The general performance in the main experiment of that outlier participant was, however, one of the best among the participants. It is possible, that the actual response time of that participant in the main experiment was smaller than the one measured in the original response time test. This would then result in a significant negative correlation in Figure 4.5. The gaze direction distribution in Condition 3 is similar to that in Condition 4, being more constrained towards the center of the visual field compared to conditions 0, 1, and 2. Even though in the current experiment no significant differences were found between the conditions with the fixed time-lead (Condition 3) and individual response-time-based condition (Condition 4), the individual abilities to use augmentation can play a role in how optimal in terms of performance the augmentation is. Thus, when designing augmentation, one should keep in mind potential individual differences.

4.5.4 Condition 2

Condition 2 was aimed to probe a time-lead of augmentation longer than the time-lead based on the response time, but shorter than the time-lead for the constantly present motion stimulus. In this condition, participants were expected to perform either similar to when the constantly moving augmentation was present (Condition 1), or a decrease in performance was expected. In particular, a lower performance was expected due to possible inhibition of return [Klein, 2000]. This term refers to a phenomenon when attention is maintained in a certain location for a while, but if no relevant information is shown in that location, attention is driven away, and the visual performance in the previously attended location is inhibited for a short period of time. We suggest that in Condition 2 of the current study, the participant's attention was likely to initially shift towards the augmentation. Thereafter, if the switch to another task hasn't occurred, after some time attention is shifted away from the augmentation, and the optimal moment of the switch could be potentially missed, leading to decreasing of the total score. From the results of Condition 2, few participants had a drop in performance possibly originating from the inhibition of return. However, on average the mean duration on the blue task as well as the total performance in Condition 2 did not significantly differ from other augmented conditions. It is suggested that the timescale of the inhibition of return effect is individually varied [Klein, 2000], and was potentially resolved only for few participants in the current experiment. Beyond the scope of the current study, further investigation of the impact of the inhibition of return is needed to test this hypothesis. The gaze direction distribution in Condition 2 also did not significantly differ from natural gaze behavior (Condition 0). It is suggested, that as participants had an additional second to use the augmentation while performing the blue task, a rapid motion onset detection was not crucial, thus, maintaining the gaze closer to the center of the visual field was not necessary in contrast to Conditions 3 and 4.

4.5.5 Conclusion and Outlook

To conclude, visual augmentation can support people's strategy in a challenging dual-task-interleaving scenario with the unequal importance of the tasks. In particular, the individual response-time-based augmentation with an abrupt peripheral motion stimulus onset, and augmentation with constantly present peripheral motion stimulus, showed advantage compared to when augmentation was not present. A visual augmentation based on the individual response time was more beneficial in terms of the mean duration on the less important task before switching to the more important one. However, when a rapid detection of peripheral augmentation was crucial, the natural gaze direction was altered being confined towards the center of the visual field indicating the allocation of attentional resources covertly towards the periphery. A continuously moving stimulus in Condition 1 also required attention, however, given a relatively long presentation time of the augmentation, it did not interfere with the gaze direction.

A possible confounding factor affecting performance is the general learning of the task, which could have impacted the performance level. From the trial-by-trial data evaluation, there was only a small and not significant learning effect, thus, the findings of the current study originate indeed mostly from the impact of augmentation. It is also important to note, that in the current experiment, the sub-tasks of the dual-task setting, even though implemented in a dynamic 3D environment, were relatively simple for the participants and involved mostly a motor component. In this task setting, visual augmentation did not significantly capture participants' attention as no reduction of performance was observed when augmentation was present. However, when the subtasks in a dual-task scenario are more cognitively demanding, capturing attentional resources by the augmentation can potentially lead to a drop in the total performance. Beyond the scope of the current work, future studies should address this question. It is also important to mention, that in this experiment, the parameters of the augmentation were selected based on the known dynamics of the task. In real life, the task environments are not always well understood and in those scenarios application of such an augmentation with predefined parameters would be challenging. Nonetheless, such an augmentation could find its application in well-understood environments, such as performing a routine task at work, where the strategy of the user would not be overruled by an algorithm but rather supported by a subtle augmentation system. Furthermore, while the proposed 3D VR-based paradigm is comparable to 2D screen-based experiments in terms of capturing performance in a task-interleaving scenario, it also permits a more natural experimental setting. In particular, it allows free head movement which enables evaluation of natural head and eye movements. This advantage as well as the possibility to implement dynamic interactive tasks makes virtual reality a powerful tool for studying task interleaving in future studies. The results of this study provide an insight into potential visual augmentation designs aiming to improve user's performance in a challenging dual-task interleaving setting.

4.6 Acknowledgments

This study was conducted with support of the Integrative Augmented Reality (I-AR) intramural funding of the University of Tuebingen, Germany.

4.7 Data Availability Statement

The data for all experiments are available at `https://osf.io/p23s7/` and experiment was not preregistered.

4.8 Authors' Biography

Olga Lukashova-Sanz is a physicist with a M. Sc. in physics with a specialization in solid state physics, and a M. Sc. with a specialization in medical technology. She is a PhD student at the ZEISS Vision Science Lab within the Ophthalmic Research Institute and ZEISS Vision Care Business Group.

Dr. Siegfried Wahl is a physicist with a Ph.D. in developmental biology of the visual system. He is Honorary Professor of the University of Tübingen and Principal Investigator of the ZEISS Vision Science Lab within the Ophthalmic Research Institute and ZEISS Vision Care Business Group.

Dr. Katharina Rifai is a physicist and did her Ph.D. in the field of Visual Neuroscience. She is a member of the ZEISS Vision Science Lab and Advanced Development department within Carl Zeiss Vision International GmbH.

Table 4.1: Experimental conditions.

Condition	**Time-lead**	**Description**
Condition 0	-	Control condition, no augmentation was present.
Condition 1	5000 ms	The motion onset of the ring sector was initiated instantaneously once the participant entered the blue task. The motion was continuous throughout the whole time spent on the blue task.
Condition 2	2200 ms	The motion onset of the ring sector was initiated later while the participant was performing the blue task, however, there was sufficient time before the sphere explosion for the participant to switch not immediately after detecting the motion.
Condition 3	1200 ms	The motion onset of the ring sector was abruptly initiated with a constant time-lead of 1200 ms until the red sphere explosion. This value was estimated from the response time test (see Section 4.3.3.6), where the response time is defined as the time that the participant needed to detect the motion onset, switch to the red room, and touch the red sphere. The value of 1200 ms was selected as approximately an average among the participants.
Condition 4	Individual response times	The motion onset of the ring sector was abruptly initiated with an individual time-lead previously recorded in the response time test (see Section 4.3.3.6). The response time is defined as the time that the participant needed to detect the motion onset, switch to the red room, and touch the red sphere.

5. Saliency-aware subtle augmentation improves human visual search performance in VR

Lukashova-Sanz, O., and Wahl, S. (2021). Saliency-aware subtle augmentation improves human visual search performance in VR. *Brain Sciences*, 11(3):283

5.1 Abstract

Visual search becomes challenging when the time to find the target is limited. Here we focus on how performance in visual search can be improved via a subtle saliency-aware modulation of the scene. Specifically, we investigate whether blurring salient regions of the scene can improve participant's ability to find the target faster when the target is located in non-salient areas. A set of real-world omnidirectional images were displayed in virtual reality with a search target overlaid on the visual scene at a pseudorandom location. Participants performed a visual search task in three conditions defined by blur strength, where the task was to find the target as fast as possible. The mean search time, and the proportion of trials where participants failed to find the target, were compared across different conditions. Furthermore, the number and duration of fixations were evaluated. A significant effect of blur on behavioral and fixation metrics was found using linear mixed models. This study shows that it is possible to improve the performance by a saliency-aware subtle scene modulation in a challenging realistic visual search scenario. The current work provides an insight into potential visual augmentation designs aiming to improve user's performance in everyday visual search tasks.

5.2 Introduction

Visual search is one of the most common tasks in everyday life, be it when a person is looking for a friend in a crowd or when a doctor is analyzing an optical coherence tomography (OCT) scan from a patient [Branchini et al., 2012]. Search becomes more challenging when the time to find the target is limited. For example, when a person is searching for the keys right before leaving the house, or when a surgeon is performing a meticulous manual task during the surgery using digital surgical microscope images

[Wahl et al., 2019]. In this study we focus on how performance in a visual search under limited time conditions can be improved.

An extensive amount of research has been done to investigate how people are searching for a target among distractors, and which neural mechanisms are laying behind, where numerous search task paradigms have been implemented (for reviews, see [Chan and Hayward, 2013; Wolfe, 2020; Verghese, 2001]). The difficulty of the visual search task depends on various factors, including how similar the target and the background are, how distinct the target is from the distractors, how complex the scene is, whether the observer has seen the scene already before, and many other aspects [Chan and Hayward, 2013; Wolfe, 2020; Verghese, 2001]. The human capacity to process visual content is limited, and mainly, in complex searches, it is crucial to select and prioritize visual information to complete the task. Attention is one mechanism that supports visual search and enables the searcher to find the target more efficiently [Posner, 1980; Ferrante et al., 2018; Jiang, 2018]. The shift of attention is associated with eye movements and fixating attended locations [Verghese, 2001; Borji et al., 2014]. Among numerous aspects impacting guiding of attention in visual search is stimulus-driven saliency of different elements of the visual scene. Specifically, the observer's attention can be attracted by salient distractors, even though from a goal-oriented perspective, they are irrelevant. To which extent saliency plays a role when it comes to visual search strategy has been long debated. Some studies showed that top-down mechanisms primarily drive visual search strategy, where fixation density can be explained by saliency only for the first few fixations [Chen and Zelinsky, 2006; Henderson et al., 2007; Rothkegel et al., 2019].

Other studies demonstrated that salient goal-irrelevant distractors could attract the observer's attention, slowing down the search [Jung et al., 2019; Bertleff et al., 2017; Foulsham and Underwood, 2011; Theeuwes, 2004]. The large variability of the results reported in the literature supports the notion that a combination of factors affects human attention guiding when looking for a target [Wolfe and Horowitz, 2017]. Here we approach visual search assuming that salient regions can attract attention in visual search.

When it comes to saliency, there are many different definitions that one meets in the literature. A salient region, in its broad context, is an area of the visual scene. It has a high contrast with its surroundings in one or multiple feature dimensions, be it color, shape, spatial frequency, speed of motion, contextual meaning, location within the visual field of view. Even though saliency is often associated with bottom-up low-level feature contrast [Itti and Koch, 2000; Bahmani and Wahl, 2016], both aspects are connected, but not identical [Schütt et al., 2019b]. Therefore, a part of a scene is considered salient if it is likely to attract the observer's attention, which is usually accompanied by fixating on that region.

The vast majority of existing knowledge on visual search is based on experimental studies conducted on conventional 2D-screens often using not-realistic synthetic stimuli as search arrays in a controlled environment. The search behavior in such artificial conditions can differ from real-life scenarios. A group of studies investigated human's visual behavior in a search task using naturalistic scenes displayed in a 2D-screen [e.g. Nuthmann, 2013; Li et al., 2016; Rothkegel et al., 2019; Cajar et al., 2020; Drewes et al., 2011; Eckstein et al., 2017]. The fast development of modern technologies such as Virtual Reality (VR) and VR eye tracking enabled researchers to study visual search in more realistic 3D environments [Boettcher et al., 2018; Olk et al., 2018]. In order to increase the level of immersion of the experimental paradigm, in the current study, we focus on real-world static visual scenes displayed in virtual reality, ensuring free body and head movement.

New technological tools, such as augmented reality (AR), enable the use of additional visual cues to purposefully guide user's attention and improve user's performance [Dey et al., 2018; Coughlan and Miele, 2017; Booth et al., 2013]. On the other hand, one drawback of augmenting visual input with additional content is that it introduces a trade-off of the potential benefit between performance and overlaying real visual scene with an additional layer of information. That, in turn, captures attentional resources, which essentially becomes a bottleneck for the design of visual augmentation [Raja and Calvo, 2017]. In this study, we hypothesize that by subtly modifying the visual scene, it is possible to drive the observer's attention away from salient distracting locations, enabling the user to find the target faster. Previously, some attempts were made to apply subtle visual content modification for gaze guidance, where color, luminance, spatial frequency, and other domains were modulated [Gatys et al., 2017; Lu et al., 2012; Grogorick et al., 2017; Biocca, Frank and Owen, Charles, and Tang and Corey, 2007; Lu et al., 2014; Bailey et al., 2009; Danieau et al., 2017; Pomarjanschi et al., 2012; Sridharan et al., 2015; Lin et al., 2017]. In this work, blur was selected as a domain for modifying visual content, as it was previously shown that blur, although with limitations, can be used for gaze guidance to an extent where the observer does not even notice the modification [Hata et al., 2016; Ueda et al., 2019; Khan et al., 2011].

Furthermore, the idea of slightly defocusing parts of the scene for triggering the observer to fixate on more clear locations is widely used in photography and cinematography [Enns and MacDonald, 2013; Yamaura et al., 2018; Huynh-Thu et al., 2013]. Also, Sitzmann et al. [2018a] proposed blurring images based on saliency to downsample the resolution of the non-salient regions for further image compression. In contrast to their approach, the strongest blur was applied to the most salient regions in this study.

This study investigates whether blurring salient regions of the visual scene, which would otherwise likely attract the observer's attention, can improve visual search task's performance, enabling the observer to find the target faster. Using a psychophysical approach, we evaluate observers' ability to locate the target within a limited amount of time. Eye-tracking data was recorded to support the results. Specifically, fixations were analyzed.

5.3 Materials and Methods

5.3.1 Participants

Twenty naïve participants (14 female and 6 male), with normal or corrected to normal vision. Participants were aged between 20 and 38 years old. All procedures conformed to Standard 8 of the American Psychological Association's "Ethical Principles of Psychologists and Code of Conduct (2010)". The study was approved by the ethics committee of the Faculty of Medicine at the University of Tübingen with a corresponding ethical approval identification code 138/2017b02. Signed informed consent was obtained from each subject before the measurements. All data were stored and analyzed in full compliance with the principles of the Data Protection Act GDPR 2016/679 of the European Union.

5.3.2 Experimental setup

5.3.2.1 Hardware specifications

The visual content was displayed to the participant using HTC Vive Pro Eye (HTC Corporation, Taoyuan, Taiwan) virtual reality headset running on a Windows 10 PC with NVIDIA GeForce GTX 1070 graphics card (NVIDIA Corporation, Santa Clara, California, USA). The field of view of the headset and the refresh rate reported by the manufacturer are 110° and 90 Hz, respectively. The participant interacted with the environment via the HTC Vive controller. The position and rotation of the headset and the controller were tracked via the HTC base stations 2.0. The eye-tracking data was collected using a built-in eye tracker at a frequency of 120 Hz.

5.3.2.2 Software specifications

The experimental paradigm was generated using the Unity Game engine [Unity Technologies, 2019], Unity version 2019.3.15.f1. The eye movement data was collected using

Unity package SRanipal version 1.1.0.1. Recording of the eye movement data at a maximum sampling rate 120 Hz was realized by means of using a separate thread parallel to the main script execution [Corporation, 2020]. The data analysis was performed using Python 3.6 packages NumPy [Van Der Walt et al., 2011] version 1.19.1, SciPy [Virtanen et al., 2020] version 1.5.2 and Pandas [Mckinney, 2010] version 1.1.3. The statistical analysis was conducted using R [R Core Team, 2020] version 3.6.1, in particular, package lme4 [Bates et al., 2015]. The data visualization was performed using Python packages Matplotlib version 3.3.1 [Hunter, 2007] and Seaborn [Waskom et al., 2017] version 0.11.0.

5.3.3 Virtual environment and stimuli

5.3.3.1 Real-world scenes

The virtual environment was composed of omnidirectional images displayed in virtual reality (VR) by back-projecting it to the Skybox sphere (Figure 5.1 (C)). An omnidirectional image is a 360-degree panoramic image. An equirectangular projection is one way to represent an omnidirectional image, where the aspect ratio of the projection is 2:1. The horizontal and vertical coordinates of the projection are polar (ϕ) and azimuthal (θ) angles, respectively, where ϕ can be in the range from 0° to 360°, and θ ranges from 0° to 180° (Figure 5.1 (A)). For the main experiment, 24 scenes were selected from the Salient360! Training dataset with diverse content: indoor/outdoor, day/night time, containing people/not containing people, etc. (see full set of images in Supplementary information Figure S3 and Figure S4). The resolution of the images ranged from 3000×1500 to 10000×5000 pixels with mean resolution of $(5878 \pm 2443) \times (2939 \pm 1222)$ pixels. For the training phase, five scenes from the Salient360! Training dataset were selected, different from the scenes of the main experiment (see Supplementary information Figure S2).

5.3.3.2 Saliency maps

To evaluate which regions of the scenes are likely to attract attention, saliency prediction models are widely used [Borji, 2019]. In recent years interest in saliency models for 360-degree omnidirectional images largely grew [Xu et al., 2020]. In the current study, the saliency maps used for spatial modulation of the omnidirectional images were obtained using the method described in [Startsev and Dorr, 2018] which won in the “Head and Eye Movement Prediction” category of “Salient360!” Grand Challenge at ICME’2017 [University of Nantes and Technicolor, 2017]. Authors of the method proposed composition of the continuity-aware and the cube map approaches using a combination of saliency predictors, with applied equator bias (for details, see [Startsev

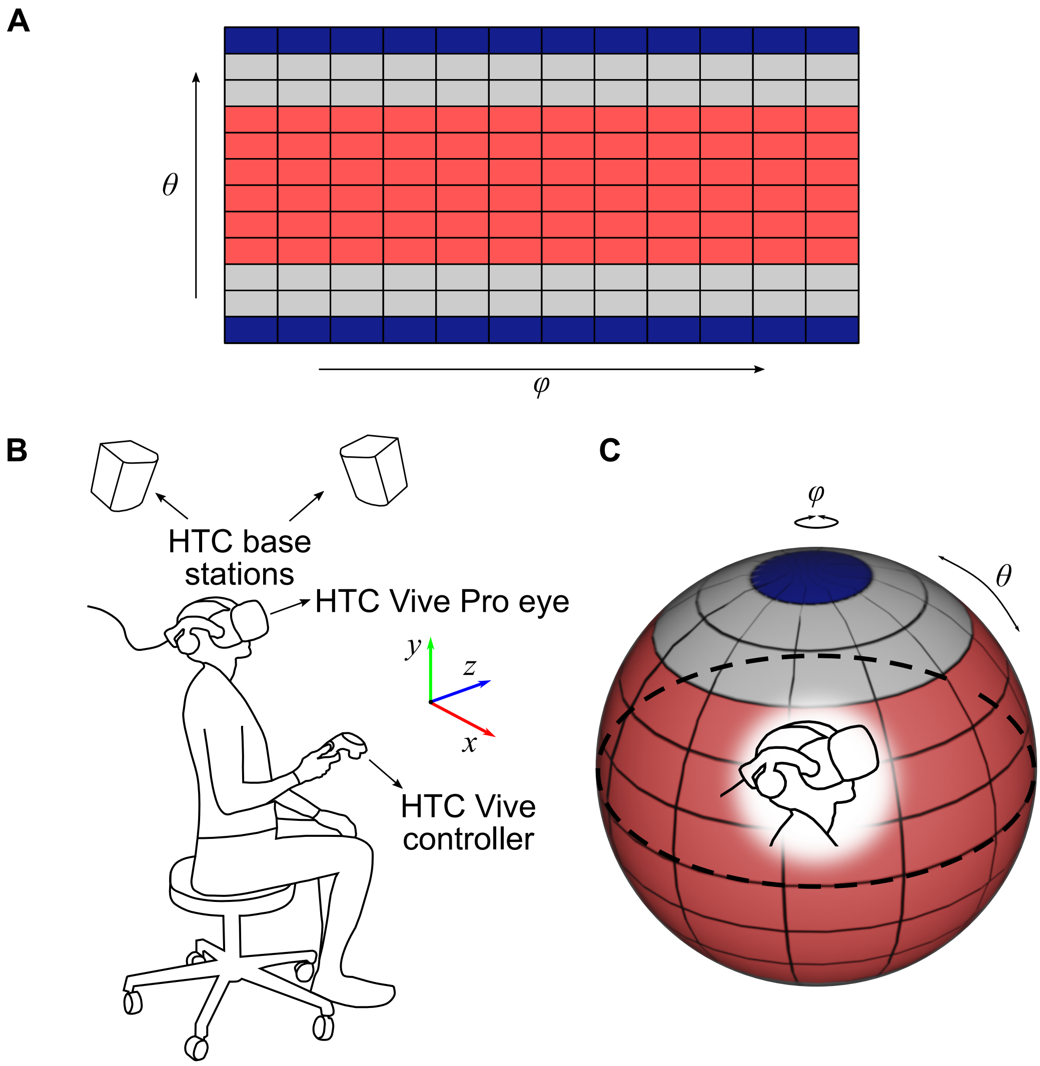

Figure 5.1: **(A)** A schematic representation of equirectangular projection. Each rectangle is 30° wide and 15° high. **(B)** The experimental setup. The experimental paradigm was implemented in Unity software and displayed in a VR headset — HTC Vive Pro Eye with a built-in eye tracker. Position and rotation tracking of the headset was realized via the HTC base stations 2.0. The participant performed the experiment in a seated position on a rotating stool. As soon as the search target was located, the participant pressed a button on the HTC Vive controller. Additionally, a left-hand coordinate system is shown, which is used to set the Unity world coordinates. For details on the experimental procedure see Section 5.3.4.1. **(C)** A schematic representation of a sphere to which the equirectangular projections (omnidirectional images) were back-projected in Unity as a Skybox. The colors indicate like-colored locations on the equirectangular projection in **(A)**.

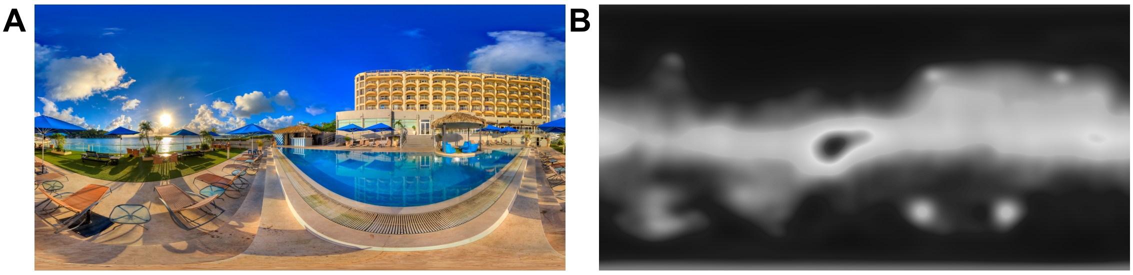

Figure 5.2: **(A)** Example of one omnidirectional image used in the main experiment. **(B)** The corresponding saliency map generated using the approach in [Startsev and Dorr, 2018]. White regions correspond to the most salient areas, whereas black color indicates the least salient locations.

and Dorr, 2018]). The gray-scale saliency maps used for the main experiment can be found in Supplementary information Figure S5 and Figure S6. The pixel color values of the saliency maps ranging from 0 (black) to 1 (white) are referred to as saliency values. In Figure 5.2 an example of one scene together with its saliency map is shown.

5.3.3.3 Blurred images

The omnidirectional images were modified by applying blur. The strength of blur was spatially varied based on the corresponding saliency map. In particular, each blurred image resulted from a convolution of the respective original image and a two-dimensional Gaussian kernel. The kernel's size was fixed to one degree, whereas the standard deviation σ of the kernel was varied, determining the blur strength. A set of blurred 360-degree images using two different values of the Gaussian kernel standard deviation was generated. In particular, $\sigma_1 = 17\%$ of kernel size, and $\sigma_2 = 34\%$ of the kernel size, were used defining two of the experimental conditions (Figure 5.3). At a given experimental condition, the blur strength for each pixel was weighted with its saliency value. Specifically, the standard deviation of the Gaussian kernel for each pixel was computed by multiplying the standard deviation corresponding to the experimental condition (σ_1 or σ_2) with the respective saliency value of that pixel (values from 0 to 1, see Section 5.3.3.2). The size of the kernel in pixels, as well as the standard deviation, were scaled considering the individual image size in pixels. An example of a blurred part of an image is shown in Figure 5.3.

5.3.3.4 Search target

As a search target, a Gabor cross was used. Similar to [Rothkegel et al., 2019], a Gabor cross is a sum of vertical and horizontal Gabor stimuli, where each of the Gabors is a product of 8 cyc/° cosine function with Gaussian envelope with standard deviations of 0.06° and 0.32° (Figure 5.4). The total width and height of the cross was approximately

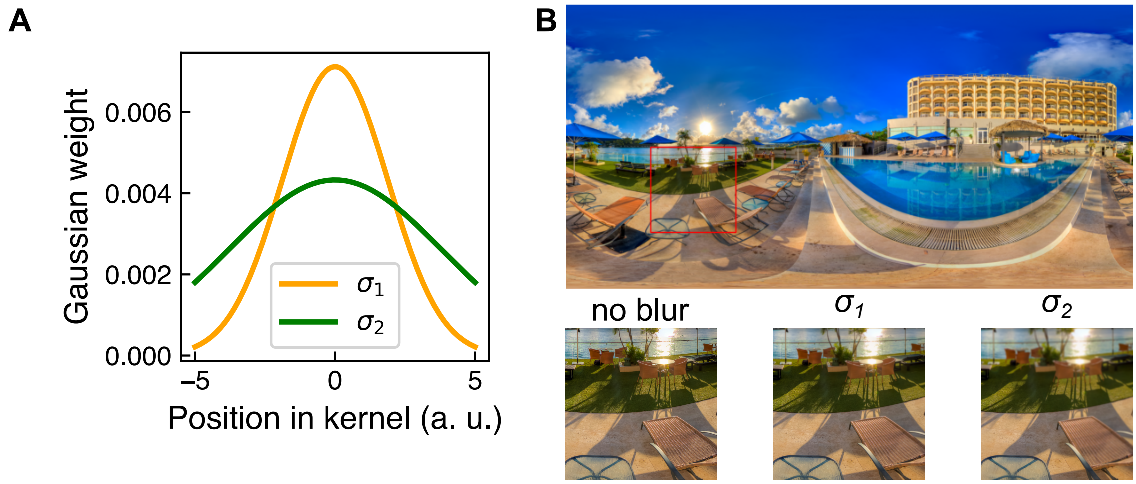

Figure 5.3: **(A)** A cross-section of a 2D Gaussian kernel used to blur images, with two standard deviations: σ_1 = 17% of kernel size, and σ_2 = 34% of the kernel size. The kernel size was fixed to a number of pixels corresponding to 1° visual angle. For schematic illustration, the kernel size in the plot equals 10 a. u. **(B)** Example of blurred part of one scene generated as a convolution of respective 2D Gaussian kernel and original image, considering its saliency map. In the lower row of **(B)**, a part of the original image which is indicated by a red rectangular is shown in three blur conditions: no blur, Gaussian blur with maximum standard deviation σ_1, and Gaussian blur with maximum standard deviation σ_2.

3° visual angle. The target was overlaid with the visual scene by positioning it in Unity at a pseudorandom location on top of the omnidirectional image (see Section 5.3.3.5). To blend the target with the omnidirectional image, the transparency of the cross was decreasing towards the edges of the Gabor cross at a rate of the Gaussian envelope used to generate the cross.

5.3.3.5 Search target locations

In each trial, the search target was positioned at a pseudorandom location defined by spherical coordinates (r, ϕ, θ). Specifically, r was set to a fixed value, whereas ϕ and θ were varied. Prior to the experiment, a set of nine possible target locations (ϕ, θ) was generated for each visual scene, where ϕ could be in the range [0°, 135°] and [225°, 360°], and θ in the range [45°, 135°]. This range was selected to ensure that subjects do not have to rotate their head in an uncomfortable position trying to look too much up or down. Additionally, taking into account that in two experimental conditions the visual scenes were spatially blurred based on the corresponding saliency values, the regions with low saliency were selected as possible search target locations to avoid displaying the target on very differently blurred backgrounds in different experimental conditions. Specifically, possible (ϕ, θ) were limited to regions with low saliency for each visual scene — locations with saliency values under 17% of maximum saliency

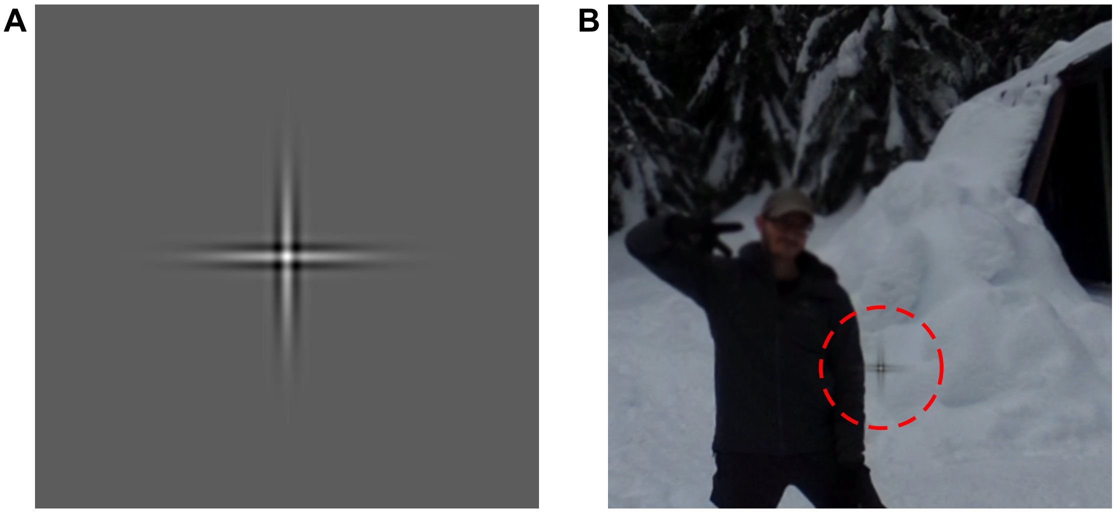

Figure 5.4: **(A)** The Gabor cross was used as a search target in the visual search task. It is a sum of vertical and horizontal Gabor stimuli, where each of the Gabors is a product of 8 cyc/° cosine function with Gaussian envelope with standard deviations of 0.06° and 0.32°. **(B)** An example of the search target is overlaid with a visual scene at a pseudorandom location. For simplicity, the cross is highlighted by a surrounding red dashed line. In the experiment, the red line was not present.

value on corresponding saliency map. During the experiment, each participant had the same set of pseudorandom locations for corresponding visual scenes and respective conditions. The size of the search target was was kept constant in visual angle in all visual scenes. The distribution of the final set of possible search target locations for all visual scenes can be seen in Supplementary information Figure S1.

5.3.4 Experimental procedure

5.3.4.1 General procedure

Each participant performed a visual search task wherein each trial they had to search for the search target located at a pseudorandom location within a limited amount of time set to 20 s. Participants were instructed to find the target as fast as possible, naturally moving their head and gaze. Once the target was found, participants were asked to fixate the target and press a button on the controller. As soon as the button was pressed, the trial was terminated. Participants performed the experiment in a seated position on a rotating stool enabling free head movement (Figure 5.1 (B)). The experiment was performed in a single one-and-a-half-hour experimental session. First, each participant had a training phase where they got acquainted with the virtual environment and the search target. Next, the main part of the experiment was conducted. Every time the participant put on the headset, a five-point built-in calibration procedure of the eye tracker was performed. Each scene was explored starting from the same spatial point $\phi = 180°, \theta = 90°$. To do so, before starting a new trial, participants had to fixate a reference fixation point of size 1° visual angle on a gray background located in the corresponding direction. The gaze position was controlled using the eye tracker. Once the gaze position condition was satisfied, the trial was started. That is, the scene with the search target was onset.

5.3.4.2 Training phase

The training phase consisted of 10 trials with no time limit. In each trial, one of five different visual scenes was shown with the search target overlaid on top of the scene. Thus, each scene was presented twice. The scenes used in the training phase were different from the ones used in the main experiment (see Supplementary information Figure S2). By the end of the training phase, all participants verbally responded that the task was straightforward and that they felt comfortable to continue with the main experiment.

5.3.4.3 Main experiment

In the main experiment, participants were instructed to perform the search task the same as during the training phase. Furthermore, they were informed that some scenes are partly blurred, but that it does not affect the task instructions. Participants were also informed that each trial's time is limited to 20 s, but that in some trials, it is too difficult to find the target in the background. Thus, it is normal not to find the target within the given time in some trials. Each trial finished pressing a button if the target was found or after the time limit was reached. The fixation on the target was not controlled during the trial. The main experiment was split into three blocks with short three-to-five minutes breaks. During each break, the participant took off the headset and rested. Before starting each next block, the eye tracker calibration procedure was conducted. Each block consisted of 72 trials. All 24 scenes were presented in each block three times: as an original image with no blur, blurred with standard deviation σ_1, and blurred with standard deviation σ_2. The blur strength determined three experimental conditions: Condition 0 (no blur), Condition 1 (blurred with σ_1), and Condition 2 (blurred with σ_2). During each block, all three conditions were presented in random order. Each block's duration varied depending on how long it took the participant to find the target in each trial, ranged between 10 and 20 minutes. During the main experiment, each participant performed a total of 216 trials divided into three blocks, with each block containing the three conditions above-mentioned for every scene (24). For each scene, the search target location was always different selected from the set of nine possible locations (see Section 5.3.3.5).

5.3.5 Analysis

The behavioral data as well as the eye movement data were analyzed using Python as well as lme4 library of R, and compared using linear mixed model analysis [Baayen et al., 2008].

5.3.5.1 Behavioral performance metrics

The *search time* and the *proportion of missed trials* were selected as the main behavioral performance metrics. The search time is defined as the time since a trial started until the button was pressed. Thus, it is a continuous variable ranging between 0 ms and 20 000 ms. Only trials where the target was found were used to compute the search time. The proportion of missed trials is computed as the number of trials where the target was not found within 20 s divided by the total amount of trials in the respective experimental condition. As for each condition, there were 72 trials from each participant. The proportion of missed trials is a discrete variable with a minimum step of $1/72 = 0.014$. Better performance is defined by a shorter search time as well as a lower proportion of missed trials. The impact of blur was estimated by fitting the linear mixed models to the data, fitting a separate model for each of the metrics, where *search time* and *proportion of missed trials* are dependent variables, *blur condition* is fixed effects, and *subject* is a random factor. Furthermore, to evaluate learning throughout the experiment, the *block number* was introduced as an additional fixed effect in the model. To further explore whether the differences between blur conditions evolve with time, an extension of the model could be done where one could account for possibly different slopes in data subgroups at a given blur condition and given block number. However, the limited amount of data points collected in this study does not support a more complex model. Therefore, a proposed linear mixed model with two fixed effects and one random factor was selected. An extensive quantitative analysis of performance evolution over the course of time is out of the scope of the present work and is a subject of future studies

5.3.5.2 Eye movement metrics

As a secondary set of metrics to characterize visual search performance, the number of fixations until the target was found (*NumOfFix*), the proportion of fixations within the area of interest (*PropFixInAOI*), and duration of fixations (*FixDuration*) were computed. The area of interest (AOI) is defined, similar as in [Sitzmann et al., 2018b], as 5% most salient pixels of the original image based on its saliency map (see example in Figure 5.5). The proportion of fixations within the area of interest is defined as the number of fixations within the area of interest divided by the total number of fixations

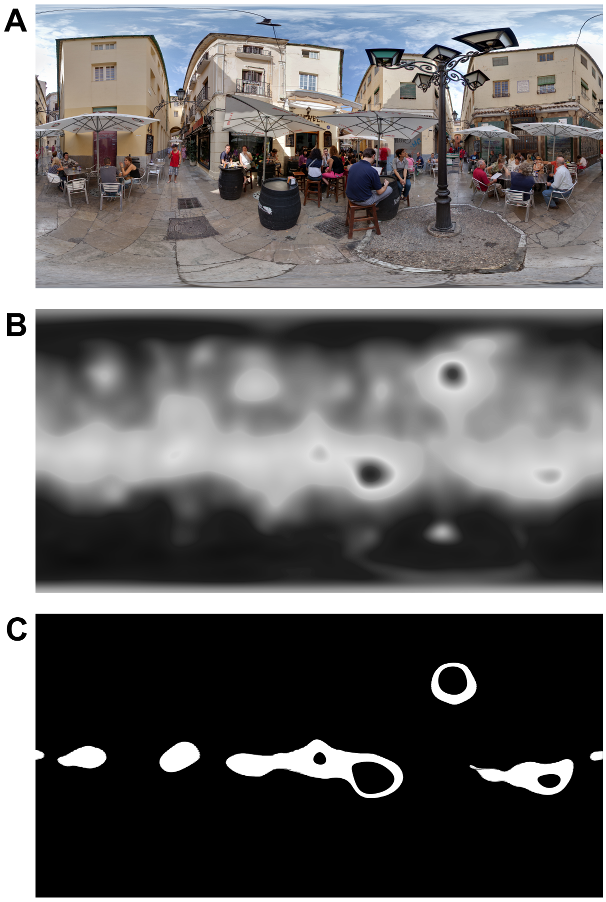

Figure 5.5: **(A)** Example of one omnidirectional image, **(B)** its saliency map in gray scale, and **(C)** selected 5% most salient pixels indicated by white color. The area of interest is the white region in **(C)**.

for each trial. The number of fixations until the target was found was defined as the total number of fixations in each trial where the target was found. Finally, the fixation duration was computed as the duration of fixations in each trial. The effect of blur condition was evaluated for each metric by fitting a separate linear mixed model where *NumOfFix*, *PropFixInAOI*, and *FixDuration* are dependent variables, *blur condition* is fixed effect, and *subject* is a random factor. To account for possible learning throughout the experiment, the *block number* was introduced as an additional fixed effect in the model.

5.3.5.3 Eye movement raw data pre-processing

The eye movement data were recorded at a frequency 120 Hz. The gaze position data was accessed using a customized written Unity script utilizing the HTC SRanipal SDK package functions. The time variable was taken using the system time, as it was previously shown that the time function from SRanipal package returns not always reliable values [Imaoka et al., 2020]. In Table 5.1 the main recorded variables are described. All variables were recorded for left and right eyes.

Table 5.1: Main eye- and head-movement-related raw variables recorded during the experiment.

Variable	**Units**	**Meaning**
Timestamp	any integer number	The system time in ms at the moment of sample recording.
Eye data validity bit mask	an integer from 0 to 31	Indicates the validity of the data. A value of 31 indicates the highest validity of the recorded data. This parameter is used to filter the raw data where the eye tracker lost the pupil, including filtering blinks.
Gaze normalized direction vector	A three-coordinates vector *(x, y, z)* with each coordinate ranging from -1 to 1	A gaze vector indicating the direction of gaze in the headset right-hand coordinate system. To convert it to the left-hand coordinate system (Figure 5.1 (B)), the x coordinate was multiplied by -1.
Head rotation	a rotation quaternion *(x, y, z, w)* of head	A quaternion describing the rotation of the headset in Unity world coordinates. The position of the headset was always fixed to the origin (0, 0, 0).

To prepare the data for further processing, first, similar to [Imaoka et al., 2020], the raw data were filtered based on the eye data validity bit mask value, which represents the bits containing all validity for the current frame. After the filtering, only the data where the eye data validity bit mask had value 31 for both eyes, were selected. Doing so, the data where the eye tracker partly or completely lost the pupil (including blinks) was filtered out. Next, the gaze position was calculated in spherical coordinates. In particular, the polar ϕ and azimuthal θ angles were computed using Equation (5.1) and Equation (5.2). In Unity, the z-axis corresponds to the depth dimension.

$$\phi = \arctan \frac{x}{z}, \tag{5.1}$$

$$\theta = \text{arctan2}\,(y, \sqrt{x^2 + z^2}), \tag{5.2}$$

where (x, y, z) are coordinates of normalized gaze directional vector in headset coordinates. Note that SRanipal returns the gaze direction vector in the right-handed coordinate system. To convert the coordinates in the left-hand coordinate system (same as Unity world coordinate system, see Figure 5.1 (B)), the x-coordinate was multiplied by -1. To compute the gaze position in Unity world coordinate system, the gaze position in headset coordinate system was multiplied by the head rotation quaternion. In Figure 5.6 an example of gaze position for one subject in one trial in

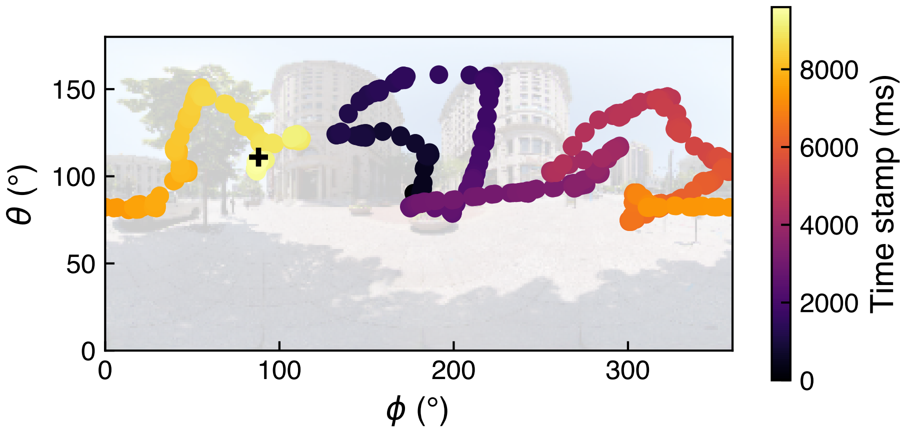

Figure 5.6: Example of raw gaze position in spherical coordinates for one subject in one trial. The background image is the visual scene shown in this particular trial — for easier visualization of the gaze positions, the omnidirectional image is washed out in the figure. The black cross with coordinates $(88°, 111°)$ indicates position of the search target in this trial. The color of the scatter points range from dark violet to light yellow and indicates the time stamp of each gaze position sample starting from the beginning of the trial: here, the participant started the search from the center of the image and ended around the target position. This trial lasted approximately ten seconds and the search target was successfully found by the participant.

spherical coordinates (ϕ, θ) is shown.

5.3.5.4 Fixation detection algorithm — I-VT

Fixations were identified using *velocity threshold algorithm for fixation identification* (I-VT) [Salvucci and Goldberg, 2000]. The algorithm was implemented following the description in [Kübler, 2020] and [Olsen, 2012]. The gaze velocity v was computed in °/s between each two consecutive samples (Equation (5.3)).

$$v = \frac{\sqrt{(\phi_i - \phi_{i-1})^2 + (\theta_i - \theta_{i-1})^2}}{t_i - t_{i-1}}, \tag{5.3}$$

where (ϕ_i, θ_i) and $(\phi_{i-1}, \theta_{i-1})$ are consecutive gaze positions in degrees visual angle in headset coordinates, and t_i and t_{i-1} are respective time stamps. To reduce the noise level of the data, a running average filter was applied with the window size of three samples which is $\sim 25\,\mathrm{ms}$. An eye movement was considered to be a fixation if the gaze velocity did not exceed a threshold 60 °/s [Leube et al., 2017]. Two fixations were merged in a single fixation if the time between them was under 75 ms [Komogortsev et al., 2010], and the angular distance was under 1° [Komogortsev et al., 2010; Over et al., 2007]. Too short fixation with a duration under 60 ms were filtered out [Komogortsev et al., 2010; Over et al., 2007]. In Figure 5.7 the eye movement data

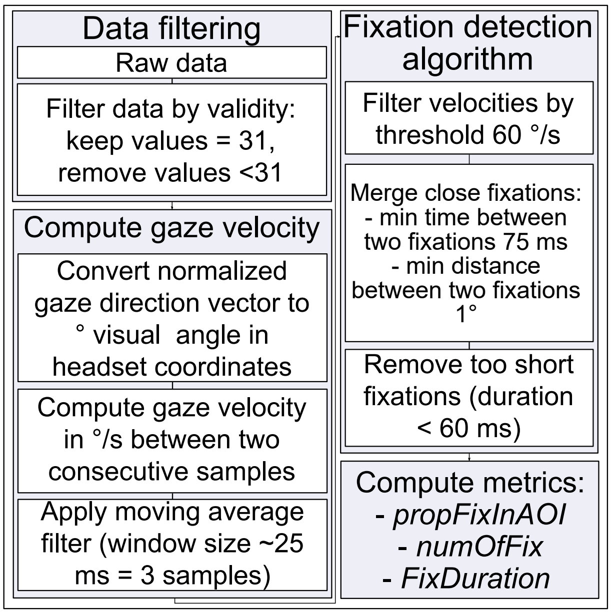

Figure 5.7: Eye movement data processing algorithm. For details see Section 5.3.5.4.

processing algorithm is summarized in a flow chart.

5.4 Results

5.4.1 Behavioral data

In Figure 5.8 the mean search time and the proportion of trials where the target was not found, estimated across all subjects, are shown. The mean value of the search time considering trials where the target was found was 8423 ms $\pm$ 1358 ms, 8146 ms $\pm$ 1258 ms, and 7670 ms $\pm$ 1550 ms for conditions 0, 1, and 2, respectively. The mean value of the proportion of trials where the target was not found, was 0.17, 0.11, and 0.10 for conditions 0, 1, and 2, respectively. The individual data for each subject can be found in Supplementary information Section S5. From linear mixed model analysis (see Section 5.3.5.1) over the course of all trials, a significant effect of blur was found for both behavioral metrics. Specifically, for the search time, a significant difference between the no-blur and σ_2-blur conditions was found with $p < 0.001$. The difference between the no-blur and σ_1-blur conditions was close to but not significant with $p = 0.09$. In terms of the proportion of trials where the target was not found, significant differences between the no-blur and σ_1-blur, as well as between no-blur and σ_2-blur conditions, were found with $p < 0.001$ for both conditions.

To check for the learning effect, the mean search time and proportion of missed trials were evaluated over the course of the experiment (see Supplementary Information Section S7, Figure S12(A, B)). A significant effect of block number was found for mean

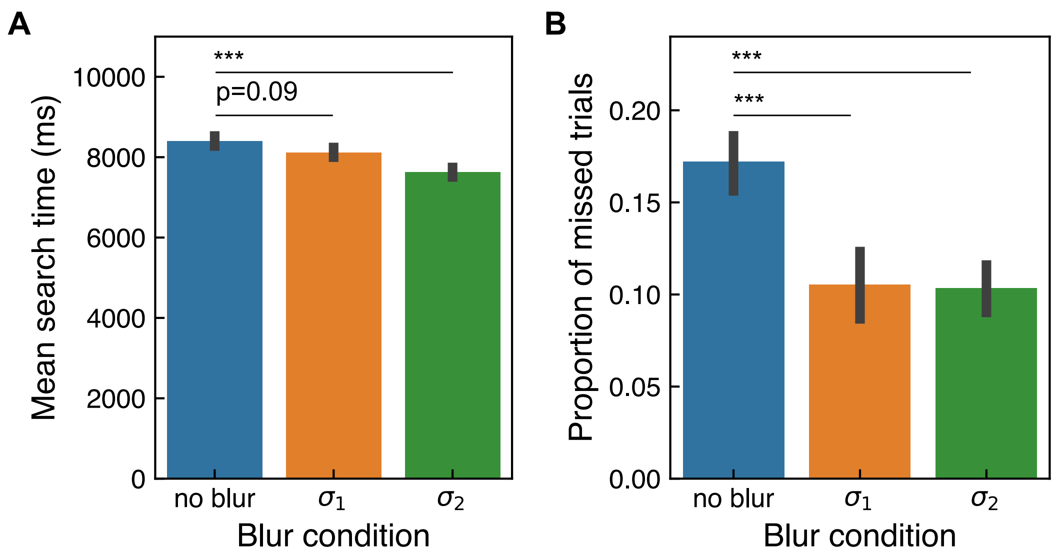

Figure 5.8: **(A)** An estimate of mean search time in each blur condition. The mean search time was computed for each subject by averaging the search time over the number of trials where the target was found. **(B)** An estimate of the proportion of trials where the target was not found in each blur condition. The bar plots represent data from 20 subjects. The error bars show the standard error of the mean. The indicators of significant differences obtained from the linear mixed model analysis are *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. For non-significant differences, the p-value is shown.

search time as well as for proportion of missed trials with $p < 0.001$. This can also be observed as a downtrend of both behavioral metrics for all three blur conditions indicating a general learning of the task. This is expected, as with time participants get more familiar with the VR headset, as well as get more acquainted with the visual scenes since the same images were presented in each experimental block. Despite general learning, the difference in performance metrics between the no-blur condition and other two blur conditions was observed already in the first block of the experiment, most prominently, for the proportion of missed trials.

5.4.2 Eye movement data

In Figure 5.9 (A) distribution of fixation duration is visualized as a kernel density estimate plot which is an alternative to a histogram plot. The plot represents all fixations computed for all subjects and all trials regardless of whether the target was found or not in a respective trial. The total number of detected fixations was 21504, 18059, and 16092 for the Conditions 0 (no blur), 1 (σ_1), and 2 (σ_2), respectively. The mean duration of fixation averaged over all subjects across all trials is 133 ms $\pm$ 76 ms. This value is comparable with typically reported fixation duration considering that in a visual search task, fixations are usually shorter than in a free-viewing task [Nuthmann et al., 2010]. In Figure 5.9 (B) an estimate of the number of fixations until the target was found ($NumOfFix$) is shown. To estimate $NumOfFix$ only trials where the target was found were considered. The mean values of $NumOfFix$ are 11.9, 10.8, and 9.6 for Conditions 0, 1, and 2, respectively. Figure 5.9 (C) demonstrates an estimate of the proportion of fixations within the area of interest ($PropFixInAOI$). To estimate $PropFixInAOI$, all trials were considered regardless of whether the target was found or not in the respective trial. The mean values of $PropFixInAOI$ are 0.23, 0.20, and 0.19 for Conditions 0, 1, and 2, respectively. Finally, in Figure 5.9 (D) an estimate of fixation duration ($FixDuration$) is shown. To compute the estimate of $FixDuration$ all fixations were considered regardless of whether the target was found or not and whether a fixation was within AOI or not. See details of metrics' definitions in Section 5.3.5.2. The individual data for each subject can be found in Supplementary information Section S6.

From the linear mixed model analysis (see Section 5.3.5.2) over the course of individual trials and fixations, a significant effect of blur was found for $NumOfFix$ and $PropFixInAOI$ for all blur conditions. Specifically, for both metrics ($NumOfFix$ and $PropFixInAOI$), a significant difference between the no-blur and σ_1-blur, as well as between the no-blur and σ_2-blur conditions, were found with $p < 0.001$ for both conditions. For $FixDuration$ a significant difference was found between the no-blur and σ_2-blur conditions with $p < 0.05$. The $FixDuration$ difference between the no-blur and σ_1-blur conditions was not significant with $p = 0.3$. The differences in $FixDuration$ only for fixation within AOI were not significant with $p = 0.61$ and $p = 0.16$ for Condition 0 vs Conditions 1, and Condition 0 vs Condition 2, respectively.

Evolution of $PropFixInAOI$ was assessed across the time course of the experiment to check for learning (see Supplementary information Section S7 Figure S12(C)). A significant effect of block number was found with $p < 0.001$. This indicates a general learning effect across three blocks of measurements which as mentioned in the previous section, can be expected due to familiarizing of subjects with the setup and the scenes. Nonetheless, the proportion of fixations in the AOI in the blurred conditions compared to the non-blur condition appears to be smaller already in the first block of the experiment.

5.5 Discussion

In a visual search task, it was investigated whether blurring salient regions of the visual scene can drive the observer's attention away from those regions and enable the observer to find the target faster. Using a psychophysics approach implemented in

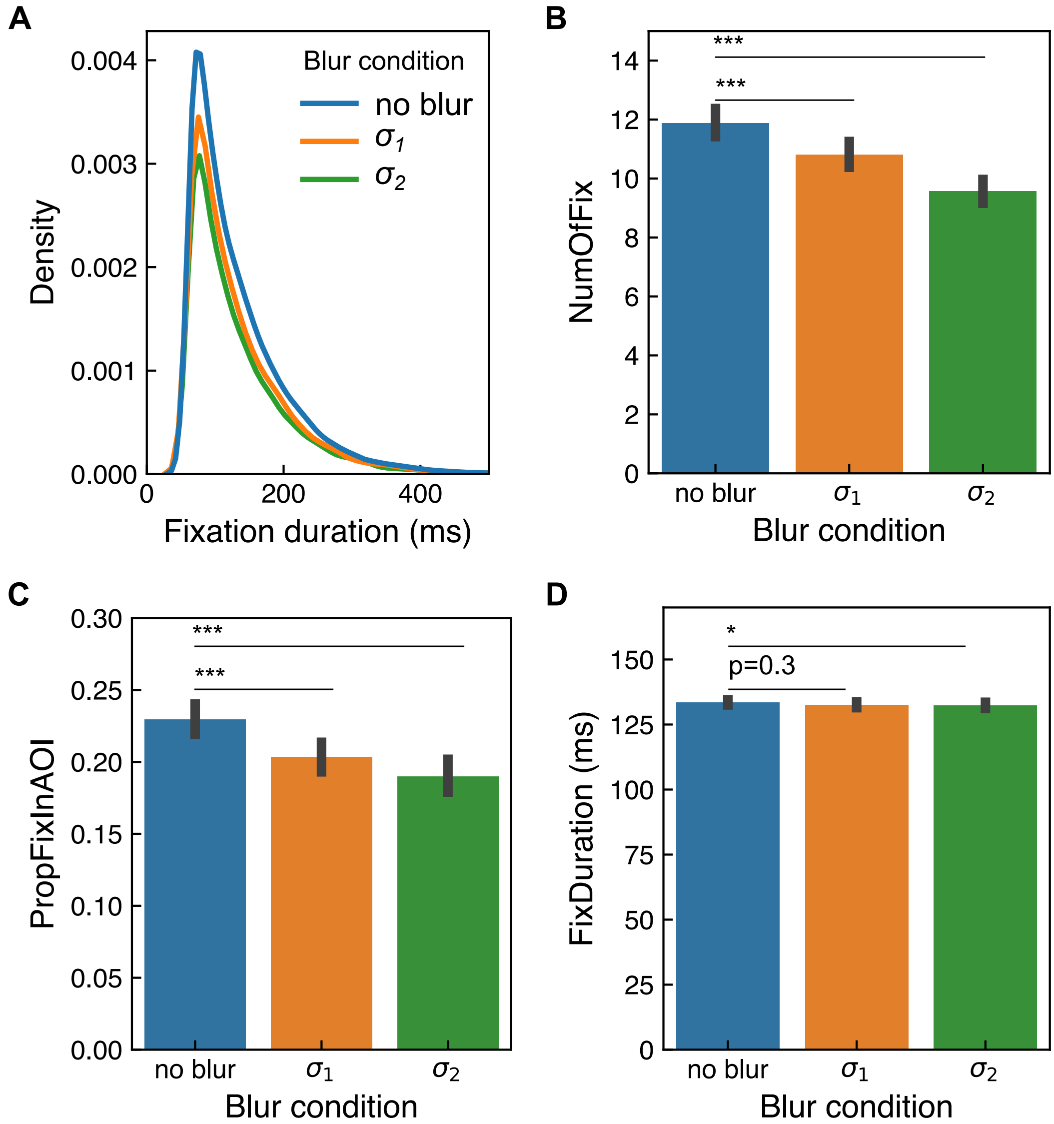

Figure 5.9: **(A)** A distribution of fixation duration as a kernel density estimate plot. Only fixations with duration up to 500 ms are shown to better resolve differences between different blur conditions. The mean duration of fixation averaged over all subjects across all trials is 133 ms $\pm$ 76 ms. The plot represents all fixations computed for all subjects and all trials. Each curve is normalized to the number of observations such that the total area under all densities sums to 1. **(B)** An estimate of the number of fixations until the target was found in each blur condition. The bar plots represent data from 20 subjects considering only trials where the target was found. **(C)** An estimate of the proportion of fixations within the area of interest in each blur condition. The bar plots represent data from 20 subjects considering all trials. **(D)** An estimate of fixation duration in each blur condition. The bar plots represent data from 20 subjects considering all trials. The error bars show the standard error of the mean. The indicators of significant differences obtained from the linear mixed model analysis are *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. For non-significant differences, the p-value is shown.

VR, we evaluated the observer's ability to find the target in a real-world scene within a limited amount of time, as well as how long it took the observer to find the target. The eye movement data were evaluated by accounting for the number of fixations within the most blurred areas and the total number of fixations until the target was found. Similarly, the mean duration of fixations were computed for a non-blur and two blur conditions.

Overall, the experimental paradigm captured well a challenging visual search task in a realistic 3D environment. The possibility of freely moving head and gaze brought the controlled experimental setting closer to a real-life scenario compared to traditional screen-based paradigms.

The significant decrease found in the search time in conditions when salient regions were blurred shows that modulating the visual scene by spatially applying saliency-aware blur can lead to a faster locating of the target. In particular, in the condition with the strongest blur (σ_2), the gain in search time was almost 10% compared to the no-blur condition. A more noticeable difference was found for the proportion of trials where the target was not found. Specifically, compared to the no-blur condition, in both blur conditions (σ_1 and σ_2), participants missed fewer trials with a significant drop of 40%. These results indicate a possibility to improve performance in a challenging visual search task via partial scene modulation using blur.

Further analysis of the eye movement data, in particular, fixations, revealed a significant decrease of both: the number of fixations until the target was found, and the proportion of fixations within the area of interest. These results correlate well with the trend observed in the behavioral metrics. Specifically, participants needed to make fewer fixations to find the target in blurred conditions than the non-blur condition, resulting in shorter search times and fewer missed trials. The fact that participants fixated fewer times within the salient areas when those areas were blurred supports our hypothesis that the observer's gaze was driven away from the scenes' blurred locations, contributing to a more successful search in terms of search time and ability to find the target. One possibility is that participants learned over the course of the experiment that the target is located in not blurred areas, which could result in a tendency to search non-blurred regions. However, the emerging difference in performance metrics and $PropFixInAOI$ between the no-blur and two blur conditions already in the first block of the experiment illustrates that, although some learning took place indicated by a significant effect of block number, participants did not simply learn to search only non-blurred regions. These findings are in line with existing knowledge on eye movements during real-world scene viewing. Several authors [Cajar et al., 2016a,b, 2020] showed that when a low-pass spatial frequency filter is applied across the visual field, saccades are preferentially initiated to unfiltered scene regions, both in free-viewing memorization task and visual search task. Precisely, the authors showed that when a low-pass filter, or blur, is applied to the central field of view, saccadic amplitudes tend to increase, and the number of fixations decreases. Accordingly, our results also indicate that fewer saccades landed in blurred areas, and subsequently, fewer fixations fell in those regions. In accordance with the literature available, our results support the notion that eye movements are adjusted to increase the potential usefulness of inspected visual information, moreover when foveating higher spatial frequency regions of the scene [Cajar et al., 2016a,b; Laubrock et al., 2013; Shioiri and Ikeda, 1989; Ojanpää and Näsänen, 2003; Nuthmann, 2014; Najemnik and Geisler, 2005, 2008].

Fixation duration, however, did not change much across different blur conditions. Cajar et al. [2016a] suggested that, concerning the central field of view, fixations were longer when the available spatial frequencies matched the necessary function of foveal vision, which is the analysis of details, compared to the low-pass filtered central field. On the other hand, Nuthmann [2014] argued that fixations become longer when the task requires higher processing capacity, for example, when the image is slightly blurred, and it is more difficult to distinguish details. Contrarily, Cajar et al. [2016a], found that fixations increase in duration when the visual task becomes more difficult due to spatial-frequency filtering only when the task complexity is moderate, that is when the viewer still can make use of fixating on some locations of the scene for a longer time. In the current study, different parts of the scene were blurred to a different extent. We suggest that observers fixated mostly on more sharp areas, as is also evident from the data ($PropFixInAOI$), and did not need to adjust their fixation duration significantly. It is important to note that the task nature and the type of search target also play a role in fixation duration [Nuthmann et al., 2010; Rothkegel et al., 2019; Becker et al., 2020]. In the current study, the search target was a simple Gabor-cross in contrast to some sophisticated objects. The target was always known and was constant throughout the whole experiment. Therefore, the fact that fixation duration did not change much across the different experimental conditions indicates that the task difficulty did not vary significantly, but participants just needed fewer fixations to locate the target.

When studying visual search in VR, several challenges arise, producing some limitations in the current study. One aspect is the variability in the target visibility across the different combinations of scene-target location. It is easier to find the target in some trials because it is more visible in the background, whereas it might be more difficult in others. The target's visibility depends on many factors such as contrast, spatial frequency, and retinal eccentricity, as also discussed by Rothkegel et al. [2019]. Simultaneously, the difficulty of finding the target represented by how long it took the participant to locate it in each trial depends not only on the visibility of the target but also in which direction (clockwise or counterclockwise) the observer started the search relative to the target location. In the current study, we did not control for the target visibility. Furthermore, by keeping the search target abstract rather than a meaningful object, we avoided spatial scene-contextual bias of the search where observers would likely fixate only the scene's locations containing anchors for the searched object [Boettcher et al., 2018]. However, in real life, we search for real objects. Thus, it is of interest to test our approach on realistic objects as the search target, although that may induce further bias, as some objects are expected in specific locations but not others. Also, in our study a high spatial frequency target was used. By blurring some parts of the image the distribution of the spatial frequencies of visual scene shifts towards lower frequencies. Considering feature-based guidance, this could cause the searchers to fixate less on low pass filtered locations resulting in a better performance in blurred conditions. Testing more search targets also of low spatial frequencies could provide further insight into the relation between the spatial frequency distribution of the scene and the target. In the present study, however, by selecting a certain range of possible target locations, the immediate proximity of the target was not altered by blurring. Based on the results of the study, blurring salient locations potentially can improve visual search performance at least when the search target is of high spatial frequency. Another particularity of the current study is the variability of visual scenes. While the different scenes eliminate the spatial bias of the scenes' search and learning, it introduces a scene dependence of the salient areas. Nonetheless, it is reasonable to use multiple diverse, realistic scenes for the visual search task [Rothkegel et al., 2019; Nuthmann, 2013], as that is what we observe in real life. One more aspect to keep in mind is that we based the spatial blur pattern on a saliency map computed by a model in this study. The ultimate goal of using saliency maps, as aforementioned, is to know which locations are likely to be fixated by the viewer. The vast majority of fixation prediction models, including those we used in this study, are developed for a free-viewing task and not a search task. Therefore, for further improvement of our method, it would be interesting to generate saliency maps using more fitting models, among which can be deep-neural-network based models trained not only on ground-truth spatial locations of the fixations but also taking into account the temporal dimension of fixations [Xu et al., 2020; Assens et al., 2018]. In addition to fixation analysis, it is of interest to evaluate saccadic behavior. But compared to a screen-based paradigm, it becomes more challenging to detect saccades in the eye-tracking data stream due to a high noise level caused by the headset and rapid head movement. Thus, further development of methods for a more reliable saccade detection in VR eye-tracking data is necessary such as, for example, proposed by Diaz et al. [2013]. Also, when applying a blur to the visual scene, one has to keep in mind that the visual information is partly impaired in blurred regions. This study shows that even a small blur (σ_1) can improve visual search performance. However, when the task is more demanding in processing small details, the blur strength may affect the task. Future studies analyzing how much blur can be applied improving performance in search tasks while not disturbing other performance are required. Finally, it is worth mentioning that the levels of blur used in this study were selected based on a subjective judgment of the experimenters. Specifically, a relatively small (σ_1) and a medium (σ_2) blur levels were chosen. Although no systematic analysis of subjective blur perception was conducted, some subjects reported that they noticed some parts of the scenes to be blurred more than others. Beyond the scope of the present work, future studies on subjective blur awareness can be conducted [e.g. Hata et al., 2016].

In the present study, a setting was used, where the target was located in not salient, and consequently, not blurred regions. This way it was intended to avoid a potential pop-out effect of the target when blurring the background. In the scope of this study, the results support the notion that blurring most salient regions can make the search more efficient at least when the target is situated in not salient locations. For subsequent studies it would be interesting to investigate a broader range of settings such as blurring non-salient regions, or blurring locations not based on saliency.

Regarding the guiding mechanisms in visual search, there are multiple factors which affect the search to a different extent, including bottom-up saliency, top-down goal-oriented feature guidance, scene guidance, recent search history of the observer as well as effects of value [Wolfe, 2020]. Various models of attentional guidance in visual search have been proposed describing the interplay of those factors [for review see Frintrop et al., 2010]. A commonly used framework to describe the neurophysiological mechanism of combining the guidance elements is a priority map which is a weighted average of different factors. The priority map is considered to combine the representation of bottom-up object's distinctiveness and its top-down relevance to the observer [Fecteau and Munoz, 2006; Liesefeld and Müller, 2019]. The impact of saliency in visual search guidance has been debated, where the contribution of saliency is apparent in some tasks at hand [Jung et al., 2019; Bertleff et al., 2017; Foulsham and Underwood, 2011; Theeuwes, 2004], but not in others [Chen and Zelinsky, 2006; Henderson et al., 2007; Rothkegel et al., 2019]. The neurophysiological studies also demonstrate a varied saliency contribution in the priority map building across different tasks [Fecteau and Munoz, 2006]. In the present study, by demonstrating more efficient search upon suppression of salient regions, our findings, among others, implicitly show that saliency can indeed play a role in visual search strategy not only in case of synthetically generated search sets, but also in real-world visual scenes. Importantly, results of the current work illustrate saliency impact in a 360-degree scenario where a natural free head rotation is enabled as well the visual field of view is significantly larger in contrast to screen-based studies. Even so, from previous studies it is clear that the role of saliency in visual search guidance is limited [Wolfe, 2020]. Such, if the search target would be more naturalistic and context-dependent, a smaller impact of saliency and a larger contribution of goal-oriented factors such as scene context, search history or value-based guidance can be expected. Nevertheless, we believe that through a 360-degree setting, multiple diverse visual scenes and use of a recent promising saliency model, our study contributes to understanding of saliency role in a real-world scenario of visual search.

This study served as a proof-of-principle of using a scene modulation by blurring salient regions to make the search more efficient. On a long term, the concept could potentially be implemented using, for example, see-though augmented reality head-mounted displays where the real-time video content could be modified. Another potential realization of the approach could be implementation into VR experiences such as VR simulations for professional training, e.g. in surgical procedures [Lewis et al., 2011], driving simulators [Goedicke et al., 2018], or flight simulators [Oberhauser and Dreyer, 2017].

5.6 Conclusions

To conclude, this work shows that, in a challenging realistic visual search scenario, it is possible to improve the task's performance by a saliency-aware scene modulation, specifically, partial blur. The approach's subtle nature is prone to support the user's search strategy and not be disturbing. This study provides insight into potential visual augmentation designs aiming to improve user's performance in challenging everyday visual search tasks.

5.7 Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: `osf.io/83kdh`.

5.8 Additional information

5.8.1 Author contributions

Conceptualization, O.L-S. and S.W.; methodology, O.L-S. and S.W.; software, O.L-S.; validation, O.L-S. and S.W.; formal analysis, O.L-S.; investigation, O.L-S. and S.W.; resources, S.W.; data curation, O.L-S.; writing–original draft preparation, O.L-S.; writing–review and editing, S.W.; visualization, O.L-S.; supervision, S.W.; project administration, S.W..; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

5.8.2 Funding

This study was conducted with support of the Integrative Augmented Reality (I-AR) intramural funding of the University of Tübingen, Germany.

5.8.3 Acknowledgements

We would like to thank Mikhail Startsev for providing the saliency maps used in the experiment for blurring images. Furthermore, we kindly appreciate the methodological advice of the Institute for Clinical Epidemiology and Applied Biometry of the University of Tübingen. We would like to express our gratitude towards Dr. Gunnar Blumenstock for his support in the statistical analysis of the data.

5.8.4 Conflict of interest

We declare that Siegfried Wahl is a scientist at the University of Tübingen and employee of Carl Zeiss Vision International GmbH, as detailed in the affiliations. There is no conflict of interests regarding the current study.

6. Summary

The primary focus of the present dissertation was to explore perceptually optimal visual augmentation solutions to support users' performance in challenging tasks. To generate guidelines for subtle augmentation, which still significantly improves users' performance, it was investigated how manipulation of spatial and temporal properties of visual cues affects human behavior. The individual studies provided insights into how gradual manipulation of cue complexity represented by cue discriminability affects performance in attended and unattended locations. Additionally, the impact of individual cue processing time on the performance was addressed. Moreover, the work assessed the effect on the performance of a subtle augmentation represented by a peripheral progress indicator in a challenging dual-task interleaving scenario. Finally, the potential benefit for the performance of a subtle saliency-aware augmentation represented by blur was evaluated in a visual search task. The results of the presented work led to the following conclusions:

(1) The first study explored whether attentional performance and, thus, behavior can scale gradually in space and time by systematically varying the properties of augmentation content. Also, can the individual differences in cue processing be addressed through the cue design? Yes, attentional benefits and costs in attended and unattended spatial locations can be modulated in a gradual manner if the augmenting cue is selected appropriately. The more discriminable the cue is, the larger the attentional effect becomes, even for a supra-threshold visual cue. Furthermore, based on the performance differences at two distinct ISI and their correlation with individual cue processing, it was found that participants who presumably needed more time to process less discriminable cues performed better at a longer ISI in contrast to participants who required less time for cue processing. Thus, it is possible to address individual processing differences of augmentation content using a temporally flexible visual augmentation design. The present work shows the possibility of gradual performance modulation through systematically manipulating the cue properties. Specifically, the evolution of attentional benefits and costs combined was demonstrated. Future research should further explore behavioral flexibility using perceptual cue properties as a tool. It is of particular interest to investigate if one can tailor spatial and temporal properties of augmenting content to increase the attentional benefits-to-cost ratio in a controlled manner. It, in turn, would bring visual augmentation one step further towards perceptually optimal design.

(2) The second study asked whether the peripheral locations can be exploited by augmentation. If so, how the augmentation stimulus should look like? Yes, a peri-

pheral stimulus can support the user's strategy in a challenging dual-task-interleaving scenario while maintaining the foveal processing uninterrupted. In particular, augmentation with a constantly present peripheral motion stimulus indicating the task progress showed an advantage performance-wise compared to when augmentation was not present. The proposed design only works if the augmentation is represented by a simple broad, peripherally processable stimulus. Can this augmentation design be individually tailored to the user? Yes, the individual response-time-based augmentation with an abrupt peripheral motion stimulus onset appeared to be more beneficial for the individual performance than other designs. But is individual tailoring of visual augmentation necessarily a reasonable choice? It depends on how demanding the task is. Although the response-time-based augmentation showed to be slightly more beneficial for the performance than the continuously present augmentation, it altered the natural gaze direction distribution constraining it more to the center of the visual field. Future advancements can be achieved via systematic assessment of variations of dual-task, including more complex cognitively demanding tasks where potential eye movement strategies could differ from the examined setting. Furthermore, other challenging task interleaving settings, such as scenarios with dynamic task reward functions, are of interest. This could provide further insights into subtle augmentation design in such settings as when a doctor is performing complex surgery or a new employee is practicing a new task at work. Moreover, the apparent individual variability in performance when manipulating augmentation properties indicates the importance of research on individual differences when developing visual augmentation. Such, a systematic comparison of augmentation impact on the performance of an expert and a novice across various tasks will have strong implications for the augmentation design.

(3) The third study asked whether subtle filtering of the visual information can be helpful for the user's performance. Yes, in a challenging naturalistic visual search scenario, it is possible to improve the task's performance by a saliency-aware scene modulation. Specifically, partially blurring task-irrelevant salient areas of the scene led to a more efficient search for the target in less salient regions. Better performance was achieved through fewer but not shorter fixations in the salient areas indicating the possibility to guide the gaze to support the user's strategy subtly. Naturally, the developed filtering-out approach has a potential danger of canceling out crucial visual information. Thus, the proposed augmentation design can be applied in scenarios where the environment and the task dynamics are known, and the relevance of the visual content can be evaluated. A potential realization of the approach could be implementing the suggested design into VR experiences, such as VR simulations for professional training of a predefined task. Taking it one step further, future improvements of the proposed augmentation design can focus on further investigating the usability of different factors

affecting visual search strategy for subtle augmentation development. Saliency-aware blur is one path, but other modalities such as scene guidance can be explored in the future. Moreover, bringing the setting to a more interactive scenario where natural hand and body movements are enabled will allow a higher level of realism and, therewith, provide fruitful outcomes for future augmentation designs. Finally, similar to the second study, the proposed augmentation solutions were developed based on the known dynamics of the task setting. While in many real-life scenarios, this is valid, such as in practicing a predefined task, other environments are not necessarily well understood. Therefore, other dynamic solutions based on predicting algorithms are of particular interest. These findings will present an essential step towards developing integrative augmenting visual content, which supports the user's strategy in a range of tasks in a subtle and not overruling manner.

As a closing remark, in the present dissertation, I developed a set of tools to create and examine augmentation content. I hope to have set directions for characterizing visual augmentation stimuli based on their perceptual properties. I wish for the present work to open new opportunities for subtle perception-based augmentation solutions, which will help the users in daily tasks in the long run. I believe that considering primarily human visual processing factors for the augmentation design is a natural route of success towards generating efficient visual augmenting content. Despite the considerable activity in the field of augmented reality, none of the currently existing wearable see-through solutions is widely accepted in our society. Based on previous research and the findings of this dissertation, we can say with certainty that up until now, there has been no universal augmentation which can improve human performance in all possible challenging tasks. From my standpoint, future solutions should be task-specific and clearly define what an improvement means in a particular scenario to ensure efficiency and increase trust in the technology. Furthermore, I believe that the possibility of customizing augmentation parameters to the user's individual needs is the key to the adoption of the technology in this exciting field.

7. Zusammenfassung

Der Fokus der vorliegenden Dissertation lag auf der Erforschung von für die Wahrnehmung optimalen visuellen Augmentationslösungen, um die Leistung der Benutzer bei anspruchsvollen Aufgaben zu unterstützen. Um Richtlinien für subtile Augmentationsreize zu generieren, die die Leistung der Benutzer erheblich verbessert, wurde untersucht, wie sich die Manipulation räumlicher und zeitlicher Eigenschaften des visuellen Augmentationsstimulus auf das menschliche Verhalten auswirkt. Die einzelnen Studien lieferten Erkenntnisse darüber, wie sich die graduelle Manipulation der Komplexität von Hinweisen, die in Form der Unterscheidbarkeit der Hinweise dargestellt wird, auf die Leistung an beaufsichtigt und unbeaufsichtigten Orten auswirkt. Zusätzlich wurde der Einfluss der individuellen Hinweisverarbeitungszeit auf die Leistung untersucht. Darüber hinaus bewertete die Arbeit die Auswirkung eines subtilen Augmentationsstimulus, der durch einen peripheren Fortschrittsindikator in einem herausfordernden Szenario mit ständigem Wechsel zwischen zwei unterschiedlichen Aufgaben dargestellt wurde. Schließlich wurde der potenzielle Nutzen für die Leistung einer durch Unschärfe präsentierten subtilen Augmentation in einer visuellen Suchaufgabe bewertet. Die Ergebnisse der vorgestellten Arbeit führten zu folgenden Schlussfolgerungen:

(1) Die erste Studie untersuchte, ob die Aufmerksamkeitsleistung und damit das Verhalten allmählich in Raum und Zeit skalieren kann, indem die Eigenschaften von Augmentationsinhalten systematisch variiert werden. Können die individuellen Unterschiede in der Verarbeitung des Augmentationsreizes auch durch das Design des Stimulus berücksichtigt werden? Aufmerksamkeitsnutzen und -kosten an beachteten und unbeachteten räumlichen Orten können schrittweise moduliert werden, wenn der augmentierende Hinweis entsprechend ausgewählt wird. Je unterscheidbarer der Hinweis ist, desto größer wird der Aufmerksamkeitseffekt, sogar für einen Hinweis unterhalb der Wahrnehmungsschwelle. Darüber hinaus wurde basierend auf den Leistungsunterschieden bei zwei unterschiedlichen ISI und ihrer Korrelation mit der individuellen Hinweis-Verarbeitung festgestellt, dass Teilnehmer, die vermutlich mehr Zeit benötigten, um weniger unterscheidbare Hinweise zu verarbeiten, bei einer längeren ISI besser abschnitten als Teilnehmer, die weniger Zeit zur Verarbeitung benötigen. Somit ist es möglich, mit einem zeitlich flexiblen visuellen Augmentationsdesign auf individuelle Verarbeitungsunterschiede von Augmentationsinhalten einzugehen. Die vorliegende Arbeit zeigt die Möglichkeit einer allmählichen Leistungsmodulation durch systematische Manipulation der Hinweis-Eigenschaften. Insbesondere wurde die Entwicklung der kombinierten Aufmerksamkeitsvorteile und -kosten demonstriert. Zukünftige Forschung sollte die Verhaltensflexibilität unter Verwendung von auf die Wahrneh-

mung bezogenen Eigenschaften des Hinweises als Werkzeug weiter untersuchen. Von besonderem Interesse ist die Untersuchung, ob räumliche und zeitliche Eigenschaften von Augmentationsinhalten angepasst werden können, um das Kosten-Nutzen-Verhältnis der Aufmerksamkeit auf kontrollierte Weise zu erhöhen. Dies wiederum würde die visuelle Erweiterung einen Schritt weiter in Richtung eines für die Wahrnehmung optimierten Designs bringen.

(2) Die zweite Studie erforschte, ob die Peripherie des Gesichtsfeldes für Augmentationsstimuli verwendet werden kann. Wenn ja, wie soll der Augmentationsreiz aussehen? Ja, ein peripherer Stimulus kann die Strategie des Benutzers in einem herausfordernden Szenario mit ständigem Wechsel zwischen zwei unterschiedlichen Aufgaben unterstützen, während die foveale Verarbeitung ununterbrochen aufrechterhalten wird. Insbesondere die Augmentation mit einem durchgängig angezeigten Bewegungsstimulus in der Peripherie, der den Aufgabenfortschritt anzeigt, zeigte einen leistungsmäßigen Vorteil gegenüber einer fehlenden Augmentation. Das vorgeschlagene Design funktioniert nur, wenn die Augmentation durch einen einfachen, breiten, peripher verarbeitbaren Stimulus dargestellt wird. Kann dieses Augmentationsdesign individuell auf den Nutzer zugeschnitten werden? Ja, die individuelle, auf der Reaktionszeit basierende Augmentation mit einem abrupten Bewegungsbeginn des peripheren Stimulus schien für die individuelle Leistung vorteilhafter zu sein als andere Designs. Aber ist eine individuelle Anpassung der visuellen Augmentation unbedingt eine gute Wahl? Es kommt darauf an, wie anspruchsvoll die Aufgabe ist. Obwohl sich die reaktionszeitbasierte Augmentation als etwas vorteilhafter für die Leistung erwies als die kontinuierlich vorhandene Augmentation, veränderte sie die natürliche Blickrichtungsverteilung und beschränkte sie mehr auf die Mitte des Gesichtsfelds. Zukünftige Fortschritte können durch systematische Bewertung von Variationen von der Doppel-Aufgabe erreicht werden, einschließlich komplexerer, kognitiv anspruchsvoller Aufgaben, bei denen potenzielle Augenbewegungsstrategien von der untersuchten Umgebung abweichen könnten. Darüber hinaus sind andere herausfordernde Situationen mit Wechsel zwischen Aufgaben von Interesse wie beispielsweise Szenarien mit dynamischen Aufgabe-Belohnungsfunktionen. Dies könnte weitere Einblicke liefern in das subtile Augmentationsdesign in Situationen, in denen beispielsweise ein Arzt eine komplexe Operation durchführt oder ein neuer Mitarbeiter eine neue Aufgabe bei der Arbeit erlernt. Darüber hinaus zeigt die offensichtliche individuelle Leistungsvariabilität bei der Manipulation von Augmentationseigenschaften die Bedeutung der Erforschung individueller Unterschiede bei der Entwicklung visueller Augmentation. Ein systematischer Vergleich des Einflusses der Augmentation auf die Leistung eines Experten und eines Anfängers bei verschiedenen Aufgaben wird starke Implikationen auf das Augmentationsdesign haben.

(3) Die dritte Studie erforschte, ob eine subtile Filterung der irrelevanten visuellen Informationen für die Leistung des Benutzers hilfreich sein kann. Ja, in einem anspruchsvollen naturalistischen visuellen Suchszenario ist es möglich, die Leistung der Aufgabe durch eine auffallende Szenenmodulation zu verbessern. Insbesondere führte ein teilweise angewandter Unschärfeeffekt irrelevanter hervorstechender Bereiche der Szene zu einer effizienteren Suche nach dem Ziel in weniger hervorstechenden Bereichen. Eine bessere Leistung wurde durch weniger, aber nicht kürzere Fixierungen in den markanten Bereichen erreicht, die die Möglichkeit anzeigen, den Blick zu lenken, um die Strategie des Benutzers subtil zu unterstützen. Natürlich birgt der entwickelte Ausfilterungsansatz die potenzielle Gefahr, dass wichtige visuelle Informationen ausgelöscht werden. Somit kann das vorgeschlagene Augmentationsdesign in Szenarien angewendet werden, in denen die Umgebung und die Aufgabendynamik bekannt sind und die Relevanz des visuellen Inhalts bewertet werden kann. Eine mögliche Umsetzung des Ansatzes könnte die Implementierung des vorgeschlagenen Designs in VR-Erlebnisse wie VR-Simulationen für das professionelle Training einer vordefinierten Aufgabe sein. Geht man noch einen Schritt weiter, können sich zukünftige Verbesserungen des vorgeschlagenen Augmentationsdesigns auf die weitere Untersuchung der Verwendbarkeit verschiedener Faktoren konzentrieren, die die visuelle Suchstrategie für die Entwicklung subtiler Augmentationen beeinflussen. Szenenabhängige Unschärfe ist ein Weg, aber andere Modalitäten wie Szenenführung könnten in Zukunft erforscht werden. Darüber hinaus wird die Einstellung auf ein interaktiveres Szenario, in dem natürliche Hand- und Körperbewegungen ermöglicht werden, ein höheres Maß an Realismus ermöglichen und damit fruchtbare Ergebnisse für zukünftige Augmentationsdesigns liefern. Schließlich wurden ähnlich wie in der zweiten Studie die vorgeschlagenen Augmentationslösungen auf der Grundlage der bekannten Dynamik der Aufga-benstellung entwickelt. Während dies in vielen realen Szenarien gültig ist, beispielsweise beim Üben einer vordefinierten Aufgabe, werden andere Umgebungen nicht unbedingt gut verstanden. Daher sind andere dynamische Lösungen, die auf Vorhersagealgorithmen basieren, von besonderem Interesse. Diese Erkenntnisse stellen einen wesentlichen Schritt zur Entwicklung integrativer, augmentierender visueller Inhalte dar, die die Strategie des Benutzers bei einer Reihe von Aufgaben auf subtile und nicht aufdringliche Weise unterstützen.

Zusammengefasst, habe ich in der vorliegenden Dissertation eine Reihe von Werkzeugen entwickelt, um Augmentationsinhalte zu erstellen und zu untersuchen. Ich hoffe, dass ich Richtlinien zur Charakterisierung visueller Augmentationsreize basierend auf ihren Wahrnehmungseigenschaften gegeben habe. Ich wünsche mir, dass die vorliegende Arbeit neue Möglichkeiten für subtil wahrnehmungsbasierte Augmentationslösungen eröffnet, die den Anwendern langfristig bei ihren täglichen Aufgaben helfen werden. Ich glaube, dass die hauptsächliche Berücksichtigung von menschlicher visueller Verarbeitungsfaktoren für das Augmentationsdesign ein natürlicher Erfolgsweg zur Generierung effizienter visueller Augmentationsinhalte ist. Trotz der beachtlichen Aktivität im Bereich Augmented Reality findet keine der derzeit existierenden Mixed-Reality-Brillen breite Akzeptanz in unserer Gesellschaft. Basierend auf der bisherigen Forschung und den Erkenntnissen dieser Dissertation können wir mit Sicherheit sagen, dass es bisher keine universelle Augmentation gibt, die die menschliche Leistungsfähigkeit in allen möglichen herausfordernden Aufgaben verbessern kann. Aus meiner Sicht sollten zukünftige Lösungen aufgabenspezifisch sein und klar definieren, was eine Verbesserung in einem bestimmten Szenario bedeutet, um die Effizienz zu gewährleisten und das Vertrauen in die Technologie zu stärken. Darüber hinaus glaube ich, dass die Möglichkeit, Augmentationsparameter an die individuellen Bedürfnisse des Benutzers anzupassen, der Schlüssel zur Einführung der Technologie in diesem spannenden Bereich ist.

List of Figures

1.1 Representation of "virtuality continuum" (adapted and adjusted from [Milgram and Kishino, 1994]). 13

1.2 The illustration of the concept of a subtle supporting visual augmentation where the human performance is defined by the ratio of the benefits of visual augmentation to its costs. The more prominent (less subtle) the visual augmentation becomes, the higher the benefit-to-cost ratio. However, a very prominent visual augmentation can become overloading for the user leading to a decrease in the benefits-to-costs ratio. An "optimal" performance is attributed to a "sweet spot" for the augmentation, which is supportive for the performance but does not significantly disturb the user perceptually. 16

1.3 Representation of basic Posner attentional paradigm [Posner, 1980]. Here, as an example, a symbolic central cue is shown. During the experiment, first, a cue is displayed, cueing the observer's attention to a specific location. Thereafter, the stimulus appears either in the cued location (valid cueing condition) or in the uncued location (invalid cueing condition). Finally, the observer is requested to respond to the stimulus as soon as possible. The visual performance in the attended location or unattended location can be then compared. A neutral cue is typically exploited for a control condition, where the cue does not guide the observer's attention to any particular location. 18

1.4 Illustration of lower performance due to aversion to risk in task interleaving scenario with unequal payoff function for each task. For simplification, only two parallel tasks are considered: Task 1 and Task 2. Here, Task 2 is considered to be significantly more critical to the user. . 20

3.1 The shape-based cue construction (**A**) at different cue discriminability levels: low (CD1), medium (CD2) and high (CD3), together with the neutral cue; (**B**) Schematic representation of cue direction identification. The "largest" lobe of the left and right lobes, estimated by the area of the lobe, indicates cueing direction. In this figure, as an example, all presented shape-based cues indicate the direction to the right. In the experiment, the right and the left directions were cued equally often. . 30

3.2 Experimental flow of the main experiment. 33

3.3 Results of cue sensitivity test with a shape-based cue: mean proportion correct averaged over all subjects as a function of the cue discriminability. The grid lines correspond to three cue discriminability levels in arbitrary units selected for the main experiment. 35

3.4 An example of a family of curves obtained from one session (ISI 150 ms) for one participant. The proportion of answers "tilted CW" as a function of the target stimulus tilt angle is fitted with a psychometric function. Green, blue and red curves correspond to the valid, neutral, and invalid conditions, respectively. The data sets correspond to cue discriminability levels (**A**) CD1, (**B**) CD2, and (**C**) CD3. Note that data points for some tilt angles are superimposed - all tilt angles were equally shown to the participants. 37

3.5 The average slope of response curves over all participants (dashed lines) for valid (green), invalid (red), and neutral (blue) conditions at different cue discriminability levels, together with single-participant data (continuous semi-transparent lines). 38

3.6 The individual response time recorded in cue sensitivity test as a function of cue discriminability. The N tick label corresponds to the neutral cue. On the right, the line cue response times are depicted, denoted by a tick label L. The straight solid lines are the linear fits of the measurement points for each participant. Different colors correspond to different participants. 39

3.7 The difference between individual slopes at ISI 250 ms and ISI 150 ms as a function of the cue response time normalized to the line cue response time at a given cue discriminability and validity. Red crosses and green triangles represent invalid and valid conditions, respectively. (**A**) and (**D**) correspond to CD1 cue discriminability; (**B**) and (**E**) CD2; (**C**) and (**F**) CD3. A positive Pearson correlation for the valid and invalid conditions at CD1 was obtained with Pearson coefficients 0.83 and 0.93, respectively. 40

4.1 Experimental paradigm: a dual-task interleaving setting with the unequal importance of each sub-task. The unequal task importance was implemented through 1:10 penalty ratio for the explosion of each blue and red sphere, respectively. During each 60 sec-trial the participant interleaved between the blue and the red tasks performing one sub-task at a time. The participant switched between two tasks by pressing a button on the HTC Vive controller. The example scene views demonstrate the case when augmentation was present, thus, the head-contingent ring is overlaid with the visual scene of the blue task. In the control baseline condition, the ring was not present. Note, that the example scene views demonstrate a larger field of view than the actual one in the headset — when performing the experiment in the virtual environment with the headset, the outer edge of the augmentation ring was not visible. For more details on the experimental paradigm see the description in the text. 51

4.2 The spatial configuration of the experiment. As an example, the blue room configuration is shown. The head-mounted device represents the head position of the participant, the rest of the figure schematically represents the virtual environment designed for the experiment. The details about the augmentation stimulus are described in Section 4.3.3.2. 52

4.3 Visual augmentation: a head-contingent peripheral stimulus consisting of a background ring and an additional ring sector moving clockwise. The motion speed was set to a constant value of 68°/sec resulting in a full 360° turn in approximately 5 sec. The ring sector is indicated in (**B**) by a striped pattern. The size of the sector was arbitrarily set to 20% of the full ring circumference resulting in 18°. The ring was always concentric with the heading direction of the participant regardless of the head movements. The inner and outer radii of the ring were set to 30° and 50° in visual angle (va), respectively. The transparency of the ring sector was gradually decreasing such that one edge of the sector is maximally discriminable on the background ring, whereas the other edge is blended into the background ring. For more details, see the description in the text. 53

4.4 The normalized mean duration on the blue task (dark blue), and the normalized total score (magenta), averaged over 14 participants. The x-axis represents experimental augmentation conditions. The error bars represent the standard error of the mean. The level of significance represents the results of a multicomparison post-hoc test with a Tukey correction. The indicators of significant differences are *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. 59

4.5 The difference in total score in Conditions 3 and 4 as a function of individual response time. Note, that the score in Condition 3 was subtracted from the score in Condition 4. The blue line corresponds to the best linear regression estimate with a confidence interval of 95%. The black dashed vertical line indicates the fixed value of the time-lead 1200 ms in Condition 3. 60

4.6 Gaze direction distribution: (**A**) distribution of the gaze direction in horizontal and vertical visual angle in head coordinates for one participant for one experimental condition (Condition 1). Data is visualized through a joint histogram plot with hexagonal bins where the depth of the color represents the count of data points in the corresponding region. The grey ring represents the interval of visual angles in head coordinates overlaid with the head-contingent augmentation ring in experimental conditions with augmentation present. (**B**) the same data set as in (**A**) but presented in world coordinates. (**C**) the density of the horizontal gaze direction distribution in head coordinates combined for participants. The curve describes the Gaussian kernel density estimate. Different colors represent different experimental conditions. (**D**) the density of the horizontal gaze direction distribution in world coordinates combined for participants. 62

5.1 **(A)** A schematic representation of equirectangular projection. Each rectangle is 30° wide and 15° high. **(B)** The experimental setup. The experimental paradigm was implemented in Unity software and displayed in a VR headset — HTC Vive Pro Eye with a built-in eye tracker. Position and rotation tracking of the headset was realized via the HTC base stations 2.0. The participant performed the experiment in a seated position on a rotating stool. As soon as the search target was located, the participant pressed a button on the HTC Vive controller. Additionally, a left-hand coordinate system is shown, which is used to set the Unity world coordinates. For details on the experimental procedure see Section 5.3.4.1. **(C)** A schematic representation of a sphere to which the equirectangular projections (omnidirectional images) were back-projected in Unity as a Skybox. The colors indicate like-colored locations on the equirectangular projection in **(A)**. 74

5.2 **(A)** Example of one omnidirectional image used in the main experiment. **(B)** The corresponding saliency map generated using the approach in [Startsev and Dorr, 2018]. White regions correspond to the most salient areas, whereas black color indicates the least salient locations. 75

5.3 **(A)** A cross-section of a 2D Gaussian kernel used to blur images, with two standard deviations: $\sigma_1 = 17\%$ of kernel size, and $\sigma_2 = 34\%$ of the kernel size. The kernel size was fixed to a number of pixels corresponding to 1° visual angle. For schematic illustration, the kernel size in the plot equals 10 a. u. **(B)** Example of blurred part of one scene generated as a convolution of respective 2D Gaussian kernel and original image, considering its saliency map. In the lower row of **(B)**, a part of the original image which is indicated by a red rectangular is shown in three blur conditions: no blur, Gaussian blur with maximum standard deviation σ_1, and Gaussian blur with maximum standard deviation σ_2. 76

5.4 **(A)** The Gabor cross was used as a search target in the visual search task. It is a sum of vertical and horizontal Gabor stimuli, where each of the Gabors is a product of 8 cyc/° cosine function with Gaussian envelope with standard deviations of 0.06° and 0.32°. **(B)** An example of the search target is overlaid with a visual scene at a pseudorandom location. For simplicity, the cross is highlighted by a surrounding red dashed line. In the experiment, the red line was not present. 77

5.5 **(A)** Example of one omnidirectional image, **(B)** its saliency map in gray scale, and **(C)** selected 5% most salient pixels indicated by white color. The area of interest is the white region in **(C)**. 80

5.6 Example of raw gaze position in spherical coordinates for one subject in one trial. The background image is the visual scene shown in this particular trial — for easier visualization of the gaze positions, the omnidirectional image is washed out in the figure. The black cross with coordinates $(88°, 111°)$ indicates position of the search target in this trial. The color of the scatter points range from dark violet to light yellow and indicates the time stamp of each gaze position sample starting from the beginning of the trial: here, the participant started the search from the center of the image and ended around the target position. This trial lasted approximately ten seconds and the search target was successfully found by the participant. 82
5.7 Eye movement data processing algorithm. For details see Section 5.3.5.4. 83
5.8 **(A)** An estimate of mean search time in each blur condition. The mean search time was computed for each subject by averaging the search time over the number of trials where the target was found. **(B)** An estimate of the proportion of trials where the target was not found in each blur condition. The bar plots represent data from 20 subjects. The error bars show the standard error of the mean. The indicators of significant differences obtained from the linear mixed model analysis are *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. For non-significant differences, the p-value is shown. 84
5.9 **(A)** A distribution of fixation duration as a kernel density estimate plot. Only fixations with duration up to 500 ms are shown to better resolve differences between different blur conditions. The mean duration of fixation averaged over all subjects across all trials is 133 ms ± 76 ms. The plot represents all fixations computed for all subjects and all trials. Each curve is normalized to the number of observations such that the total area under all densities sums to 1. **(B)** An estimate of the number of fixations until the target was found in each blur condition. The bar plots represent data from 20 subjects considering only trials where the target was found. **(C)** An estimate of the proportion of fixations within the area of interest in each blur condition. The bar plots represent data from 20 subjects considering all trials. **(D)** An estimate of fixation duration in each blur condition. The bar plots represent data from 20 subjects considering all trials. The error bars show the standard error of the mean. The indicators of significant differences obtained from the linear mixed model analysis are *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. For non-significant differences, the p-value is shown. 86

List of Tables

4.1 Experimental conditions. 68

5.1 Main eye- and head-movement-related raw variables recorded during the experiment. 81

Bibliography

Abrams, J., Barbot, A., and Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, Perception & Psychophysics*, 72(6):1510–1521.

Abrams, R. A. and Christ, S. E. (2003). Motion Onset Captures Attention. *Psychological Science*, 14(5):427–432.

Akoumianakis, D. and Stephanidis, C. (2005). Building consensus in human-computer interaction design: Integrated activity-oriented design environments. *International Journal of Human-Computer Interaction*, 18(1):85–103.

Assens, M., Giro-i Nieto, X., McGuinness, K., and O'Connor, N. E. (2018). PathGAN: Visual Scanpath Prediction with Generative Adversarial Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11133 LNCS:406–422.

Awan, O. A. (2021). Virtual Radiology Readouts After the COVID-19 Pandemic. *AJR. American journal of roentgenology.*

Awh, E., Belopolsky, A. V., and Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8):437–443.

Aydındoğan, G., Kavaklı, K., Şahin, A., Artal, P., and Ürey, H. (2021). Applications of augmented reality in ophthalmology [Invited]. *Biomedical Optics Express*, 12(1):511.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Bahmani, H. and Wahl, S. (2016). Distorted Low-Level Visual Features Affect Saliency-Based Visual Attention. *Frontiers in computational neuroscience*, 10:124.

Bailey, R., McNamara, A., Sudarsanam, N., and Grimm, C. (2009). Subtle gaze direction. *ACM Transactions on Graphics*, 28(4):1–14.

Barras, C. and Kerzel, D. (2017). Salient-but-irrelevant stimuli cause attentional capture in difficult, but attentional suppression in easy visual search. *Psychophysiology*, 54(12):1826–1838.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Beck, D. M. and Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, 8(8).

Becker, S. I., Atalla, M., and Folk, C. L. (2020). Conjunction search: Can we simultaneously bias attention to features and relations? *Attention, Perception, & Psychophysics*, 82(1):246–268.

Bertleff, S., Fink, G. R., and Weidner, R. (2017). Attentional capture: Role of top-down focused spatial attention and the need to search among multiple locations. *Visual Cognition*, 25(1-3):326–342.

Biocca, Frank and Owen, Charles, and Tang, A. and Corey, B. (2007). Attention Issues in Spatial Information Systems: Directing Mobile Users' Vi...: EBSCOhost. *Journal of Management Information System*, 23(4):163—-184.

Boettcher, S. E., Draschkow, D., Dienhart, E., and Võ, M. L. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, 18(13):1–13.

Booth, T., Sridharan, S., McNamara, A., Grimm, C., and Bailey, R. (2013). Guiding attention in controlled real-world environments. In *Proceedings of the ACM Symposium on Applied Perception - SAP '13*, page 75, New York, New York, USA. ACM Press.

Borji, A. (2019). Saliency Prediction in the Deep Learning Era: Successes and Limitations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1.

Borji, A., Sihite, D. N., and Itti, L. (2014). What/Where to Look Next? Modeling Top-Down Visual Attention in Complex Interactive Environments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5):523–538.

Branchini, L., Regatieri, C. V., Flores-Moreno, I., Baumann, B., Fujimoto, J. G., and Duker, J. S. (2012). Reproducibility of choroidal thickness measurements across three spectral domain optical coherence tomography systems. *Ophthalmology*, 119(1):119–123.

Brignani, D., Guzzon, D., Marzi, C. A., and Miniussi, C. (2009). Attentional orienting induced by arrows and eye-gaze compared with an endogenous cue. *Neuropsychologia*, 47(2):370–381.

Brumby, D. P., Salvucci, D. D., and Howes, A. (2009). Focus on driving. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1629, New York, New York, USA. ACM Press.

Cajar, A., Engbert, R., and Laubrock, J. (2016a). Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vision Research*, 127:186–197.

Cajar, A., Engbert, R., and Laubrock, J. (2020). How spatial frequencies and color drive object search in real-world scenes: A new eye-movement corpus. *Journal of Vision*, 20(7):8.

Cajar, A., Schneeweiß, P., Engbert, R., and Laubrock, J. (2016b). Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision*, 16(2):8–8.

Carmigniani, J. and Furht, B. (2011). Augmented Reality: An Overview. In *Handbook ofAugmented Reality*, chapter 1, pages 3–46. Springer.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13):1484–1525.

Carrasco, M. (2018). How visual spatial attention alters perception. *Cognitive Processing*, 19(S1):77–88.

Carrasco, M., Loula, F., and Ho, Y.-X. (2006). How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. *Perception & Psychophysics*, 68(6):1004–1012.

Carrasco, M., Williams, P. E., and Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, 2(6):4.

Chan, L. K. and Hayward, W. G. (2013). Visual search. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(4):415–429.

Cheal, M. and Lyon, D. R. (1991). Central and Peripheral Precuing of Forced-Choice Discrimination. *The Quarterly Journal of Experimental Psychology Section A*, 43(4):859–880.

Chen, X. and Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, 46(24):4118–4133.

Cipresso, P., Giglioli, I. A. C., Raya, M. A., and Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, 9(NOV).

Corporation, H. (2020). Getting VerboseData at the fastest rate possible. - Vive Eye Tracking SDK - Community Forum.

Coughlan, J. M. and Miele, J. (2017). AR4VI: AR as an Accessibility Tool for People with Visual Impairments. *International Symposium on Mixed and Augmented Reality : (ISMAR) [proceedings]. IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2017:288–292.

Dabbish, L., Mark, G., and González, V. M. (2011). Why do i keep interrupting myself?: Environment, habit and self-interruption. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 3127–3130.

Dalmaijer, E. S., Mathôt, S., and Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, 46(4):913–921.

Danieau, F., Guillo, A., and Dore, R. (2017). Attention guidance for immersive video content in head-mounted displays. In *Proceedings - IEEE Virtual Reality*, pages 205–206. IEEE Computer Society.

Dey, A., Billinghurst, M., Lindeman, R. W., and Swan, J. E. (2018). A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5(37):1.

Diaz, G., Cooper, J., Kit, D., and Hayhoe, M. (2013). Real-time recording and classification of eye movements in an immersive virtual environment. *Journal of Vision*, 13(12):5–5.

Drewes, J., Trommershauser, J., and Gegenfurtner, K. R. (2011). Parallel visual search and rapid animal detection in natural scenes. *Journal of Vision*, 11(2):20–20.

Dux, P. E. and Rentmarois (2009). The attentional blink: A review of data and theory.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5):14–14.

Eckstein, M. P., Koehler, K., Welbourne, L. E., and Akbas, E. (2017). Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes. *Current Biology*, 27(18):2827–2832.e3.

Enns, J. T. and MacDonald, S. C. (2013). The role of clarity and blur in guiding visual attention in photographs. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2):568–578.

Farmer, G. D., Janssen, C. P., Nguyen, A. T., and Brumby, D. P. (2018). Dividing Attention Between Tasks: Testing Whether Explicit Payoff Functions Elicit Optimal Dual-Task Performance. *Cognitive Science*, 42(3):820–849.

Favela, J., Tentori, M., and Gonzalez, V. M. (2010). Ecological validity and pervasiveness in the evaluation of ubiquitous computing technologies for health care. *International Journal of Human-Computer Interaction*, 26(5):414–444.

Fecteau, J. H. and Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, 10(8):382–390.

Ferrante, O., Patacca, A., Di Caro, V., Della Libera, C., Santandrea, E., and Chelazzi, L. (2018). Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex*, 102:67–95.

Folk, C. L., Remington, R. W., and Johnston, J. C. (1992). Involuntary Covert Orienting Is Contingent on Attentional Control Settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):1030–1044.

Foulsham, T. and Underwood, G. (2011). If Visual Saliency Predicts Search, Then Why? Evidence from Normal and Gaze-Contingent Search Tasks in Natural Scenes. *Cognitive Computation*, 3:48–63.

Frintrop, S., Rome, E., and Christensen, H. I. (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, 7(1):1–39.

Fuller, S., Park, Y., and Carrasco, M. (2009). Cue contrast modulates the effects of exogenous attention on appearance. *Vision Research*, 49(14):1825–1837.

Gatys, L. A., Kümmerer, M., Wallis, T. S. A., and Bethge, M. (2017). Guiding human gaze with convolutional neural networks.

Georgiou, Y. and Kyza, E. A. (2017). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human-Computer Studies*, 98:24–37.

Gibson, B. S. and Bryant, T. A. (2005). Variation in cue duration reveals top-down modulation of involuntary orienting to uninformative symbolic cues. *Perception & Psychophysics*, 67(5):749–758.

Gibson, J. J. (1978). The Ecological Approach to the Visual Perception of Pictures. *Leonardo*, 11(3):227–235.

Giordano, A. M., McElree, B., and Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision*, 9(3):30–30.

Goedicke, D., Li, J., Evers, V., and Ju, W. (2018). VR-OOM: Virtual reality on-road driving simulation. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 2018-April, pages 1–11, New York, NY, USA. Association for Computing Machinery.

Goldsmith, M. and Yeari, M. (2012). Central-Cue Discriminability Modulates Object-Based Attention by Influencing Spatial Attention. *Experimental Psychology*, 59(3):132–137.

Grogorick, S., Stengel, M., Eisemann, E., and Magnor, M. (2017). Subtle gaze guidance for immersive environments. In *Proceedings - SAP 2017, ACM Symposium on Applied Perception*, pages 1–7, New York, New York, USA. Association for Computing Machinery, Inc.

Guzzon, D., Brignani, D., Miniussi, C., and Marzi, C. A. (2010). Orienting of attention with eye and arrow cues and the effect of overtraining. *Acta Psychologica*, 134(3):353–362.

Hata, H., Koike, H., and Sato, Y. (2016). Visual guidance with unnoticed blur effect. In *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, volume 07-10-June, pages 28–35, New York, New York, USA. Association for Computing Machinery.

Hayward, D. A. and Ristic, J. (2013). Measuring attention using the Posner cuing paradigm: the role of across and within trial target probabilities. *Frontiers in Human Neuroscience*, 7:1–11.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., and Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye Movements*, pages 537–562. Elsevier Ltd.

Howes, A., Lewis, R. L., and Vera, A. (2009). Rational Adaptation Under Task and Processing Constraints: Implications for Testing Theories of Cognition and Action. *Psychological Review*, 116(4):717–751.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3):99–104.

Huynh-Thu, Q., Vienne, C., and Blondé, L. (2013). Visual storytelling in 2D and stereoscopic 3D video: effect of blur on visual attention. In Rogowitz, B. E., Pappas, T. N., and de Ridder, H., editors, *Human Vision and Electronic Imaging XVIII*, volume 8651, page 865112. SPIE.

Iatsyshyn, A. V., Kovach, V. O., Lyubchak, V. O., Zuban, Y. O., Piven, A. G., Sokolyuk, O. M., Iatsyshyn, A. V., Popov, O. O., Artemchuk, V. O., and Shyshkina, M. P. (2020). Application of augmented reality technologies for education projects preparation. In *CEUR Workshop Proceedings*, volume 2643.

Imaoka, Y., Flury, A., and de Bruin, E. D. (2020). Assessing Saccadic Eye Movements With Head-Mounted Display Virtual Reality Technology. *Frontiers in Psychiatry*, 11:1–19.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. In *Vision Research*, volume 40, pages 1489–1506. Pergamon.

Janssen, C. P. and Brumby, D. P. (2010). Strategic Adaptation to Performance Objectives in a Dual-Task Setting. *Cognitive Science*, 34(8):1548–1560.

Janssen, C. P. and Brumby, D. P. (2015). Strategic adaptation to task characteristics, incentives, and individual differences in dual-tasking. *PLoS ONE*, 10(7):1–32.

Janssen, C. P., Brumby, D. P., and Garnett, R. (2012). Natural Break Points. *Journal of Cognitive Engineering and Decision Making*, 6(1):5–29.

Janssen, C. P., Everaert, E., Hendriksen, H. M. A., Mensing, G. L., Tigchelaar, L. J., and Nunner, H. (2019). The influence of rewards on (sub-)optimal interleaving. *PLOS ONE*, 14(3):e0214027.

Janssen, C. P., Gould, S. J., Li, S. Y., Brumby, D. P., and Cox, A. L. (2015). Integrating knowledge of multitasking and interruptions across different perspectives and research methods. *International Journal of Human Computer Studies*, 79:1–5.

Janssen, C. P. and Gray, W. D. (2012). When, What, and How Much to Reward in Reinforcement Learning-Based Models of Cognition. *Cognitive Science*, 36(2):333–358.

Jiang, Y. V. (2018). Habitual versus goal-driven attention. *Cortex*, 102:107–120.

John Jonides (1980). Voluntary vs. automatic control over the mind's eye's movement. *Attention and performance*.

Jung, K., Han, S. W., and Min, Y. (2019). Search efficiency is not sufficient: The nature of search modulates stimulus-driven attention. *Attention, Perception, & Psychophysics*, 81(1):61–70.

Khan, R. A., Dinet, E., and Konik, H. (2011). Visual attention: Effects of blur. In *2011 18th IEEE International Conference on Image Processing*, pages 3289–3292. IEEE.

Kim, D. and Choi, Y. (2021). Applications of smart glasses in applied sciences: A systematic review.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4(4):138–147.

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., and Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering*, 57(11):2635–2645.

Kübler, T. C. (2020). The perception engineer's toolkit for eye-tracking data analysis. In *Eye Tracking Research and Applications Symposium (ETRA)*, pages 1–4, New York, NY, USA. Association for Computing Machinery.

Laubrock, J., Cajar, A., and Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, 13(12):1–20.

Lee, F. J. and Taatgen, N. A. (2002). Multitasking as Skill Acquisition Frank. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, pages 572–577.

Leube, A., Rifai, K., and Wahl, S. (2017). Sampling rate influences saccade detection in mobile eye tracking of a reading task. *Journal of Eye Movement Research*, 10(3):1–11.

Lewis, T. M., Aggarwal, R., Rajaretnam, N., Grantcharov, T. P., and Darzi, A. (2011). Training in surgical oncology - The role of VR simulation. *Surgical Oncology*, 20(3):134–139.

Li, C. L., Aivar, M. P., Kit, D. M., Tong, M. H., and Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision*, 16(8):9–9.

Liesefeld, H. R. and Müller, H. J. (2019). Distractor handling via dimension weighting. *Current Opinion in Psychology*, 29:160–167.

Lin, Y. C., Chang, Y. J., Hu, H. N., Cheng, H. T., Huang, C. W., and Sun, M. (2017). Tell me where to look: Investigating ways for assisting focus in 360° video. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 2017-May, pages 2535–2545, New York, NY, USA. Association for Computing Machinery.

Liu, T., Abrams, J., and Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychological science*, 20(3):354–62.

Liu, T., Stevens, S. T., and Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, 47(1):108–113.

Lu, W., Duh, B.-L. H., and Feiner, S. (2012). Subtle cueing for visual search in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 161–166. IEEE.

Lu, W., Duh, H. B. L., Feiner, S., and Zhao, Q. (2014). Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):404–412.

Lukashova-Sanz, O. and Wahl, S. (2021). Saliency-Aware Subtle Augmentation Improves Human Visual Search Performance in VR. *Brain Sciences*, 11(3):283.

Madsen, A., Rouinfar, A., Larson, A. M., Loschky, L. C., and Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Physical Review Special Topics - Physics Education Research*, 9(2):020104.

Marois, R. and Ivanoff, J. (2005). Capacity limits of information processing in the brain.

Mckinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference T (SCIPY 2010)*, pages 56–61.

Mena Vargas, Y. A., Millán Rojas, E. E., and Sánchez Castillo, V. (2019). Application of augmented reality as a means of interdisciplinary learning. *Scientia et Technica*, 24(3).

Milgram, P. and Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12).

Mone, M. A. and Shalley, C. E. (1995). Effects of Task Complexity and Goal Specificity on Change in Strategy and Performance Over Time. *Human Performance*, 8(4):243–262.

Montagna, B., Pestilli, F., and Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, 49(7):735–745.

Müller, H. J. and Rabbitt, P. M. (1989). Reflexive and Voluntary Orienting of Visual Attention: Time Course of Activation and Resistance to Interruption. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):315–330.

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., and Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *Journal of Neuroscience*, 23(9):3561–3565.

Najemnik, J. and Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391.

Najemnik, J. and Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3):1–14.

Neugebauer, A., Rifai, K., Getzlaff, M., and Wahl, S. (2020). Navigation aid for blind persons by visual-to-auditory sensory substitution: A pilot study. *PLOS ONE*, 15(8):e0237344.

Nguyen, N. Q., Cardinell, J., Ramjist, J. M., Androutsos, D., and Yang, V. X. D. (2020). Augmented reality and human factors regarding the neurosurgical operating room workflow. In *Proceeding — SPIE AR VR MR, 2020*.

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, 21(7):803–837.

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):342–360.

Nuthmann, A., Smith, T. J., Engbert, R., and Henderson, J. M. (2010). CRISP: A Computational Model of Fixation Durations in Scene Viewing. *Psychological Review*, 117(2):382–405.

Oberhauser, M. and Dreyer, D. (2017). A virtual reality flight simulator for human factors engineering. *Cognition, Technology and Work*, 19(2-3):263–277.

Ojanpää, H. and Näsänen, R. (2003). Utilisation of spatial frequency information in face search. *Vision Research*, 43(24):2505–2515.

Olk, B., Dinu, A., Zielinski, D. J., and Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *Royal Society Open Science*, 5(5):172331.

Olsen, A. (2012). The Tobii I-VT Fixation Filter Algorithm description. Technical report, Tobii Technology.

Orlosky, J., Liu, C., Kalkofen, D., and Kiyokawa, K. (2019). Visualization-guided attention direction in dynamic control tasks. In *Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2019*, pages 372–373. Institute of Electrical and Electronics Engineers Inc.

Over, E. A., Hooge, I. T., Vlaskamp, B. N., and Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, 47(17):2272–2280.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203.

Pestilli, F. and Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*, 45(14):1867–1875.

Pierce, C. A., Block, R. A., and Aguinis, H. (2004). Cautionary Note on Reporting Eta-Squared Values from Multifactor ANOVA Designs. *Educational and Psychological Measurement*, 64(6):916–924.

Pomarjanschi, L., Dorr, M., and Barth, E. (2012). Gaze guidance reduces the number of collisions with pedestrians in a driving simulator. *ACM Transactions on Interactive Intelligent Systems*, 1(2):1–14.

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

Pratt, J. and Fischer, M. H. (2002). Examining the role of the fixation cue in inhibition of return. *Canadian Journal of Experimental Psychology*, 56(4):294–301.

R Core Team (2020). R: A language and environment for statistical computing.

Raja, V. and Calvo, P. (2017). Augmented reality: An ecological blend. *Cognitive Systems Research*, 42:58–72.

Remington, R. and Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics*, 35(4):393–399.

Rothkegel, L. O., Schütt, H. H., Trukenbrod, H. A., Wichmann, F. A., and Engbert, R. (2019). Searchers adjust their eye-movement dynamics to target characteristics in natural scenes. *Scientific Reports*, 9(1):1–12.

Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium 2000*, pages 71–78, New York, New York, USA. Association for Computing Machinery (ACM).

Schütt, H. H., Rothkegel, L. O., Trukenbrod, H. A., Engbert, R., and Wichmann, F. A. (2019a). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, 19(3):1–23.

Schütt, H. H., Rothkegel, L. O., Trukenbrod, H. A., Engbert, R., and Wichmann, F. A. (2019b). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, 19(3):1–1.

Shioiri, S. and Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception*, 18(3):347–361.

Simons, D. J. and Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1).

Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., and Wetzstein, G. (2018a). Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642.

Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., and Wetzstein, G. (2018b). Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642.

Sridharan, S., Pieszala, J., and Bailey, R. (2015). Depth-Based Subtle Gaze Guidance in Virtual Reality Environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, New York, NY, USA. ACM.

Startsev, M. and Dorr, M. (2018). 360-aware saliency estimation with conventional image saliency predictors. *Signal Processing: Image Communication*, 69:43–52.

Takahashi, K. and Watanabe, K. (2013). Gaze cueing by pareidolia faces. *i-Perception*, 4(8):490–2.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review*, 11(1):65–70.

Trachel, R. E., Clerc, M., and Brochier, T. G. (2015). Decoding covert shifts of attention induced by ambiguous visuospatial cues. *Frontiers in Human Neuroscience*, 09(JUNE):358.

Ueda, T., Iwai, D., and Sato, K. (2019). IlluminatedFocus: Vision augmentation using spatial defocusing. In *SIGGRAPH Asia 2019 Emerging Technologies, SA 2019*, pages 21–22, New York, NY, USA. Association for Computing Machinery, Inc.

Unity Technologies (2019). Unity.

University of Nantes and Technicolor (2017). ICME'17 — Salient360! – Visual Attention Modeling for 360-degree Content.

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30.

Verghese, P. (2001). Visual Search and Attention: A Signal Detection Theory Approach. *Neuron*, 31(4):523–535.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, l., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G. L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Wahl, S., Dragneva, D., and Rifai, K. (2019). Digitalization versus immersion: performance and subjective evaluation of 3D perception with emulated accommodation and parallax in digital microsurgery. *Journal of Biomedical Optics*, 24(10):1.

Wang, Y. H. (2017). Using augmented reality to support a software editing course for college students. *Journal of Computer Assisted Learning*, 33(5):532–546.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Miles, A., Ram, Y., Yarkoni, T., Williams, M. L., Evans, C., Fitzgerald, C., Brian, Fonnesbeck, C., Lee, A., and Qalieh, A. (2017). mwaskom/seaborn: v0.8.1 (september 2017).

Wickens, C. (2021). Attention: Theory, Principles, Models and Applications. *International Journal of Human–Computer Interaction*, 37(5):403–417.

Wilkinson, F., Wilson, H. R., and Habak, C. (1998). Detection and recognition of radial frequency patterns 1This research was first reported at the annual meeting of the Association for Research in Vision and Ophthamology, 1996.1. *Vision Research*, 38(22):3555–3568.

Wolfe, J. M. (2020). Visual Search: How Do We Find What We Are Looking For? *Annual Review of Vision Science*, 6(1):2.1 – 2.24.

Wolfe, J. M. and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501.

Wolfe, J. M. and Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3):0058.

Wu, H. K., Lee, S. W. Y., Chang, H. Y., and Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers and Education*, 62:41–49.

Xu, M., Li, C., Zhang, S., and Callet, P. L. (2020). State-of-the-Art in 360-dgree Video/Image Processing: Perception, Assessment and Compression. *IEEE Journal of Selected Topics in Signal Processing*, 14(1):5–26.

Yamaura, H., Tamura, M., and Nakamura, S. (2018). Image blurring method for enhancing digital content viewing experience. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10901 LNCS, pages 355–370. Springer Verlag.

Yeshurun, Y. and Carrasco, M. (2008). The effects of transient attention on spatial resolution and the size of the attentional cue. *Perception & Psychophysics*, 70(1):104–113.

Zarraonandia, T., Aedo, I., Díaz, P., and Montero Montes, A. (2014). Augmented Presentations: Supporting the Communication in Presentations by Means of Augmented Reality. *International Journal of Human-Computer Interaction*, 30(10):829–838.

Conference contributions and talks related to this work

Peer reviewed conference contributions

Lukashova-Sanz, O., Wahl, S., and Rifai, K. (2019). The impact of shape-based cue discriminability on attentional performance. ECVP, Leuven (Belgium): Poster contribution

Lukashova-Sanz, O., Wahl, S., and Rifai, K. (2020). Determining performance in an interactive VR based challenging task-interleaving scenario. MüTüZü, Tübingen (Germany): Poster contribution

Lukashova-Sanz, O., Wahl, S., and Rifai, K. (2020). Time for a change: the impact of temporal indicators on performance in a challenging task-interleaving scenario. V-VSS, Virtual meeting: Poster contribution

Lukashova-Sanz, O., and Wahl, S. (2021). Did you find it? Visual search tends to be faster when applying saliency-aware subtle scene modulation in VR-based realistic scenario. V-VSS, Virtual meeting: Poster contribution

Non peer reviewed contributions

Lukashova-Sanz, O. (2021). Visual search enhancement via saliency-aware augmentation in VR. Young Researchers Vision Camp, Virtual meeting: Talk contribution

Statement of own contribution

Publication 1 — The impact of shape-based cue discriminability on attentional performance

Lukashova-Sanz, O., S. A. Wallis, T., Wahl, S., and Rifai, K. (2021). The impact of shape-based cue discriminability on attentional performance. *Vision*, 5(2):18

Contribution of the first author:

I actively contributed to the development of the initial study idea for the experiment and designing the study. I developed and implemented the experimental paradigm, performed the measurements with participants and analyzed the data. I wrote the initial version of the manuscript and improved it based on my ideas and feedback from the other authors.

Contribution of the other authors:

Wahl, S., provided materials and supervised the study. Furthermore, S., Wahl contributed to the development of the study idea, the methodological approach, and helped with editing the manuscript.

S. A. Wallis, T., contributed to the study design and methodological approach, as well as helped with editing the manuscript.

Rifai, K., actively contributed to the development of the study idea, the study design, the methodological approach, and helped with editing the manuscript.

Publication 2 — Augmentation impacts strategy and gaze distribution in a dual-task interleaving scenario

Lukashova-Sanz, O., Wahl, S., and Rifai, K. (2022). Augmentation impacts strategy and gaze distribution in a dual-task interleaving scenario. *International Journal of Human-Computer Interaction*, 38(4):383-394

Contribution of the first author:

I actively contributed to the development of the initial study idea for the experiment and designing the study. I developed and implemented the experimental paradigm, performed the measurements with participants and analyzed the data. I wrote the initial version of the manuscript and improved it based on my ideas and feedback from the other authors.

Contribution of the other authors:

Wahl, S., provided materials and supervised the study. Furthermore, S., Wahl helped with the initial study idea, the methodological approach and editing the manuscript.

Rifai, K., contributed to the initial study idea, the methodological approach and editing the manuscript.

Publication 3 — Saliency-aware subtle augmentation improves human visual search performance in VR

Lukashova-Sanz, O., and Wahl, S. (2021). Saliency-aware subtle augmentation improves human visual search performance in VR. *Brain Sciences*, 11(3):283

Contribution of the first author:

I actively contributed to the development of the initial study idea for the experiment and designing the study. I developed and implemented the experimental paradigm, performed the measurements with participants and analyzed the data. I wrote the initial version of the manuscript and improved it based on my ideas and feedback from the other authors.

Contribution of the other authors:

Wahl, S., provided materials and supervised the study. Furthermore, S., Wahl helped with the initial study idea, the methodological approach and editing the manuscript.

Acknowledgements

Above all, I would like to express my most profound appreciation to my advisers Prof. Dr. Siegfried Wahl and Prof. Dr. Martin Butz, for their scientific guidance, continuous support and expert knowledge. Many thanks for the extensive opportunities I have been given throughout the last years.

Dr. Katharina Rifai for the direct mentoring during the first part of the thesis and continuous support after that. Thank you for our multiple hours of profound scientific discussions and thorough, constructive feedback on my work.

I am also thankful to all my former and present colleagues at Zeiss Vision Science Lab. I am grateful for the environment of fruitful discussions, opening up new questions after finding solutions to existing obstacles, and always being curious and highly encouraging. Thank you for being outstanding support in my scientific path. I further thank all the volunteers who participated in my experiments.

Finally, I would like to express my deepest gratitude to my family and friends, Tony and Shuri, and, most importantly, my husband David for continuous support and helping me experience and overcome the challenges throughout the last years.