Fall 2017

Identification of key visual areas that guide an assembly process in real and virtual environments

Salvador Rojas-Murillo
University of Iowa

Recommended Citation
Rojas-Murillo, Salvador. "Identification of key visual areas that guide an assembly process in real and virtual environments." PhD (Doctor of Philosophy) thesis, University of Iowa, 2017.
https://doi.org/10.17077/etd.9k5zfnv3

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Industrial Engineering Commons

IDENTIFICATION OF KEY VISUAL AREAS THAT GUIDE AN ASSEMBLY PROCESS IN REAL AND VIRTUAL ENVIRONMENTS

by

Salvador Rojas-Murillo

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Industrial Engineering
in the Graduate College of
The University of Iowa

December 2017

Thesis Supervisors: Assistant Professor Priyadarshini R. Pennathur
Professor Geb W. Thomas

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D.THESIS

This is to certify that the Ph.D. thesis of

Salvador Rojas-Murillo

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Industrial Engineering at the December 2017 graduation

Thesis Committee:

Priyadarshini R. Pennathur, Thesis Supervisor

Geb W. Thomas, Thesis Supervisor

Daniel V. McGehee

Andrew Hollingworth

Clarence D. Kreiter

DEDICATION

This work is dedicated to all my Industrial Engineering Professors, at The Tecnológico de Morelia, Lehigh University and The University of Iowa who inspired me to pursue the field of industrial engineering and who helped me grow as an engineering student, an academic and more important as an individual.

I also want to dedicate this work to my Father and the loving memory of my Mother, and my sisters and brothers for their love and unconditional support.

Most especially, I dedicate this work to my beloved wife Paola, my daughter Paula and my son Diego, who have been my companions during this adventure. Their absolute support, love and generous sacrifice made this work possible.

"…because you are not experienced in the matters of the world,
all the things that have some degree of difficulty seem impossible to you.
Trust in time, because it usually provides sweet outcomes to
many bitter adversities."

Don Quixote de la Mancha
By Miguel de Cervantes Saavedra

ACKNOWLEDGMENTS

I would like to acknowledge and to thank Dr. Priyadarshini Pennathur and Dr. Geb Thomas for their significant support, guidance, insight, patience, and feedback during the development of this project, as well as throughout my graduate career. I am very thankful for having the privilege of working with you, and from learning the relevance of scientific work.

I also would like to recognize the valuable advice, guidance, and feedback that I received from the dissertation committee members Dr. Clarence Craiter, Dr. Andrew Hollinworth, and Dr. Daniel MacGehee, their substantial input had a very positive impact in the direction of my research.

In addition, I would like to recognize the technical support of Mr. Oscar Alejandro Vitella-Ramirez who developed the coding of the virtual environment.

I also want to express my thankfulness to all the wonderful people that make The University of Iowa an outstanding academic institution, and who provided a great support during my graduate studies: Jennifer Romping, George Patterson, Sarah Williams, Andrea Flaherty, Dan Mentzer, Deirdre Egan, Donald Anderson, Matthew Karam, Audrey Bahrick, Tom Persoon, Randy Fry, Sue Zaleski, Mark Thomason, Jeff Vande Berg, Leslie Bailey, Karla Stout, Tara Hoadley, Kevin Roiseland and Andrés Alfaro.

My Ph.D. studies would not have been financially possible without the support of the Mexican Council of Science and Technology (CONACYT), and The University of Iowa.

Finally, I want to thank my classmate and lab friends, Marco Prada, Drew Kern, David Heckelsmiller, Howard Chen, Brian Jones, Steven Long, Caitlyn Hagarty, Juana New, Jaqueline Pereira da Silva, Hamed Salehi, Amirmasoud Momenipour, and Brennan Ayres. For their friendship, motivation and support during challenging times.

ABSTRACT

Today's assembly operations represent about 15-70% of all manufacturing time and about 40% of all manufacturing costs, and manual assembly processes are still a significant portion of today's assembly operations. Furthermore, today's manufacturing environment requires a well-trained and flexible workforce that can easily adapt to changing products and processes. Unfortunately, manufacturing training is often performed using the master-apprentice model in the assembly line resulting in unsafe and expensive training conditions as this model is a slow and expensive process. Previous research has considered the use of virtual environments (VEs) for training purposes in different fields such as aviation, driving, construction, medicine, and manufacturing among many others. However, to this date, no assembly studies have been successful in providing a positive transfer of knowledge between virtual environments and real environments.

On the other hand, several eye-tracking studies in radiology, air-traffic control, driving, and reading show that participants with higher levels of experience have different eye-scan patterns than participants with lower levels of experience. However, it is unknown how visual scans are affected by practice. Furthermore, several empirical visuomotor studies of task-oriented processes in real environments show that observers fixated their eyes on the areas that are crucial to the required task. However, we do not know the necessary visual elements to observe when performing and when learning how to perform an assembly task, nor the effects of following visual instructions and having visual distractors during this process. Finally, we have yet to establish what observation differences may exist between real and virtual environments with regards to these unknowns.

This work presents the results of an assembly task which required participants to follow visual instructions and to select assembly objects among similar distractors. This assembly task was performed for ten cycles in real and virtual environments, and we used an eye-tracking device to register participants' visual scans. We successfully identified the areas that are needed to observe for an assembly task in both environments and the effect of visual instructions and distractors in a visual scan. We found statistically significant differences for visual scans by assembly cycle and environment, with a p-value of <0.05. We also identified a connection between learning curves and participant eye scan, showing a significant decrease in the incidence of eye tracking metrics (visit count, visit duration, fixation count and fixation duration) between the first and the tenth cycles (ΔM), particularly for visual distractors ranging from 37.36% to 48.77%, and for visual instructions ranging from 35.17% to 54.82%. We found that participants' observations became more efficient with practice, not only in terms of identifying distractors and following visual instructions but also in terms of developing an ability to observe key visual elements. For the RE we found a positive Pearson correlation between the proportion of fixation duration and assembly cycle for the key visual areas with p-values<0.002 and a negative Pearson correlation between the proportion of fixation duration for the non-key visual areas with p-values<0.046. Similar results were obtained for the VE.

PUBLIC ABSTRACT

Although assembly operators are still frequently trained on the job, their training is often slow, expensive, and sometimes unsafe. One proposed solution to this problem is the use of virtual environments (VEs). Virtual environments have been successfully applied in different fields, such as aviation and navigation. Unfortunately, to this date, no assembly studies have been successful in providing a positive transfer of knowledge between virtual and real environments.

On the other hand, previous research in real and virtual environments identified that observers fixate their eyes on the areas that are task relevant for the required task. Furthermore, several studies have shown a connection between eye fixations and previous knowledge and experience. However, to this date it is unknown what the key visual areas that are needed to perform an assembly task are, and if these areas are the same in real and virtual environments. Moreover, the effect of having distractors and following visual instructions on the observation of key visual areas is also unknown.

This work presents the results of an assembly task that required participants to follow visual instructions and to select assembly objects among similar distractors. This assembly task was performed for ten cycles in real and virtual environments, and we used an eye-tracking device to register participants' visual scans.

We successfully identified the areas that needed to be observed for an assembly task in both environments and the effect of visual instructions and distractors on the visual scan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1 - INTRODUCTION

1.1 Overview

1.1.1 Manufacturing Systems and Assembly Operations

With regards to human participation, manufacturing systems can be divided into three categories: manual work systems, worker-machine systems, and automated systems [1]. Manual work systems are characterized by the lack of powered tools, while worker-machine systems use powered tools operated by humans. Automated work systems perform without the direct intervention of a human operator.

Since the introduction of worker-machine systems in the 1800s [2], the goal of these systems has been to manufacture products efficiently, productively, and reliably. Today's manufacturing companies still pursue the same goal of being successful in a highly competitive manufacturing environment. One way to develop manufacturing competitiveness is by making an assortment of different products using the same assembly line [3, 4], and by increasing manufacturing responsiveness through the implementation of lean manufacturing and six-sigma principles [5, 6]. These principles promote manufacturing flexibility, cycle time reduction, inventory reduction, cost reduction while increasing product quality [7].

However, increasing product variety often increases the complexity of manufacturing operations [8]. As described by Hu et al. [9], this results in a higher level of difficulty for the management of the supply chain. Moreover, it also adds to the likelihood of human error. Although using automated assembly systems can counterbalance some of the negative impacts from the increase in product complexity, this introduction of automated solutions is not always feasible due to technical difficulties, and automation costs.

Furthermore, while automated assembly systems are more accurate and faster than manual assembly systems, they are not as flexible for all types of products, particularly custom or low quantity products or products with a low life cycle [10]. Therefore, even when it is technically possible to automate most manufacturing operations, it is not financially viable to automate all manufacturing operations.

On the other hand, a significant component of manufacturing operations is assembly operations (manual and automated). Assembly operations add value by putting together two or more components [11]. According to Hoedt [12], assembly operations represent about 15 - 70% of all manufacturing time, and about 40% of all manufacturing costs. Considering that all assembly processes (manual and automated) have some degree of dependency on human input, a well-trained workforce is an essential requirement for success in a highly-competitive manufacturing environment [10].

1.1.2 Assembly Operations Training

Manual assembly operations require physical and mental capabilities, such as hand-eye coordination, and cognitive resources to assemble two or more parts using only the operator's physical power without any external power source, for example, using a hammer or a screwdriver [1]. Therefore, manual assembly training is supported by physical and cognitive components. The physical component includes learning the physical movements, and the coordination between limbs needed for the manual assembly task. For example, hammering a nail requires physical coordination between the hand that holds the nail, the hand that holds the hammer and the arm that moves the hand. Physical movements provide kinesthetic and vestibular feedback, and physical feedback is an essential learning component for psychomotor tasks. As described by Schmidt and Young [13], during assembly motions motor and sequential movements serve as memory cues.

The cognitive component involves a combination of perceptual and conceptual training [14]. Perceptual learning modifies how trainees see, observe, perceive, and respond to an environmental stimuli, such as the placing of a nail at the correct angle and correct position. Perceptual learning allows observers to visually encode task-relevant information that results in observation learning and observation efficiency [15]. For example, trained personnel can read visual instructions more efficiently. Conceptual training provides trainees with the ability to categorize and differentiate objects based on the object's specific features, such as types of nails, hammers, and boards [16].

On the other hand, since its introduction by Wright [17], the concept of learning curves in manufacturing settings has been extensively studied. Learning curves are used to measure the degree of performance improvement with regards to the level of practice [18].

Equation 1: Learning curve equation [18]

$$Y = KX^n$$

"Where:

Y = Number of direct labor hours required to produce the X^{th} unit

K = Number of direct labor hours required to produce the 1^{st} unit

X = Cumulative unit number

n = $\frac{\log \phi}{\log 2}$ = Learning index

∅ = learning rate

1-∅ = The progress ratio [18]"

Based on this standard formula different methods have been proposed to calculate learning curves because not all types of learning adjust to log-linear models [19].

1.1.3 Visual Perception and Eye-Tracking

Visual inspection is based on the observers' visual perception. Visual perception is a swift process that gathers information about a novel visual scene and provides an early understanding of a visual scene [20]. Visual perception is also connected to the concept of visual interpretation, which could be described as the way we process scene observations and construct an understanding from a visual scene. Visual interpretation is supported by the three levels of vision (high, medium and low) [21],[22],[23], [24], and it is an analysis that goes from general to specific [23], [25]. For example, in the nail and hammer example, an observer first identifies the objects in the visual scene as nails, hammers, and boards, before identifying object-specific information such as nail dimensions and specific uses.

The use of eye-tracking has allowed scientists to identify four relevant aspects of visual observation and task performance. First, Yarbus showed that the interpretation of static visual scenes is task-dependent. Experiment participants examined the same visual scene in different ways as their observation task changed [26].

Second, empirical visuomotor studies of task-oriented processes in real environments show that observers fixated their eyes on the areas that are crucial to the required task [27 – 29]. Some of these empirical visuomotor studies were of daily tasks such as driving [30], making tea [28], preparing a sandwich [31], and walking [32]. Moreover, key visual areas are often different from the most visually salient cues in the scene. For example, when preparing a sandwich participants observed at a knife that was needed to perform the task instead of seeing more prominent or more salient objects in the scene [33]. Similarly, previous studies show that people can learn what areas are task-relevant, and where to locate them. For example, radiologists know which areas to observe in a chest x-ray, and chess players are able to identify when a chess piece is under attack [34 – 36].

Also, observers developed visual strategies to perform the different tasks and acquired a direct connection between eye-movements and body movements [28] [35] [31] [37].

Third, different eye-tracking studies in radiology [38], computer tomography [39], aircraft inspection [40], and in air traffic control [41] had shown that less experienced observers benefit from learning where expert observers looked when they performed their visual inspection task.

Fourth, previous research in image analysis shows a strong connection between visual scan patterns and cognitive processing [21]. Some examples are found in radiology [54] – [56], chess [36] ,[52], air traffic control [41]–[43] and reading [57] – [59]. In these activities, researchers found that although participants were asked to perform the same task, for example, playing chess, reading, or finding an abnormality in a chest X-ray, their visual scan patterns became more efficient as their level of experience increased [44].

In musculoskeletal radiology [45], and in mammogram interpretation [46] researchers found that more experienced participants identified abnormalities faster than less experienced participants [44]. The lack of experience of novice observers hinders their ability to identify abnormalities correctly. Less experienced observers often pay attention to true lesions but fail to recognize them. [44], [47], [48], [49]. Novice radiologists also tend to observe larger areas, with more eye fixations and shorter saccades [40].

Similarly, expert chess players identified a chess-piece that was under attack faster than chess players with lower levels of experience [36], [52]. Air traffic controller research indicates that less experienced air traffic controllers perform a higher amount of visual computations and have less efficient visual scans [41]–[43]. Research on reading shows that readers' eye fixations display a high level of variability for fixation duration and saccade length. Readers had fewer fixations in repetitive and redundant information

[50] [51] [52]. Therefore, several fields show that an observer who lacks both knowledge and experience will be hindered in their ability to perform efficient visual scans and to recognize and identify the significance of what they have observed.

1.1.4 Master-Apprentice Training and VEs

Assembly training often uses a master-apprentice model to train new employees. In this model, the master transmits knowledge to the apprentice by demonstrating, directing, and commenting on the apprentices' performance, while the apprentice observes, listens and follows the master's instructions [53]. The master-apprentice method has been proven to be effective. However, it is frequently a lengthy and expensive process because it requires instructors to spend long hours next to the trainee providing coaching and guidance.

An alternative approach to the master-apprentice model is the use of virtual environment (VE) systems. Virtual environments are 3D artificial environments in which the user interacts with the environment while perceiving themselves within the interaction taking place [54], [55]. Using VE systems to train assembly processes has some advantages [12]. First, operators can begin their training before the required start-up manufacturing date. Second, VE systems do not require the use of real tools and materials and are, therefore, less costly than training in a real environment (RE). Third, VEs are available at any time of the day and do not require previous work to be operational. Fourth, RE developers can modify the level of task difficulty or can program specific environmental conditions relevant to the training task. Fifth, VEs allow users to switch between different training scenarios in a short period reducing training set-up time.

Some fields have successfully used VE for training purposes such as space [56], medical education [57], welding [58], driving [59] and navigation [1, 5]. Unfortunately, the success in the application of VE for assembly tasks remains unclear. While Adams et al. found positive knowledge transfer between RE and VEs [62], other authors have found that assembly tasks are more comfortable to learn in an RE than in a VE [26] [20][63].

According to Nat [20], learning in a VE is more difficult because of the lack of haptic and proprioceptive feedback. Also, other factors have a negative impact on VE training, including a lack of collision information, the lack of consistency between virtual and real haptic information, the scale of parts, the snap of parts, and the manipulation of the parts [55], [62].

VE developers aim to overcome these negative interaction factors by increasing the system's interaction fidelity. Interaction fidelity describes the degree to which the user perceives the virtual environment to resemble the real environment [64]. When virtual and real environments are perceived to be similar, there is higher possibility that there will be a positive transfer of knowledge between the two [65] [55].

Interaction fidelity is affected by the level of visual fidelity. Visual fidelity is a metric used to describe the level of visual similarity between real and virtual environments [64]. Therefore, a usual approach used to increase the degree of interaction fidelity is by delivering high levels of visual fidelity through high-resolution displays and high-resolution scenes. However, a counterintuitive finding in the literature indicates that delivering high-resolution visual representations in a VE or highly accurate virtual representations of real-life objects does not always increase interaction fidelity [45]–[47]. Therefore, interaction fidelity is more complicated than providing high-resolution screens and images. Also, VE developers must consider that VE interfaces are the only source of interaction that is constant for all VEs and that these interfaces are the connection between the VE and the user as they deliver and receive information [12].

On the other hand, as previously described empirical visuomotor studies in RE have shown that observers performing a task concentrated their eye movements on key visual areas. Visuomotor studies in VE show similar results. A virtual walking task found that eye fixations are task-dependent and are not related to visual saliency [69]. Also, in a pick-and-place experiment Triesch et al. [70], recorded similar results.

1.2 Problem Statement and Objective

Assembly and manufacturing research describes that assembly tasks have a significant role in manufacturing operations. Assembly tasks require proper assembly training, and assembly training is a combination of motor skills training, perceptual training, and conceptual training. Eye-tracking research indicates that eye-tracking movement is task-dependent and that there is a connection between eye-movements and cognition, as well as between eye-movements and body movements. Furthermore, radiological, chess, air controller and reading research shows that there are some activities where previous knowledge and experience heavily guide eye fixations.

Visually trained observers have fewer fixations on task-irrelevant or redundant information even if this information is visually salient. Similarly, visuomotor studies in RE and VE, show that observers tend to have fewer fixations on task-irrelevant features, and do not pay attention to object changes that are not crucial to the task. Likewise, VE object interaction research describes that observers that need to follow visual instructions rather observing at these instructions for multiple times rather than observing and memorizing them.

VE research also revealed that having a correct level of interface similarity increases the probability of positive transfer of knowledge. However, it also revealed that having a high visual resolution display or showing high-resolution images does not guarantee a correct level of interaction fidelity, and that assembly tasks are more accessible to learn in RE than in VE due to a lack of haptic information among other reasons.

On the other hand, several aspects of visual training remain unclear. First, we still do not fully know the key observation points during an assembly task.

Moreover, since operators often follow visual instructions, and they are exposed to multiple sources of visual information, it is essential to know how the observation of a visual instruction pattern affects the selection of key visual areas, and how distracting information affects these areas? Also, considering a similar RE and VE assembly task. Can we characterize if the key visual areas are similar in both environments? Can we determine what the effect of following visual instructions and having visual distractors in both environments is?

Second, we know that assembly operations have a learning curve, yet we still need to determine what the relationship is between visual scanning and practice.

The third question is whether or not there is a positive transfer of knowledge between VE to an RE for an assembly task. If there is, can we assess this transfer of knowledge by comparing the visual scans of VE trained participants when they first perform an assembly in an RE, to the visual scans of new participants when they first perform an assembly in an RE? This would allow us to observe the effect of training in visual perception.

The objective of this research was to acquire empirical information that would lead to an understanding of the selection of key visual areas during an assembly process in RE and VE. Moreover, it is also relevant to acquire an understanding of the effect of visual instruction pattern and visual distractors on the selection of key visual areas. Also, we sought to understand if visual scans can be used to evaluate the level of expertise in an assembly task and if visual scans can be used to measure the transfer of knowledge between the two environments.

Acquiring this information amplified our understanding of what visual information is required to perform an assembly task in real and virtual environments, and also to identify the effect of following visual assembly instructions and having visual distractors in the selection of the required information. It also supported our understanding of the effects of manual training on visual perception. Furthermore, it provided information about the role of visual perception in the transfer of knowledge.

The fundamental goal of this research is to acquire an understanding of the effect of practice in the development of visual perceptual learning for assembly tasks in RE and VE environments, and also to determine the connection between learning and visual scans.

1.3 Specific Aims and Central Hypothesis

Three different aims guided this work. The first aim was to identify the specific features that attract the observer's overt attention during an assembly task requiring specific visual instructions with visual distractors as performed in real and virtual environments. This study used an eye-tracking device to characterize the specific objects that attract participants' attention during a repeated assembly task in a real and virtual environment. The RE uses Lego® blocks, and the VE uses virtual models of Lego® blocks that are moved using a virtual hand.

The second aim was to generate an understanding of the relationship between learning curve and visual perceptual learning during an assembly process. This aim uses the information collected from the eye-tracking device described in the first aim to study the differences for various eye tracking metrics (visit count, visit duration, fixation count, and fixation duration) between the different assembly cycles in both the RE and VE.

The third aim explored skill transfer between the RE and VE. This aim uses the information acquired by the eye-tracking device to assess visual scan differences between trained and untrained participants in both the RE and VE.

The central hypothesis of this study is that during a manual assembly task using familiar objects people attend to the same key visual areas regardless of whether they observe the assembly scene in a real or a virtual environment. However, we hypothesized that the selection of these key visual areas is not immediate and requires perceptual training.

Therefore, we hypothesized that performing the assembly task over ten cycles would modify a participant's visual perception and would provide the cognitive information needed to identify the key visual areas required for the assembly task. Moreover, we conjecture that this modified visual perception would result in a lower dependence on the observation of the visual instructions pattern, and in fewer observations of the distractor blocks. We also hypothesize that as participants modify their visual perception, their visual scans would follow a learning curve, becoming more efficient with more practice. Therefore, visual scan patterns of participants who have completed ten cycles should be more efficient when performing the same assembly task in a different environment for the first time, as opposed to the visual scan patterns of new participants who are performing the task for the first time.

Finally, we hypothesize that the lack of haptic information from the VE will result in an increased number of eye fixations.

1.3.1 Research Questions

Aim 1

- What are the key visual areas necessary to perform an assembly process in RE and VE?
- Are these key visual areas different in REs and VEs?
- Is there a difference in the number of eye fixations between REs and VEs?
- How is the selection of key visual areas affected by having to follow a visual building pattern and by having visual distractors?

Aim 2

- What is the relationship between the participant's learning curve and the participant's visual perception during an assembly process?

Aim 3

- Can we assess the transfer of knowledge by comparing the visual scans between VE and RE?

1.4 Significance of the Study

This empirical study identifies the key visual areas required to perform an assembly task in real and virtual environments. Acquiring this knowledge supported our understanding of the roles of attention, target selection, and perceptual training towards an assembly process. This study also increased our understanding of how visual perception evolves during the performance of real and virtual assembly tasks, strengthening our knowledge of the connection between visual perception and cognition. Acquiring this knowledge could have implications for training in areas that rely on visual cognitive tasks, such as playing chess, x-ray interpretation, or visual and manual tasks like dentistry and surgery.

Furthermore, this study provides information that supports the development of virtual environments. VE developers can benefit from understanding the behavioral similarities and differences while interacting in both environments, and the possibilities for knowledge transfer between environments.

1.5 Organization of the Study

Chapter Two provides a literature review of the different concepts and theories related to human vision, cognition, learning curves and virtual environments. It describes the models and theories that inspired this work. Chapter Three describes the experimental design, the design of the two assembly environments, the equipment used in the experiments and the design considerations involved for both environments. Chapter Four presents detailed experimental results for the different variables and regions of interest. Chapter Five offers a discussion of the experimental results, including the limitations of the experimental settings and implications of the study. Chapter six presents a conclusion and a description of the future implications of this work.

CHAPTER 2 – LITERATURE REVIEW

2.1 Introduction

This chapter provides a research review of topics related to human vision, visual search, visual perception models, virtual environments, and eye-tracking. The different concepts and ideas will lay the theoretical groundwork for the hypotheses presented in this study.

2.2 Human Vision

2.2.1 Physiology of Human Vision

The human eye has a frontal horizontal binocular field of view of 120 degrees, with the field of view of one eye being about 150 degrees, and 180 degrees for both eyes [71]. The field of view extends 60 degrees upward and about 75 degrees downward. Foveal vision is the area of highest visual resolution, and it is at least 20/20 for a person with normal vision and corresponds to 5 % of the field of vision [72]. Visual acuity accounts for about 1 minute of arc or 0.15 mm at a distance of 60 cm [73]

A wide field of view allows humans to perceive objects without moving their eyes. However, objects observed using only peripheral vision have a low spatial resolution. Spatial resolution declines as foveal distance increases [74]. Consequently, humans increase their visual resolution by moving their head and eyes and positioning their foveal vision in the region of interest [75]. Human eyes can move at a speed of 600 degrees per second [76].

2.3 Visual Search

Human observers gain a quick understanding of a visual scene after observing it for a brief period. After acquiring a glimpse of a visual scene, humans move their eyes to different locations to gather more information.

2.3.1 Acquiring the Gist of a Scene

Human visual perception is a swift process that supports the understanding of a novel visual scene, even if the scene is blurred [20]. Thorpe showed that this mechanism works after the first 20 ms of observation [77]. During the first 20 ms, observers acquire a preliminary understanding of the scene's semantic information and some initial information about the image's attributes [78]. The definition of the *gist* of a scene is the amount of perceptual and semantic information that observers acquire within the first 200 ms of observation [23]. During this period observers acquire semantic information about the spatial layout of the scene [79]. Understanding the spatial layout of a visual scene is essential since it provides the global arrangement and geometry of objects, and it precedes any analysis of image details [80].

2.3.2 Visual Scan

There are two primary ways to study visual searching: measuring the time required for finding a target or recording the observer's visual scanning pattern [73]. The visual scanning pattern is a measure based on two eye-movement characteristics eye fixations and eye-saccades [81]. Eye fixations occur when an observer's eyes are positioned on a specific target or region of interest (ROI). The information acquired is then projected onto the retina and is sent to the brain for processing [78]. Eye-saccades are the eye movements that connect distinct eye fixations.

Eye-fixation and eye-saccades vary widely in duration and length and are associated with the level of detailed information that the observer requires acquiring through observation. For example, short saccades indicate that the observer is performing multiple fixations to obtain more information about a specific area. On the other hand, longer saccades suggest that the observer is looking for specific information in different areas. There are also a variety of several kinds of saccadic eye movements such as fine saccades, coarse saccades and smooth pursuit movements [26].

Eye-saccades occur on the order of tens of milliseconds, averaging about three saccades per second [76]. Saccadic movements are not entirely accurate and require fine saccades to adjust and correctly locate the foveal position on a specific target location. Only relatively isolated targets can provide optimal saccade accuracy. Surprisingly, saccade precision is not affected by the length of the saccade, or by the duration of the previous fixation [82].

Yarbus performed one of the first eye tracking studies; demonstrating that visual scans are task dependent [26]. Yarbus asked participants to answer specific questions as they observed an image (Figure 1). Some of these were estimating the ages of the people in the image or describing their attire.

Other researchers determined that there is a connection between visual scan, knowledge, and cognition [81]. Chess and radiology studies have identified that expert observers from both disciplines have different scan behaviors than observers with a lower level of experience. For instance, expert chess players identify a chess piece that is under attack faster than chess players with a lower degree of experience [36]. Similarly, during the observation of an abnormal chest X-ray expert radiologists fixate their eyes on abnormalities faster than less experienced radiologists [44]. On the other hand, less-experienced radiologists fail to correctly identify and interpret abnormalities, even when they fixate their eyes on them [83].

In the field of human-computer interaction, the number of fixations has been related to low-quality interface design. For example, in computer interface evaluation, a high number of eye fixations means the observer needs to perform a more extensive search [84]. Researchers have also described a relationship between eye scans and mental workload; they consider that shorter saccades [85] and a high number of fixations [86], is correlated to higher mental demands.

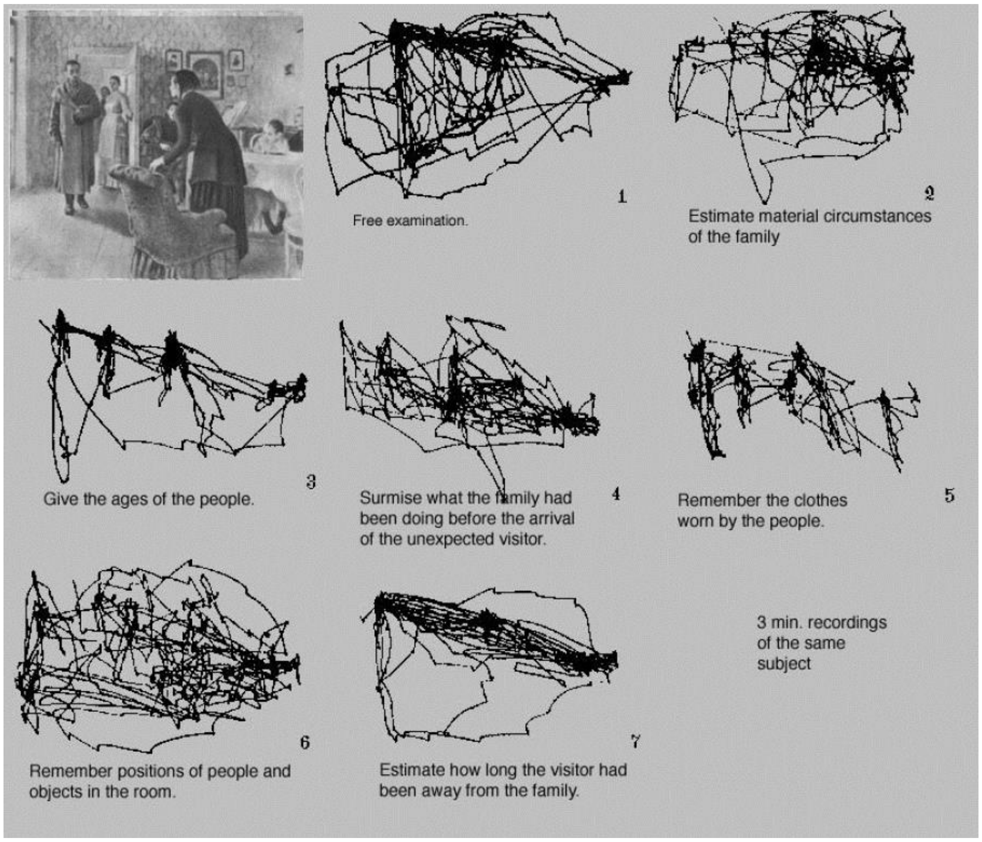

Figure 1: Scan pattern variation by task. Yarbus [26].

2.3.3 Visual Scan Strategies

As described in 1.1.3 visual scanning is relevant since there is a relationship between visual scanning and body motions, and between visual scanning and knowledge. Therefore, previous studies have considered the relevance of visual scan strategies in real and virtual environments. Kundel describes three basic visual scanning patterns that novice radiologists learn when visually scanning for a chest x-ray [73]. Similarly, in an empirical study, Kang et al. identify six different visual scan patterns for air traffic controllers observing multiple dynamic targets [87]. On the other hand, as previously described, Yarbus identifies that scan patterns are task dependent and that modifying the task parameters will change the scan pattern [26].

Ballard et al. [37] state that this finding is also valid for VEs. In his research, participants were asked to move real and virtual objects to replicate a visual instruction pattern. An eye-tracking device was utilized to register visual scan patterns to copy the visual instruction pattern. They found that participants preferred to observe and follow the visual instruction pattern rather than memorizing it. Similar results were also obtained in a navigation study in which participants attempted two different goals [69].

Furthermore, scan patterns are also affected by reward mechanisms [88], and by knowledge and expectations, as demonstrated by radiology and chess research [27]. Expert radiologists have different scan patterns as they observe smaller regions of an image, perform fewer fixations, have longer saccades, have longer fixations on areas of interest, and can connect visual information to a cognitive meaning [44], [47], [48], [89].

2.4 Visual Perception Models

Understanding a visual scene requires observers to perform different visual perception tasks. Psychological and neurological research has recognized and studied several of these tasks such as object-detection, contour/edge-detection, object-recognition, and object identification. Neurologically, visual interpretation involves different parallel visual pathways which conduct and transform visual information for visual processing [90]. Visual processing is a quick and complicated process [23], [77].

Visual processing encompasses three different vision levels: low-level vision, mid-level vision, and high-level vision. Low-level vision first involves the analysis of image-features such as object motion, object-shapes, luminance, and reflectance [21], as well as how the visual system extracts image-features and image-surfaces from retinal stimulation [91]. Low-level vision is the base of computational tasks such as pattern-recognition, image analysis, and automated processing. It is also related to the study of edges, borders, surfaces, size, orientation, contrast, symmetry, spatial frequency and shape familiarity among many others. Mid-level vision involves object groups and organization in an image [22]. High-level vision includes the recognition of objects independently of shape modifications, illumination, view point, a combination of views, and object occlusion; it is also related to the study of shape properties and spatial relations. The relevance of high-level vision is in its relationship to top-down information and its application for performing visual tasks [24].

The interpretation of a visual scene requires the interaction of these three different levels of visual processing. As described by Baars [92], the information acquired and transformed by the visual system is not only used for visual interpretation or visual processes, it is also shared with other brain areas that require visual information, such as memory or language [92], [93].

2.4.1 3D Object Recognition

There are two main approaches to studying the recognition of 3D objects. The first is called the viewpoint-invariant approach. This approach assumes that objects can be recognized regardless of the point of view. The second is the view-based approach, which suggests that objects are only recognized from specific vantage points [94].

Marr and Nishihara use an object-centered approach to describe an object's shape using the object's principal axes [95]. Similarly, Biederman describes objects as a set of volumetric primitives called geons and suggests that geons and their geometry can be used to identify the structure of an object [96].

Structure recognition is viewpoint invariant, but only for views that show the object parts that are relevant to a geon's geometry. The recognition of 3D objects should not be affected by viewpoint change as long as each view shows the same structural elements [94].

In contrast, according to view-based models, 3D objects are constructed from a 2D projection from each point of view [97] [98]. 3D objects with varying structures result in different 2D projections. Observers encode 2D features from each 2D projection, and

3D objects are recognized by comparing the current 2D projection from the previous 2D projection. New projections require the generation of a generalization from new views to the closest stored view. According to this model, the recognition of 3D objects is more inaccurate and more arduous as the distance from the following view increases from the closest stored view.

Bulthoff [99] builds on this model by adding the concept of multiple-views representation, where observers recognize 3D objects by interpolating 2D information from multiple canonical views. Canonical views are a product of learning and experience, and new 2D projections that are closer to previously learned projections are recognized faster and more accurately. Also, experience supports the development of new canonical views. An ample storage of 2D projections in memory will reduce the number of mental transformations required to interpolate between views. On the other hand, novel and deformed objects will experience rigid transformations that are more prone to inaccuracies and will take longer to interpolate [100].

Edelman and Poggio also added that the recognition of 3D objects is based on the identification of different 2D object features in various 2D projections. The selection of features and their complexity will vary depending on the object's complexity [101]. Bar also describes the importance of stored 2D projections in the observer's memory. He argues that the early recognition mechanism described by Oliva and Torralba [102] is evidence of a top-down facilitation mechanism, where previously acquired information is activated in the brain earlier than some relevant lower-level inputs. Top-down information facilitates recognition by limiting the number of stored 2D projections that need to be considered [103]. The recognition-primed decision model supports Bar's suggestion that after visual stimuli are processed during a pattern-recognition process, an expectation is generated which is confirmed by subsequent information [104]. Note that this expectation is not an unsophisticated guess as described by Cavanagh [93], and

Kosslyn [105], [106]. This pattern recognition is the most appropriate solution, and it is related to the concept of visual inference described by Marr [107].

Viewpoint-invariant and view-based approaches agree on two main points. One, there are computational and accuracy costs from changing viewpoints. Two, object structure is relevant when making a viewpoint generalization, with a complex structure being harder to generalize correctly [94]. Using these points of agreement, other researchers have developed mixed models to describe 3D recognition. Hummel and Stankiewicz proposed a model to incorporate viewpoint and structural information [108]. Foster and Gilson followed their approach and developed a model where two independent components represent objects. The first element is viewpoint invariant and is related to the structure of the object, and the second element is related to object features [109]. This new approach provides sensitive non-accidental features but is viewpoint dependent.

On the other hand, eye-tracking studies show that object eye fixation varies upon target shape, tasks objectives, and area of attention [110]–[113]. Therefore, the recognition of 3D objects is also task-dependent. Furthermore, the recognition of 3D objects still needs to solve the problem of object self- occlusion as it remains unsolved. Intricate and novel shapes are often used for these studies, and these shapes are often difficult to recognize because object shapes often self-occlude.

2.5 Virtual Environments

2.5.1 Virtual Reality versus Virtual Environments

The terms *virtual reality* and *virtual environment* are often used interchangeably; they are frequently employed to describe an artificially created environment in which the user is immersed and interacts with three-dimensional objects. However, these terms refer to two distinct concepts. *Virtual reality* refers to technologies used in the development and interaction of virtual environments such as computers, software, and display devices. On the other hand, the term *virtual environment* refers to artificially created 3D spaces where the user interacts with various objects [55].

2.5.2 VE Applications

VR technologies support the development of VEs that simulate real environments. Some of the most frequent applications are video games, entertainment (movies), and computer simulators for educational and training purposes [114]. Previous VE applications have focused on aviation [115], space [56], medical education [116], assembly tasks [63], [117], [118], welding, driving [59] and navigation [1, 5]. Ausktakalnis [119] provides a detailed description of multiple VEs for entertainment and training purposes for a diversity of applications such as architecture, construction, automotive, medicine, aerospace, and defense.

VE training offers multiple advantages. For instance, virtual training reduces the operational cost of training in real operations [58]. VE training also facilitates flexibility in choosing and customizing training scenarios [114], such as night darkness or challenging weather conditions [120], [121]. VE designers can enhance or reduce specific environmental variables such as luminance or contrast [122]. VEs are also an excellent option for training for impossible or life-threatening situations [123] [60].

2.5.3 Knowledge Transfer and VE Fidelity

Knowledge transfer occurs when a person applies previously acquired knowledge to a new task. Transfer effectiveness is determined by the amount of knowledge that is transferred from one task to the other [55]. A positive transfer of knowledge requires that both tasks have structural and surface similarities. Structural tasks are causally or functionally related to the task goal. Surface tasks are tasks that are not linked to the task goal [124].

To clarify these concepts, we can refer to the example of writing on a paper notebook versus writing on a tablet using a virtual notebook.

Using a pen-like stylus that supports the writing task (the goal) in a way that is familiar to a regular user in the real environment (pen and paper) is a structural similarity between the two tasks. On the other hand, a tablet writing display that is similar to the real notebook has a surface similarity. If the tasks share structural and surface similarities, then the probability of a positive transfer of knowledge increases. However, if the tasks do not have structural similarities, the probability that there will be a positive transfer of knowledge decreases [124]. In VE design, it is particularly important that the user should perceive not only an appropriate level of visual similarity, but also a high degree of interaction similarity between the user and the objects displayed in the VE [61], [125]. For the writing example, a virtual environment should allow the user to perceive that the writing task is similar to writing in the real environment. If the user performs the writing task using a mouse or a pointer to select letters on a virtual keyboard, there will be a lower level of interaction similarity.

As described by Waller, matching real and virtual environments requires two different kinds of fidelity: environmental fidelity and interface fidelity [61]. Environmental fidelity is related to the degree of immersion that the VE user perceives, and interface fidelity is linked to the level of similarity between feedback, control, and interaction. Interface fidelity or interaction fidelity comprises visual and motor perception [64]. VE designers have increased the levels of visual resolution in VE displays in an attempt to create high levels of environmental fidelity [126]. They also use VE systems that offer a wide field of view [127], 3D high-quality sound [128], and a "natural" manipulation of objects that at the same time provides haptic information [129]. Some authors argue that a VE with maximum fidelity will provide such a high level of perception that there could be perfect knowledge transfer because users would not be able to differentiate between real and virtual environments [61] [64].

However, improving visual fidelity does not always result in a significant environmental improvement [68]. As described by Carlson, VEs for training purposes require a VE design that is concentrated on facilitating knowledge transfer rather than on generating an exact replication of the real environment [31]. A VE design has to consider the VE training goal and artificially modify the visual representations [56] [55] [120]. The VE design must also take into account that VE training is more successful in transferring cognitive learning than motor skills [63], although immersed VE training using head-mounted displays (HMDs) have shown some potential to train gross motor skills successfully [106].

2.5.4 VE Displays

Since its development, VEs have been demonstrated using a high variety of display formats such as desktop monitors, overhead projectors and head-mounted displays (HMDs). Nemire argues, that there are visual calibration differences between VE displays and physical environments and that the artificial representation of a real environment does not always provide the same psychological responses as the actual environment [66]. Mourkoussis contends that slant perception is different in real environments than high-resolution VEs [67].

Furthermore, not all head-mounted displays offer the same level of visual fidelity. Young et al. found significant performance differences between different head-mounted displays, some of which were related to distance estimation, object interaction, simulator sickness and object-search tasks [126]. However, Mania et al. did not find any significant differences in spatial memory between participants using desktop displays and head-mounted displays [64]. Similarly, Pleban did not find any significant performance differences in combat decision making during VE training using desktop displays and overhead projectors [130]. Likewise, Patrick et al. did not find significant differences in the performance of a VE navigation task between head-mounted displays (HMDs), overhead projectors, and desktop screens [131]. However, other researchers have argued that HMDs are more efficient than desk-top displays when the observer is immersed in a

VE and uses his or her peripheral vision; e.g., large-scale navigation [132], target tracking [133], or a first-person shooter game [134].

2.5.5 VE Assembly Tasks

As previously described there are several examples of successful knowledge transfer between real and virtual environments. Unfortunately, assembly tasks are still more comfortable to learn in REs than in VEs [55] [62]. One of the reasons for this is because REs provide abundant haptic and proprioceptive feedback [62]. Haptic feedback is especially valuable for novice users [135]. Providing similar physical movements in both environments is crucial as “motor responses have a perceptual-cognitive component” [13], and provide useful proprioceptive information [55].

VE designers and scientists have been trying to improve environmental fidelity by providing artificial haptic feedback in the form of virtual gloves. Unfortunately, virtual gloves are not a perfect solution since virtual haptic systems require additional hardware, and VE users require more time to process virtual haptic feedback than virtual visual feedback [136]. A different solution for collision information and interactions with and between objects is to provide visual feedback to VE users. Some forms of visual feedback are color changes [137], arrow indicators [138], and graphic bars that show the direction and strength of collision forces [139]. Unfortunately, while virtual visual feedback provides useful information, it is unrealistic (we do not see this visual feedback in real environments), it increases mental workload, and it does not provide the same level of abundant experience provided by haptic and proprioceptive feedback [140].

2.6 Visuomotor Studies in REs and VEs

Several studies have used eye-tracking to investigate differences between VE and RE. For example, Triesch et al. studied the role of attention in VEs. They tasked participants to move artificial parallelepiped objects and used an eye-tracking device to determine if participants were able to detect changes in the object’s size. They found that participants only perceived task-relevant changes and that participants even failed to detect a change in the size of the objects that they were visually tracking [70]. A study by Johansson et al. studied gaze behavior with real objects. Participants were asked to perform a task where they had to reach for, grab, and move real objects. Researchers identified different object landmarks; they found that participants maintained a high number of eye fixations close to landmarks that were critical for the control of the task [35]. Rothkopf et al. reached similar conclusions from a walking study of real and virtual environments, in which participants fixated their view to task-relevant features, rather than observing salient background cues [69]. A motor planning study on how to prepare a sandwich in real environments also found that subjects fixated their eyes on the task-relevant areas, which were not the most salient in the visual scene [141].

2.7 Visuospatial Ability

Carroll defines visuospatial ability (VSA) as “the capacity to encode spatial information and maintain it in the working memory while transforming it” [142]. VSA is associated with individual characteristics such as gender and age [143] [144]. People with high VSA tend to pursue careers in science, technology, engineering, and mathematics [145], [146]. In contrast, individuals with lower levels of spatial ability tend to seek careers in the areas of business and education [145], [146]. VSA ability is also considered relevant in surgery [147] and dentistry [148].

Murdoch et al. found a positive correlation between a spatial relations test and surgical skill [147]. Wenzel et al. also observed a positive relationship between VSA and spatially complex surgical procedures in trainees with little experience, but not in the case of experienced practitioners [149]. In dentistry, the Dental Admissions Test (DAT) includes a section on perceptual spatial ability [148] among other criteria for selecting pre-doctoral students.

There are mixed results from studies of VSA when interacting with 3D models. Luursema et al. argue that participants with low VSA ability increased their anatomical understanding after interacting with 3D anatomical reconstructions [150]. Meijer et al. reached comparable conclusions using novel complex images [151]. However, Huk argues that participants with low VSA become cognitively overloaded with the use of 3D representations [152].

Waller found that VE interaction of males outperformed female participants [61]. Waller later argued that this is explained by gender VSA differences [153]. Likewise, Sjolinder found that VE navigation is affected by age [154]. Conversely, in an assembly task study, Hamblin found that female participants performing an assembly task in a VE had higher levels of training transfer, faster assembly times, and greater efficiency [55].

CHAPTER 3 – METHODS

The following section describes the experimental design, recruitment procedures, design considerations, the assembly process, and the pre-assembly and post-assembly procedures in both the real and virtual environments. The research methods were approved by The University of Iowa Institutional Review Board (IRB).

3.1 Experimental Description

3.1.1 Experimental Design

We selected a within-subjects design for our pilot study. This experimental design has two main advantages: 1) it reduces the number of participants required for the experiment because all participants perform all treatments [155], and 2)it reduces errors associated with individual differences.

We asked participants to perform an assembly task ten times in two different environments, real and virtual. Half of the study participants were required to perform the assembly task in the RE first before performing that same task in the VE. To prevent any order effects [156], we asked the other half of the participants to perform the assembly task in the opposite order.

3.1.2 Dependent and Independent Variables

The study was a single factor design with repeated measures for two levels (RE and VE). Table 1 shows the experimental metrics as well as the independent and dependent variables.

Table 1: Dependent variables and their metrics.

Dependent Variable	Independent Variable	Metric
Visit Count by ROI	Environment (RE or VE)	Number of visits per ROI
Visit Duration by ROI	Environment (RE or VE)	Milliseconds
Fixation Count by ROI	Environment (RE or VE)	Number of fixations
Fixation Duration by ROI	Environment (RE or VE)	Milliseconds
Assembly Time per Cycle	Environment (RE or VE)	Seconds

3.1.2.1 Relationship between Aims, Research Questions, and Experiment Variables

As previously described in section 1.2, the goal of this study is to acquire an understanding of the development of a cognitive vision for assembly tasks in RE and VE, and also to determine the connection between learning and visual scans. For this purpose, we prepared three specific aims, along with related research questions, variables, and metrics.

Aim 1. Identify the specific features that attract an observer's overt attention during an assembly task requiring assemblers to follow visual instructions while distracting objects are displayed in real and virtual environments.

Research Questions for Aim 1.

a) What are the key visual areas that are required to perform an assembly process in an REs and VEs?
b) Are these key visual areas different in REs and VEs?
c) Is there a difference in the number of eye fixations between a RE and VE?
d) How is the selection of key visual areas affected by having to follow a visual instruction pattern and by having visual distractors?

To answer the first question, we divided the task into ten snapshots. Each participant completed a total of ten repetitions, or cycles, in each environment. Each snapshot was coded for different regions of interest (ROIs) as described in section 3.1.5 -

Experimental Coding. We measured visual scan metrics such as the number of visits, visit duration, number of fixations and fixation duration time per ROI using a Tobii analyzer®. As outlined in the previous chapter, eye-tracking research has defined key visual areas as the regions that are most frequently observed, and that are needed to perform the required task. To answer the first question, we identified ROIs with a high proportion of visits, visit duration, eye fixations and eye fixation durations for each assembly task in both environments.

To answer the second and third questions we performed a linear regression analysis of the visual scan metrics for the different ROIs. The purpose was to identify contrasts between ROI observation metrics (distractors, visual instruction pattern, key visual areas) identified in question 1. In addition, we prepared a series of tables that show each observation metric and the contrasts between them.

We performed a similar analysis to answer the fourth question. Here we looked for contrasts in the visual scan metrics between the key visual areas identified in question 1 between the RE and VE.

Aim 2. Generate an understanding of the relationship between the participant's learning curve, and visual perceptual learning during an assembly process.

Research Question for Aim 2.

What is the relationship between learning curve and visual perception during an assembly process?

First, we determined the learning curve for the assembly task in each environment using the proportion of fixation duration for each assembly cycle for both environments.

Second, we generated several tables for each scan metrics that show the metric difference between the first and tenth assembly cycles.

Third, we performed a learning curve analysis based on the fixation proportion for each ROI for both environments based on previous training.

Aim 3. Explore skill transfer between RE and VE.

Research Question for Aim 3.

Can we assess the transfer of knowledge by comparing the visual scans between VE and RE?

To answer this question, our goal was to compare the results of the linear regression analysis for two different groups of participants. The first group performed the assembly task for the first time in an environment without any prior training. The second group had been trained in the previous environment for ten assembly cycles. Comparing visual scan metrics and the assembly time per cycle for both groups, we were able to determine if the initial training had any influence on participant performance.

3.1.3 Participants' Demographics and Screening

We recruited 30 male undergraduate and graduate students from the College of Engineering to reduce potential sources of variability. Previous research indicates visuospatial differences between different genders and occupations [123], [124].

Only participants with normal vision and corrected vision in the form of contact lenses were allowed to participate. To confirm their visual acuity, we performed a visual acuity test [157], which we then followed with a color blindness test [158].

All participants were required to pass both visual tests in order to participate in the study. Although we did not perform a motor skills test, we verified that participants were physically capable of performing the physical tasks required without any physical assistance.

3.1.4 Hardware and Software

The virtual environment was controlled using a computer with an Intel Core i7-6500U @ 2.50 GHz with an installed physical memory (RAM) of 16.0 GB. It was developed using Unity 5.4.2f2 (64bit).

A LEAP® Motion controller was used to move the VE blocks. The LEAP® Software was version 3.1.3+41910 (Figure 2) [159]. The controller allows users to move virtual objects with hand gestures. This motion controller has two cameras with three infrared LEDs. These cameras have wide-angle lenses which allow the controller to cover an area of about 60 centimeters above the device in four directions (left, right, superior and posterior). The controller has the following dimensions (3.1 x 1.2 x 0.5), and it has a performance of 200 Hz. [119].

Figure 2: LEAP® motion controller [145].

Figure 3 shows a user with his hands over the LEAP® controller over a desk. As observed in this image, a person in a seated position having a computer and the LEAP® motion controller positioned on a desk will see both pairs of hands at the same time, the user's real hands and the virtually generated hands.

We selected a Tobii Pro Glasses 2® eye-tracker[1] to capture eye-movements (Figure 3). The Tobii Pro Glasses 2® is a sophisticated eye-tracking device. It includes a high-definition scene camera (1920 x 1080 at 25 fps) that has a field of view of 90 degrees with a 16:9 format. The glasses also include a microphone and four eye-tracking

[1] Copyright © 2017 Tobii AB

Figure 3: Tobii pro glasses 2® eye-tracker.

sensors. The device sampling rate is 100 Hz. There is a calibration procedure of 1 point and a slippage compensation that allows 3D eye-tracking.

The glasses are connected to a small computer that is called the recording unit. The recording unit provides computational control for the glasses and also records the video and audio information acquired by the eye tracking. The recording unit also has a removable memory card slot that allows users to record and transfer the information to a computer for later analysis. The recording unit also has a battery that powers the recording unit and glasses and is connected to a computer that has a controller software. The recording unit dimensions are 130 x 85 x 27 mm [160].

Once the information is transferred to a computer, it can be coded and analyzed by using the Tobii Analyzer® Software version 1.36.1430 (x64) release date 5/11/2016 Copyright © Tobii AB 2011-2016. This software maps eye-tracking motions recorded by the glasses into one or several snapshots.

Figure 4: User with his hands over the LEAP® motion controller [160].

Software mapping is particularly useful for video recordings because the software enables users to employ different snapshots to map different steps that are of interest for the study. Figure 5 shows the Tobii analyzer®.

The analyzer software includes an automatic mapping feature that recognizes the regions of interest (ROIs) in the snapshot and maps the eye fixations from the video to the different ROIs. It also provides a feature that allows users to design various areas of interest within each snapshot, which are then mapped to required ROIs. Snapshots can be observed as the small pictures on the right panel of Figure 5.

Although the automatic feature is useful, there are some cases where the snapshot is not correctly mapped. The incorrect mapping may occur because we are mapping a video feed which is continuously changing to a still snapshot.

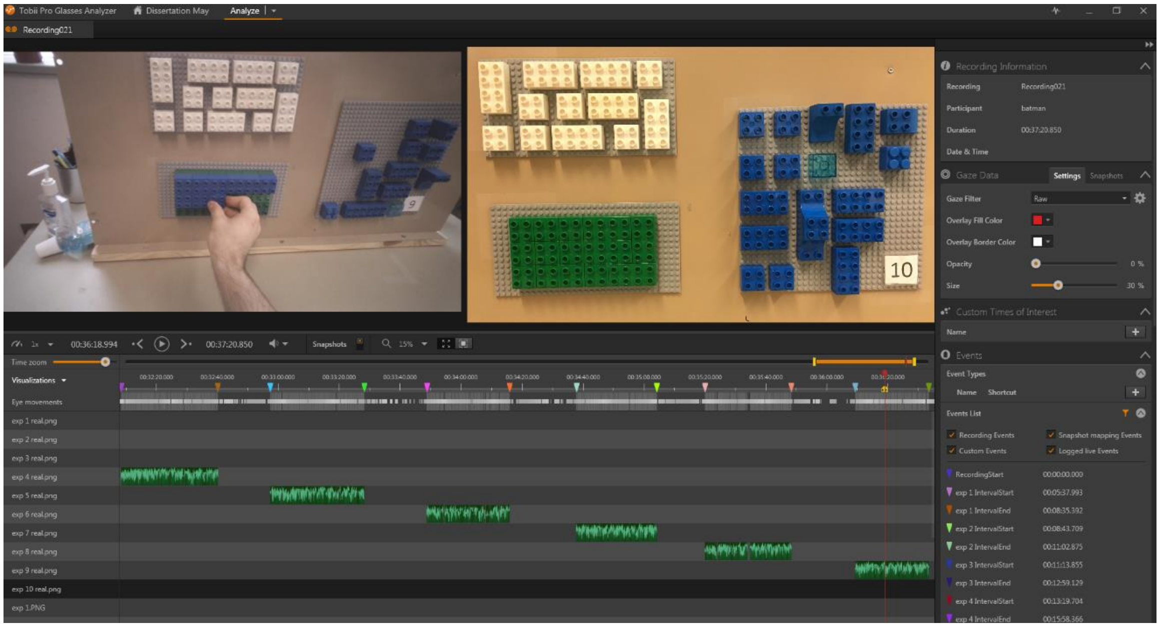

Figure 5: Tobii analyzer®.

To overcome automatic mapping errors, software designers provided a useful manual mapping tool that allows users to map the fixation manually to a specific ROI. The software also generates a detailed report that includes a variety of eye-tracking metrics such as the number of visits, visit duration, number of fixations and fixation time per snapshot and ROI.

In Figure 5, the picture frame on the top left is a video feed from one participant, while the frame on the top right shows the selected snapshot that is used to map the eye fixations to the different ROIs as defined by the research team for each snapshot. Rows below the images of the video feed and the selected snapshot represent different snapshots. The green bars represent the mapping between the video feed and the selected snapshot.

3.1.5 ROI Coding

We coded the visual scene for both environments in four main areas. Figure 6 displays images of the first assembly layout for both environments. The study maintained design consistency by using the same layout structure and block colors for all cycles for all areas. Color is frequently used as a way to distinguish similar parts in VEs [63].

The first area is the visual instruction pattern. This area is identified by the white blocks, and is located upper left. It shows the block arrangement participants must follow.

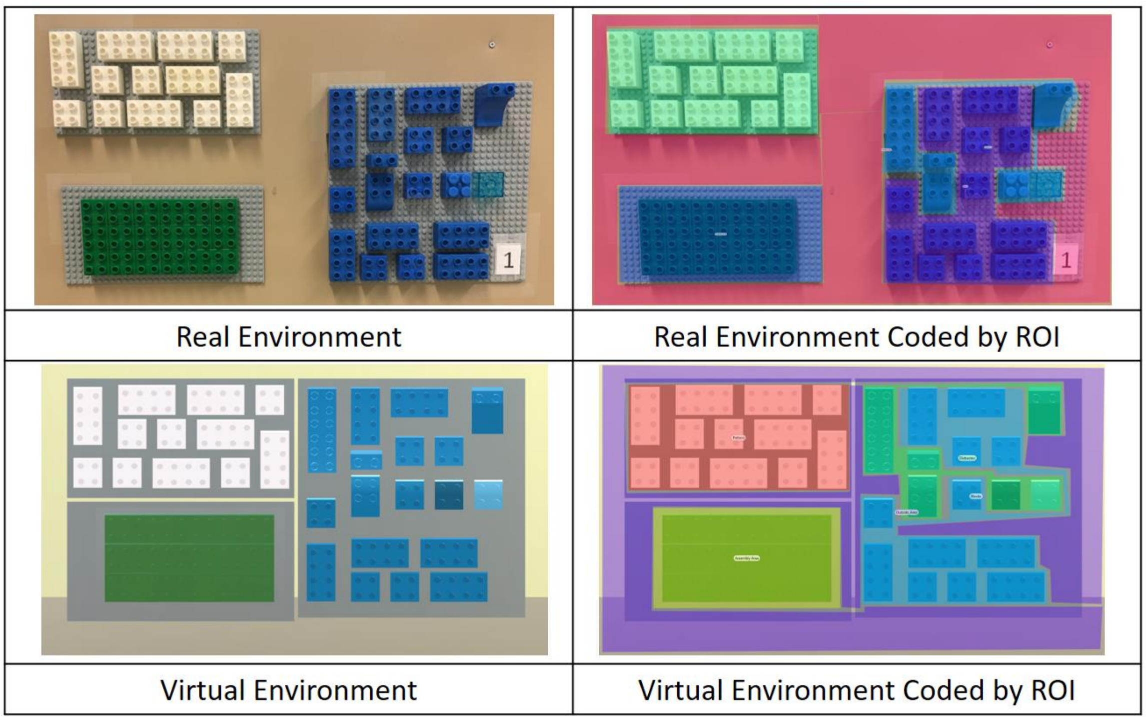

Figure 6: First assembly cycle with and without coding.

The second area is the assembly area, where building blocks are assembled. This area is located below the visual instruction pattern area and is identified by its green colored blocks.

The third area is the resource area. This area contains the building blocks and the distractor blocks. We are interested in learning the effect of having building blocks and distractor blocks in the same area.

The fourth area is the outside area. This area is the area that connects all other areas. The outside area is used to transfer the assembly blocks from their original location to the assembly area.

3.2 Experimental Design Considerations

The development of a virtual environment for assembly tasks requires developers to solve different challenges. Some of these challenges include object manipulation, scale differences between real-life and virtual objects, object assembly, object snap, and object fitting. As described by Argote and Ingram, task similarity is key to knowledge transfer [65]. Therefore, virtual environment assembly tasks should also be as similar as possible in both real and virtual environments.

3.2.1 Geometric Shapes

We selected simple geometric shapes because basic block shapes such as cubes and parallelepipeds would be familiar to all users. Previous research has shown that unfamiliar shapes are more difficult for 3D recognition, as described in section 2.4.1 [79], [100], [161], [162].

We chose blocks with distinct colors to indicate specific areas and their function. We selected Lego® Duplo blocks for our assembly task because of several advantages. First, these blocks have bright colors and familiar shapes that are easy to recognize for most participants.

Second, these blocks are easy to assemble and do not require previous training or instruction. We decided to build on top of other blocks because it was more comfortable to assemble blocks on top of other blocks than to assemble blocks on base plates. Also, the blocks added clarity for the dimensions of the assembly area.

Third, these blocks are easy to track using the eye tracker and are also easy to represent in a virtual environment. Fourth, they are easy to reproduce in a VE, and the dimensions of the virtual objects are very similar to the dimensions of the real objects. Fifth, they are double the size of regular Lego® blocks making them easy to handle in both environments.

3.2.2 Assembly Layout

3.2.2.1 Visual Instruction Pattern

We opted for a random arrangement for our assembly task. We used the same visual instruction pattern in both environments (Figure 7). This arrangement is constructed by using six cubes and six rectangular shapes and does not require any hand-twisting motions. We selected a random arrangement because it required users to examine and follow the visual instruction pattern.

This assembly design had two objectives. First, we wanted participants to initially consider this arrangement as a "random" sequence with no implicit logic.

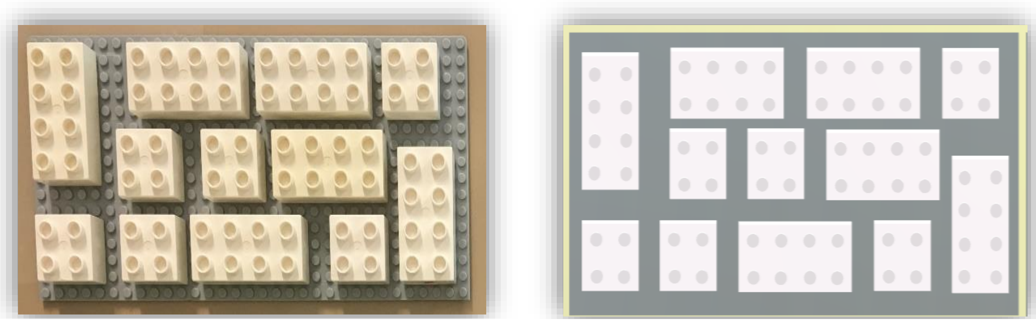

Figure 7: Visual instruction pattern white blocks for real and virtual environments.

Second, after performing several assembly cycles, participants would be able to learn the design. To achieve the second objective, we asked participants to "read" the design from top to bottom and left to right.

Following this logic, participants memorized the block sequence by observing the shape and orientation of the blocks. For example, the first and last assembly pieces were vertical rectangles. However, while this requirement facilitated the learning of the visual instruction pattern it hampered our ability to study the strategy of the visual scanning pattern.

3.2.2.2 Resource Area

The resource area contained two types of blocks: building blocks and distractor blocks. All blocks were blue with distinct shapes. Building blocks are squares and rectangles. Distractor blocks are designed to have similar shape and dimension to the building blocks.

Block arrangement in the resource area was modified for each assembly cycle in order to prevent participants from learning the location of blocks and distractors in the resource area. However, the arrangement of the blocks was consistent with each assembly cycle, meaning that the arrangement in the first and subsequent assembly cycles remained constant for all participants and environments.

Figure 8 shows each arrangement of the ten assembly cycles.

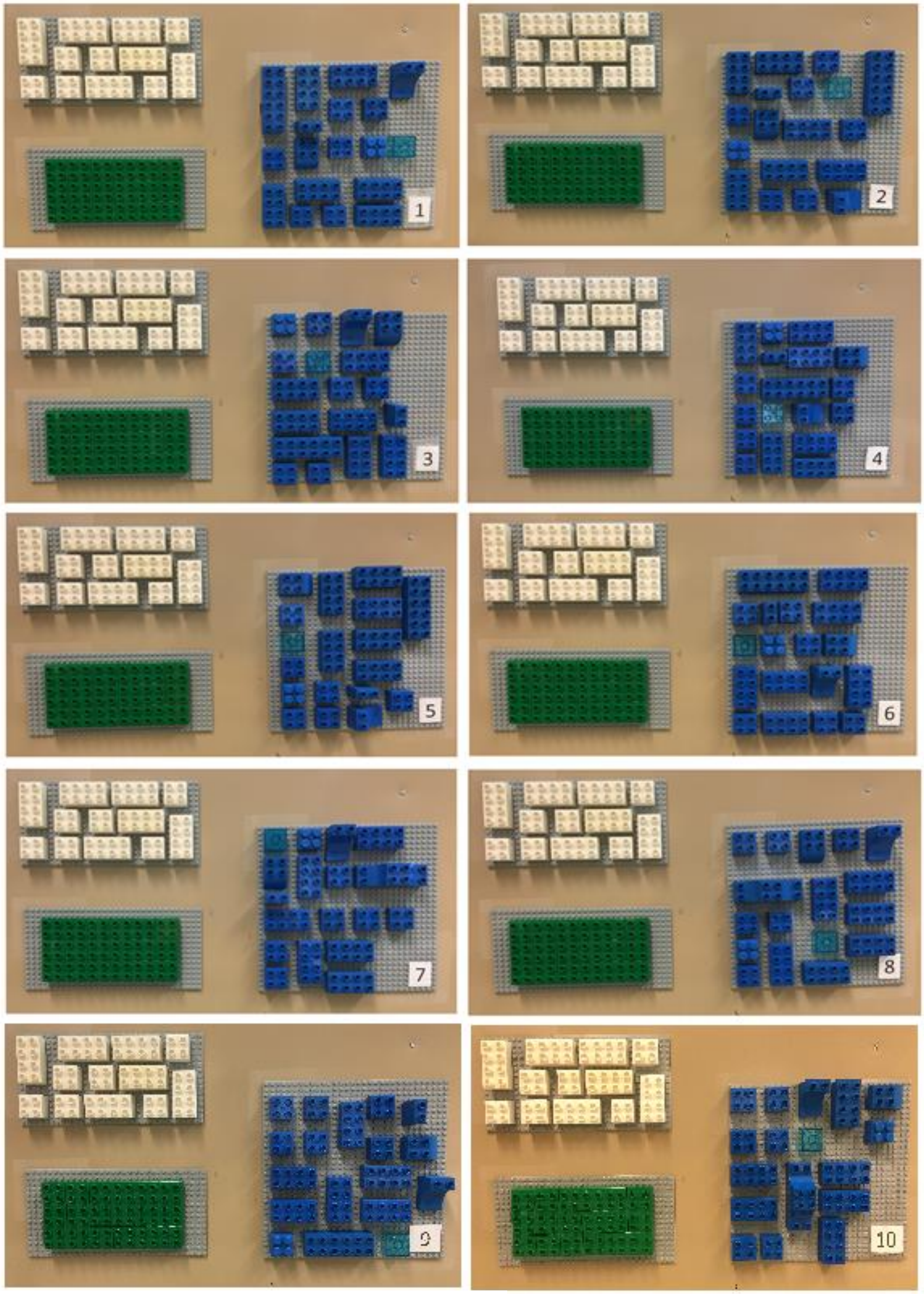

Figure 8: Experimental layouts.

3.2.3 VE Programming

The programming of the virtual environment was performed using Unity software 5.4.2f2 (64bit) ©2016 Unity Technologies ApS. Unity is a game programming engine that provides a high degree of coding flexibility. Unity can be coded in several computer languages such as C, C++, C#, JavaScript, and Boo [163]. We coded our scripts using C# because there are several libraries available for the LEAP® motion controller.

Previous studies of VE assembly provided haptic feedback and allowed users to use both hands [55], [62], [70]. We selected a LEAP® motion controller to move and assemble virtual objects because it is easy to program, and it provides proprioceptive feedback. Proprioceptive feedback is essential when learning psychomotor tasks because sequential movements serve as memory cues [13], [54].

However, one of the drawbacks of the controller is that it does not provide haptic feedback. Therefore, to prevent haptic confusion from the utilization of both hands, we asked participants to use only their dominant hand for the assembly task in both environments. Asking participants to only use one hand provided consistency between environments and prevented providing haptic advantages for the real environment.

3.2.3.1 Pinch and Snap Functions

The pinch gesture is the most relevant piece of code in our VE because it allows the user to select, control, move and place a specific block in the desired position. The LEAP® motion controller provides a pinch code for this purpose. We modified this script so the user could only activate one block at a time. The pinch gesture was only activated when the virtual hand was a short distance from the edge of a selected block. The pinch gesture is defined as positioning the tip of the index finger against the tip of the thumb while leaving the other three fingers extended. Once the pinch gesture is activated, and a virtual block is selected, the participant moves his physical arm and hand to move the virtual object to the target location.

The second most relevant function is the snap function. Using Unity's collider function, we prepared an inactive visual instruction pattern that remained “invisible” until there was a position match between the moved block and the invisible block. The solved visual instruction pattern is a set of blocks that remain inactive until the user positions the corresponding assembly block in a near correct location along the three x,y, and z-axes.

3.2.3.2 Visual and Acoustic Feedback

The design of the VE required participants to place each virtual block in a position that is close to the precise location. Therefore, participants required assistance to match the position of their real hand to their virtual hand on the z-axis. We provided visual feedback for the z-axis by changing the color of the block to red if the piece was either too close or too far from the participant. Also, we provided positive feedback when the piece was close to the target location by changing the color of the assembly block to light gray. Finally, once a block was positioned in its proper location it remained activated, and it was not possible to move it again.

We also identified that the lack of haptic feedback generated stress in many participants as there was a short lag time between performing the pinch gesture and virtually grasping the virtual block. An unintended consequence of this time delay was that participants often performed a new pinch gesture when a block was already activated. Therefore, we provided two different feedback sounds. The first sound was the pinched sound, and it was played when the participant acquired control over a block. The second was the assembly sound, which played when the block was correctly positioned at the desired location.

3.2.4 Display

3.2.4.1 VE Display Selection

To display our VE, we chose a computer monitor instead of a helmet-mounted display for two reasons. First, the Tobii Pro Glasses 2® eye-tracker frame's dimensions (179 x 159 x 57 mm) did not fit comfortably inside a helmet-mounted display. Second, research suggests that helmet-mounted displays are more efficient than desktop displays for VEs that use peripheral views or complex depth of perception, such as first person shooter video games [164].

Our assembly task did not require participants to use their peripheral view, nor did the task have a complex depth interaction between the different building elements, so we selected a desktop monitor for our purposes. The computer monitor was projecting a virtual environment at a resolution of 1080p (1920x1080 pixels, 16:9 aspect ratio).

3.2.4.2 Lay-Out Scale

As described, a high degree of similarity between the virtual and real environments increases the probability of positive knowledge transfer. Therefore, the layout and design of both environments were intended to be as similar as possible. Also, large objects allowed for better control of the LEAP® motion controller and a better eye-motion capture from the eye-tracker. The blocks in both systems had a 1:1 scale allowing observers to perceive a similar interaction.

3.3 Assembly Description

3.3.1 Assembly for Real Environments

The real environment task required participants to pick up and place twelve Lego® blocks to construct the visual instruction pattern (Figure 7). Participants performed the assembly task for ten cycles in a seated position using only their dominant hand. There was no time limit to complete the task. However, participants were encouraged to put the blocks together to their best of their ability and to complete the task as quickly as possible. Participants were allowed to modify the height of the chair so they would have a direct line of sight to the blocks and were seated 60 cm away from the Lego® panels blocks (Figure 9).

3.3.1.2 Training Exercise

Our real environment training did not include a pick and place training exercise. However, participants were told not to use both hands for the assembly and learned how to read the visual instruction pattern as well as the difference between the assembly blocks and the distractor blocks. They were instructed not to pick up any blocks if they fell to the ground and to rotate their seat 90 degrees to the left after completing each cycle. Rotating the chair allowed space for us to replace the panels that contained the blocks and also prevented participants from observing the visual instruction pattern for a more extended period.

3.3.1.3 RE Posture

In the real environment, participants performed the assembly task in a seated posture (Figure 9). This posture allowed participants to move their limbs naturally, and to adjust the seat to the required height. Furthermore, the real environment used Lego ® panels, making it easiest to position each block from a seated location.

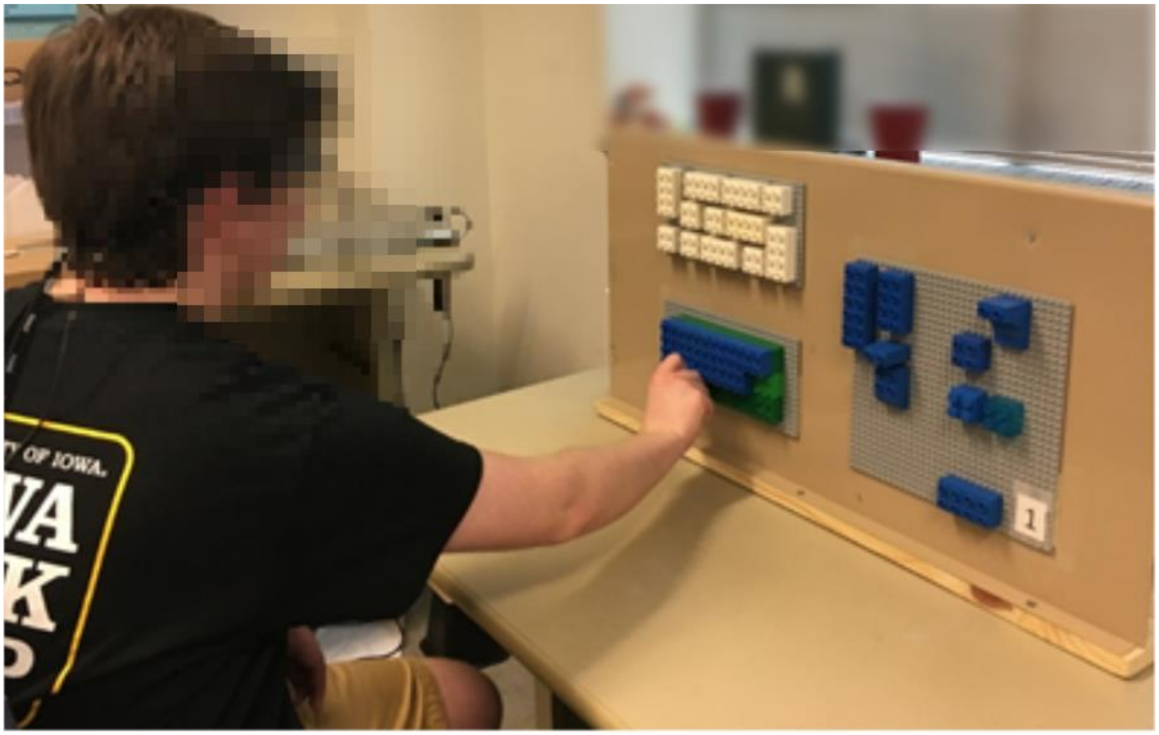

Figure 9: Participant performing assembly real environment.

To maintain the Lego® panel firmly in place on the metal panel, we glued a layer of strip magnets to the back of each panel. The magnetic strips allowed us to quickly replace assembled panels with new panels that were ready for assembly. This way participants performed the ten assembly cycles with a minimum wait time between cycles. As a final note, to compensate for the wait time in the RE between assembly cycles, the VE included a delay of five seconds between assembly cycles.

3.3.2 Assembly for Virtual Environments

The assembly in the virtual environment required participants to pick and place virtual Lego® blocks in a standing position. There was no time limit. However, participants were encouraged to perform the assembly to their best of their ability and to do the assembly as quickly as possible. Participants stood in front of a 24" computer monitor at a distance of 60 cm (Figure 11). Participants were asked to perform the assembly for ten consecutive cycles. After concluding each cycle in the VE, there was a five-second pause, and the monitor displayed a landscape image unrelated to the task.

3.3.2.1 Training Exercise

Before performing the virtual assembly task, participants were trained in how to select, move, and position virtual Lego® blocks (Figure 10). The training session required participants to place their real dominant hand over a LEAP® motion controller and to learn how to select, move, and position a block using the pinch gesture. Also, participants were instructed on how to read the visual instruction pattern, the difference between assembly blocks and distractors, and how to interpret the visual and acoustic cues.

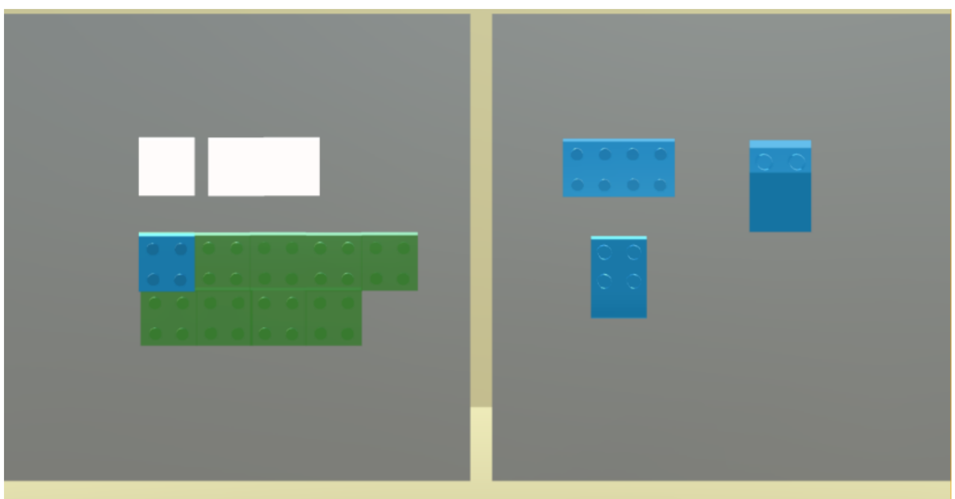

Figure 10: Virtual training exercise.

3.3.2.2 VE Posture

The purpose of the VE standing position was to prevent any visual obstruction or visual confusion. As shown in Figure 11, participants moved their real hand above the LEAP® motion controller while maintaining it below the screen. The standing posture allowed participants to keep their eyes on the screen as they observed their virtual hand moving objects on the screen without seeing their real hands (Figure 3) [165].

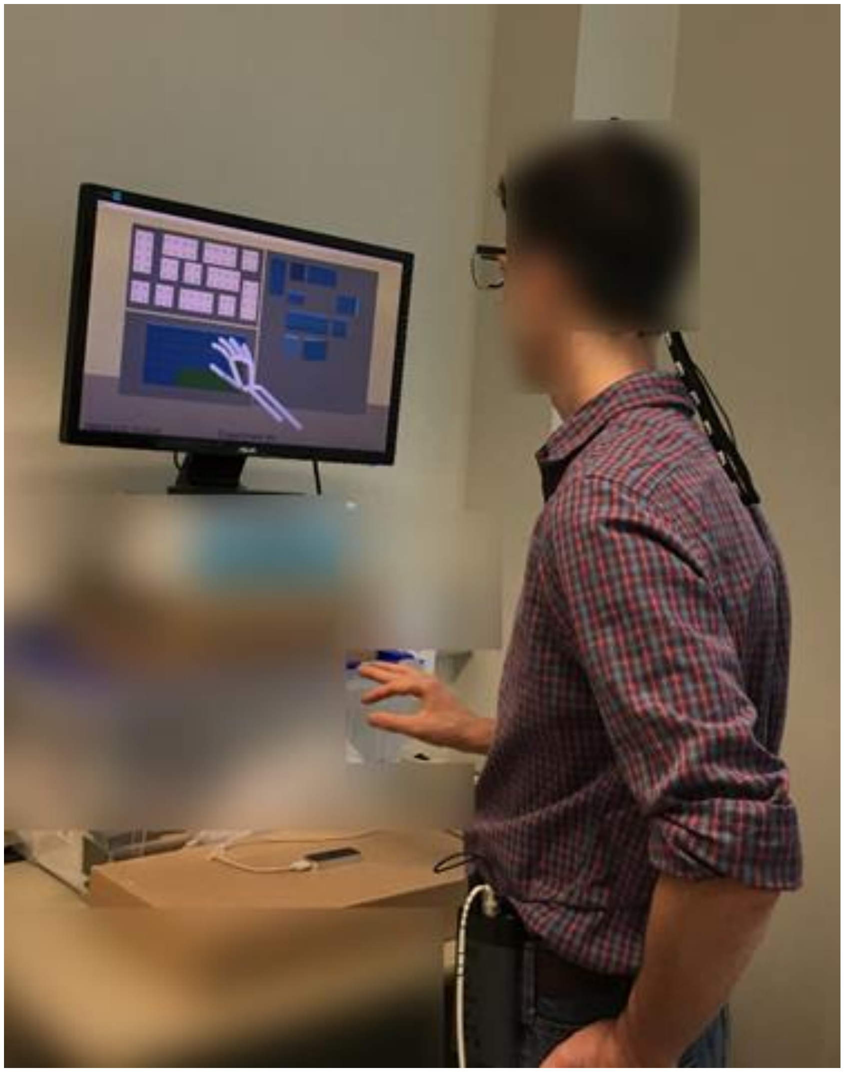

Figure 11: Participant performing an assembly task using virtual environments.

CHAPTER 4 – RESULTS

We now present the experimental results of our study. We start by providing a description of the different regions of interest and the eye tracking metrics. We then present a detailed description of the different eye tracking metrics by ROI, followed by an analysis of the assembly duration by cycle. We finish the chapter by presenting an analysis based on the proportion of fixation duration for each assembly cycles by ROI.

4.1 ROI Description and Eye Tracking Metrics

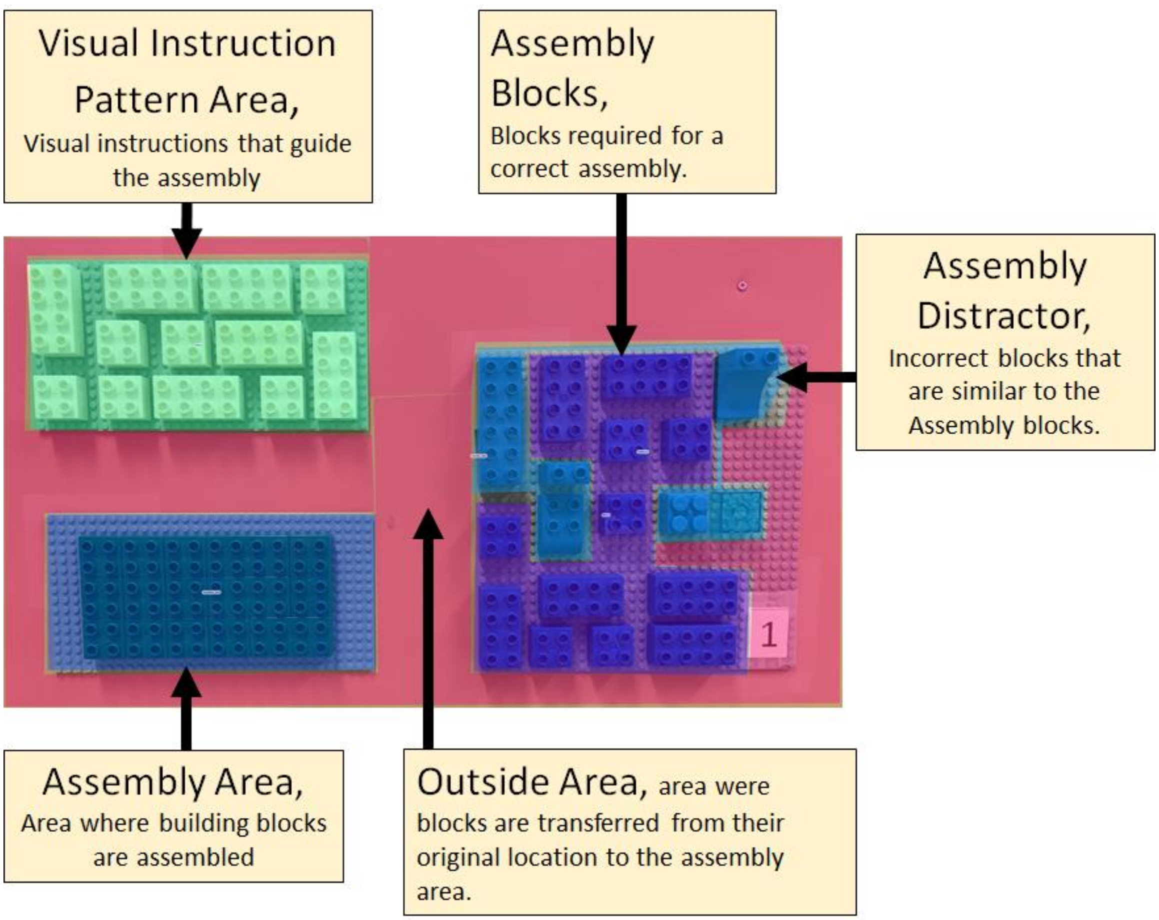

Figure 12: Regions of interest (ROI) coding and description for RE.

Figure 12 shows the ROI coding for the RE layout as previously described in Section 3.1.5 ROI coding for the RE lay-out. The different metrics for each ROI are as follows:

a) Visit count: provides the number of times the participant's foveal view is positioned over an ROI regardless of the duration of the visit.
b) Visit count duration: describes the average length of time the observer spent looking at each ROI during multiple visits.
c) Fixation count: eye fixations show a lapse when the participant's eye-gaze is relatively concentrated in one area. Therefore, this metric counts the number of eye fixations per ROI.
d) Fixation duration: this metric describes the lapse of eye fixations per ROI.

4.2 Observation Metrics by ROI

We generated four tables that describe the eye tracking metrics by ROI. Having The information contained in these tables permits the identification of the key visual areas among each ROIs as these areas capture a higher value of eye tracking metrics.

These tables use the following observation metrics:

a) The sum of the metric values by ROI. This metric is calculated as the total sum of all metric values by ROI for all 10 cycles for 30 participants.

Equation 2: Sum of metric values

$$M_n = \sum_{i=1}^{n=300} M_i$$

b) The average metric values per ROI. This metric is calculated as the total sum of all metric values by ROI for all cycles divided by 10 (number of assembly cycles per environment).

Equation 3: average metric values

$$M_{avg} = \frac{M_n}{n} \quad \text{Where:}$$

n = 300

c) The calculated difference by visual scan metric. This metric is calculated as the difference between the metric value by ROI for the sum of the tenth cycle minus the sum of the first cycle, for all participants.

Equation 4: Difference by metric

$$\Delta M = M_F - M_l \quad \text{Where:}$$

$$M_F = \sum_{i=270}^{n=300} M_i$$

$$M_l = \sum_{i=1}^{n=30} M_i$$

d) The proportion of observation metric by ROI. This metric is calculated as the total sum of each observation metric by ROI for all 10 cycles divided by the total sum all observation metrics by ROI for all 10 cycles.

Equation 5: Proportion of observation metric by ROI

$$PM_n = \frac{M_n}{\sum_{m=1}^{m=m} M_n} \qquad \text{Where}$$

m= the total number of metrics

e) Percent difference by ROI between RE and VE. This metric is calculated as a percentage of the total sum of each observation metric by ROI for all 10 cycles in the VE, divided by the total sum of each observation metric by ROI for all 10 cycles in the RE minus one divided by the sum of all observation metrics by ROI for all 10 cycles.

Equation 6: Percent difference by ROI between RE and VE

$$\Delta \mathrm{M}_E = \left(\frac{M_{nVE}}{M_{nRE}} - 1\right) * 100$$

Each eye metric table contains three different sections. The first section describes the information for the real environment; the second section describes the information for the virtual environment, and the last section shows the percent difference by ROI as described in equation 6.

Table 2 shows the visit count metric by ROI. To clarify the difference between the metrics of visit count and eye-fixation count is that visit count is calculated by adding the number of times a specific ROI is visited by the observer, while eye fixation is calculated by adding the number of fixations. Furthermore, since these ROIs are quite large, and eye fixations have an approximate range of duration of 60 ms to 250 ms milliseconds [166], an observer could perform multiple eye fixations during one visit.

Table 2: Visit Count per ROI, values in number of visits peer ROI.

Visit Count					
Real Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	10,742	15,420	10,185	19,871	6,201
M_{avg}	35.81	51.40	33.95	66.24	20.67
ΔM	25.83%	26.45%	37.36%	40.53%	51.71%
PM_n	17.21%	24.70%	16.32%	31.83%	9.93%
Virtual Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	20,888	35,872	20,258	25,568	11,531
M_{avg}	69.63	119.57	67.53	85.23	38.44
ΔM	31.54%	37.40%	48.09%	73.63%	35.17%
PM_n	18.30%	31.43%	17.75%	22.40%	10.10%
ΔM_E	94.45%	132.63%	98.90%	28.67%	85.96%

Table 3 presents the visit duration by ROI in seconds.

Table 3: Visit Duration per Count per ROI in seconds.

Visit Duration					
Real Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	2882.20	1873.79	556.40	1062.99	1588.48
M_{avg}	9.61	6.25	1.85	3.54	5.29
ΔM	37.37%	39.88%	40.63%	45.86%	54.82%
PM_n	36.19%	23.53%	6.99%	13.35%	19.95%

Virtual Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	12089.34	8163.20	1394.88	2066.23	1827.23
M_{avg}	40.30	27.21	4.65	6.89	6.09
ΔM	41.54%	26.16%	48.03%	70.04%	38.72%
PM_n	47.33%	31.96%	5.46%	8.09%	7.15%

ΔM_E	319.45%	335.65%	150.70%	94.38%	15.03%

Table 4 describes the fixation count metric by ROI.

Table 4: Fixation Count per ROI, values in number of fixations per ROI.

Fixation Count					
Real Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	13,496	8,924	2,701	4,486	7,533
M_{avg}	44.99	29.75	9.00	14.95	25.11
ΔM	36.83%	38.52%	40.86%	44.96%	54.64%
PM_n	36.34%	24.03%	7.28%	12.09%	20.27%

Virtual Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	58,496	38,753	6,773	9,057	8,653
M_{avg}	194.99	129.18	22.58	30.19	28.84
ΔM	42.06%	29.85%	48.77%	71.98%	44.49%
PM_n	48.04%	31.83%	5.54%	7.48%	7.11%

ΔM_E	333.42%	334.25%	150.78%	101.91%	14.86%

Table 5 provides results for fixation duration by ROI.

Table 5: Fixation Duration per Count per ROI in seconds.

Fixation Duration					
Real Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	2714.64	1792.35	542.03	902.12	1516.44
M_{avg}	9.05	5.97	1.81	3.01	5.05
ΔM	36.83%	38.51%	40.86%	44.96%	54.63%
PM_n	36.35%	24.00%	7.26%	12.08%	20.31%

Virtual Environment	Assembly Area	Blocks	Distractor	Outside Area	Visual instruction pattern
M_n	11696.27	7748.75	1354.56	1811.15	1736.68
M_{avg}	38.99	25.83	4.52	6.04	5.79
ΔM	42.06%	29.85%	48.77%	71.98%	44.48%
PM_n	48.04%	31.83%	5.56%	7.44%	7.13%

ΔM_E	330.86%	332.32%	149.90%	100.77%	14.52%

Figure 13 and Figure 14 show the heat maps for the first, fifth and tenth assemblies for real and virtual environments. In these maps the light green indicates a lower proportion of fixations, the yellow indicates a medium proportion of fixations, and the red indicates a high proportion of fixations.

Figure 15 presents eight different pie-charts that show the proportion metrics by ROI for Tables 2 - 5.

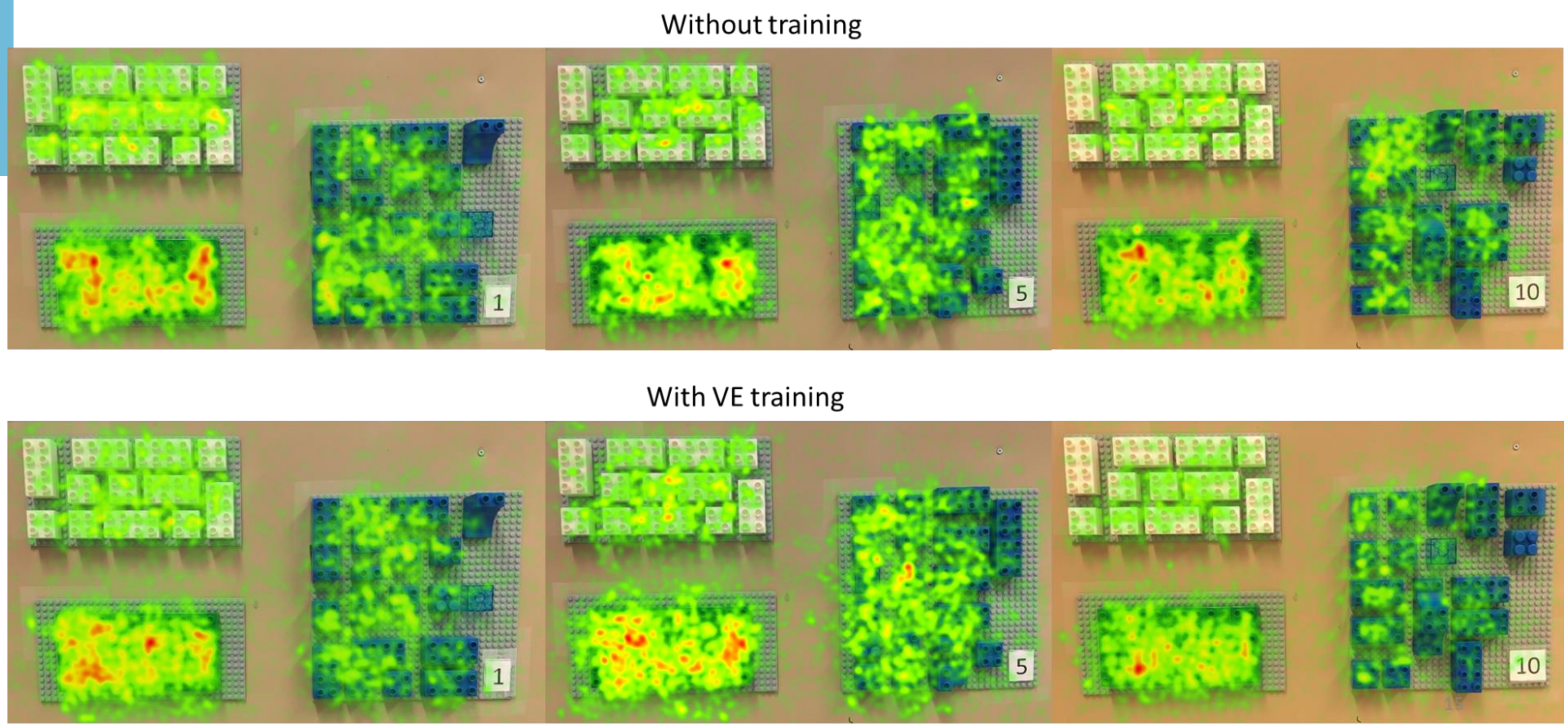

Figure 13: Heat maps for fixation duration for real environment assembly by assembly training for first, fifth, and tenth assemblies.

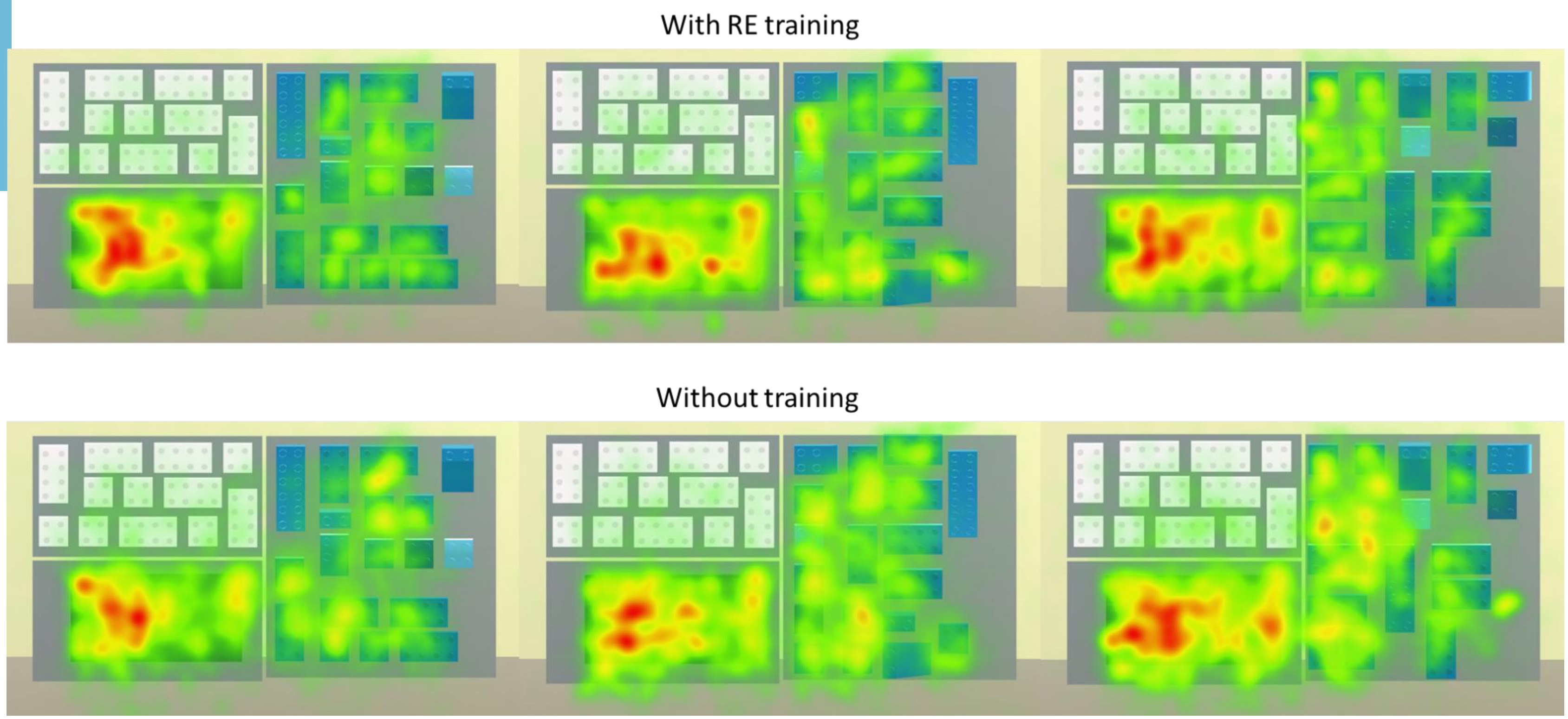

Figure 14: Heat maps for fixation duration for virtual environment assembly by assembly training for first, fifth, and tenth.

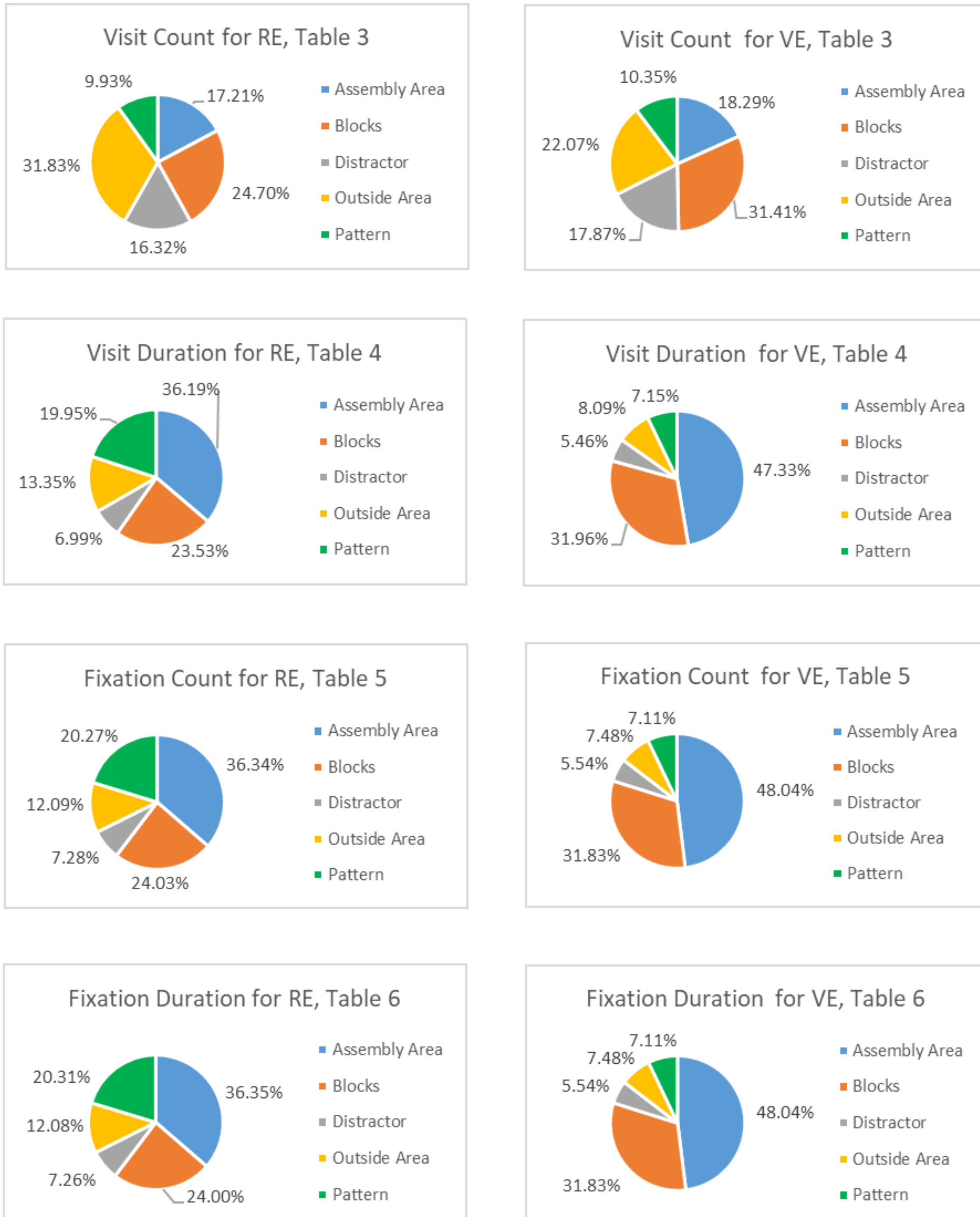

Figure 15: Metric proportion by ROI, pattern stands for visual instruction pattern.

The charts in Figure 15 indicate the key visual areas required to perform the assembly process for each metric, environment, and ROI. Two areas concentrate most of the observation proportion: assembly area and blocks, which could also be observed in the following table, Table 6.

Table 6: Cumulative percentages for Assembly Area and Blocks by Assembly environment.

	Real Environments			Virtual Environments		
Metric	Assembly Area	Blocks	Accumulative Percentages for RE	Assembly Area	Blocks	Accumulative Percentages for VE
Visit Count	17.21%	24.70%	41.91%	18.30%	31.43%	49.73%
Visit Duration	36.19%	23.53%	59.72%	47.33%	31.96%	79.29%
Fixation Count	36.34%	24.03%	60.37%	48.04%	31.83%	79.87%
Fixation Duration	36.35%	24.00%	60.35%	48.01%	31.83%	79.84%

On the other hand, although the key visual areas are the same in both environments, all eye tracking metrics have significant statistical differences for both environments with a p-value of < 0.001, having higher proportion values for the VE.

In addition, we performed a linear regression to determine whether the different eye tracking metrics (visit count, visit duration, fixation count, and fixation duration) had a statistically significant effect on the ROIs by an experimental factor.

Table 7: Summary of metrics by regions of interest for experimental factors.

			EXPERIMENTAL FACTORS						
		REGIONS OF INTEREST	Assembly Cycle	Environment (Real / Virtual)	Assembly Cycle & Environment	Order (Real 1st / Virtual 1st)	Assembly Cycle –Order	Environment - Order	Assembly-Environment-Order
METRICS	**Visit Count**	Assembly Area	√	√		√		√	
		Block	√	√	√	√		√	
		Distractor	√	√	√	√		√	
		Outside Area	√	√	√	√		√	
		Pattern	√	√	√	√		√	
	Visit Duration	Assembly Area	√	√	√				
		Block	√	√					
		Distractor	√	√	√			√	
		Outside Area	√	√	√	√		√	
		Pattern	√	√				√	
	Fixation Count	Assembly Area	√	√	√				
		Block	√	√					√
		Distractor	√	√	√			√	
		Outside Area	√	√	√	√		√	
		Pattern	√	√		√		√	
	Fixation Duration	Assembly Area	√	√	√				
		Block	√	√					√
		Distractor	√	√	√			√	
		Outside Area	√	√	√	√		√	
		Pattern	√	√		√		√	

For example, the visit count metric was recorded for each ROI (assembly area, blocks, distractors, outside area, and visual instruction pattern), and we found a significant statistical difference for this metric for all ROIs with respect to assembly cycle. Therefore, the information in Table 7 shows that the metrics of visit count, visit duration, fixation count, and fixation duration are statistically different for all ROIs for the factors of assembly cycle and environment.

4.3 Fixation Duration by Cycle by Environment and Previous Training

The information in Figures 16 and 17 describe the assembly time information by cycle for the real environment with and without previous training.

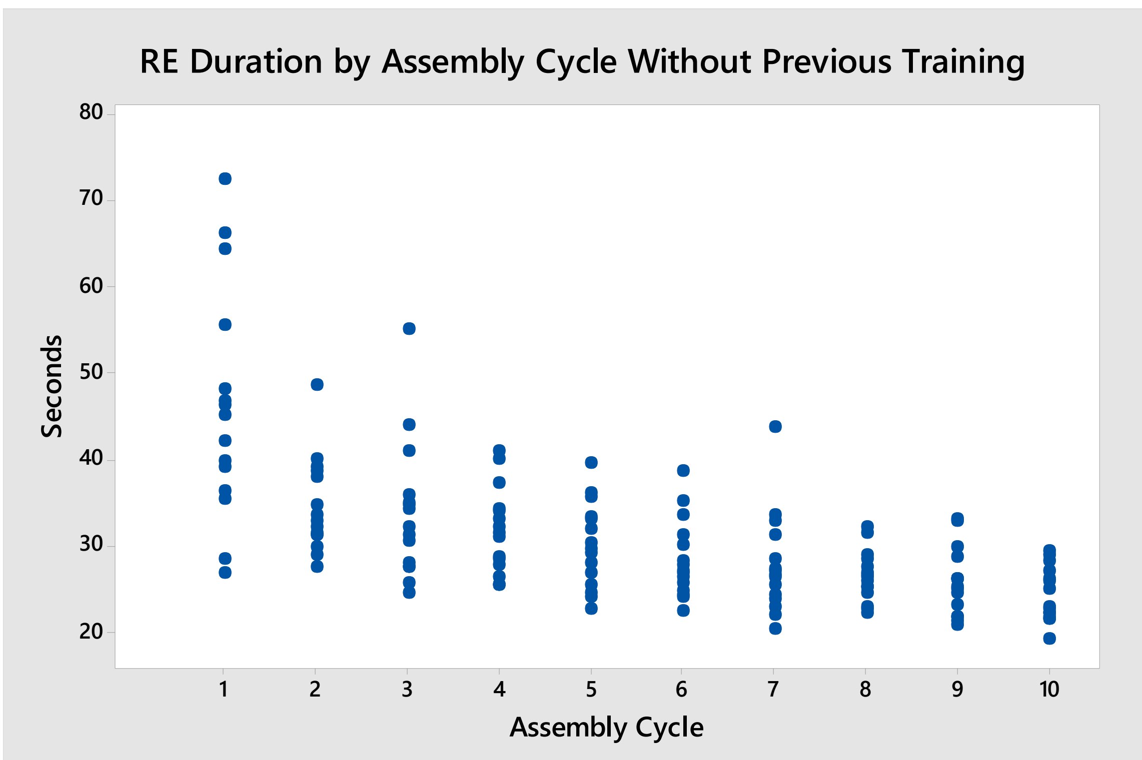

Figure 16: Assembly time by cycle for real environments without previous training.

We observe in Figures 16 and 17 that the assembly time was reduced from the first to the final assembly cycle. Moreover, we also see a lower dispersion for the duration values for the more advanced assembly cycles in comparison to the initial assembly duration cycles.

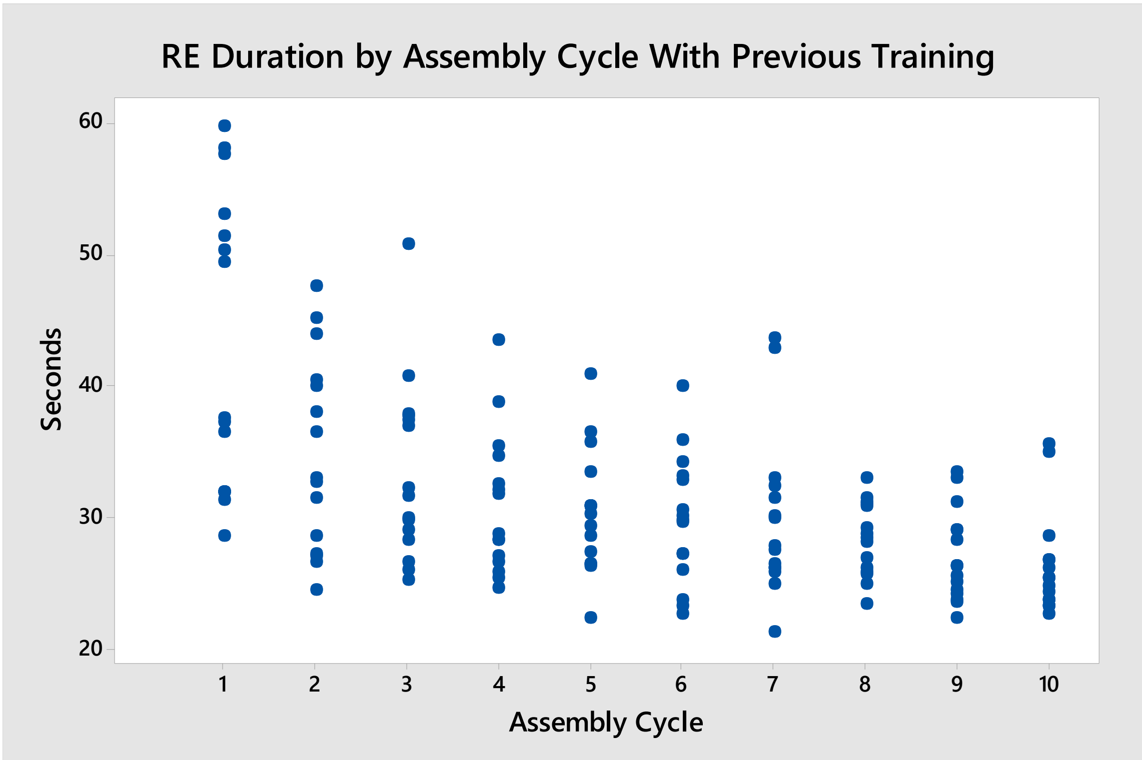

Figure 17: Assembly time by cycle for real environments with previous training.

In addition, we also note a decrease for the first assembly cycle in the range of duration values for the participants who had previous training in the virtual environment.

Similarly, Figures 18 and 19 describe the assembly time information by cycle for the virtual environment.

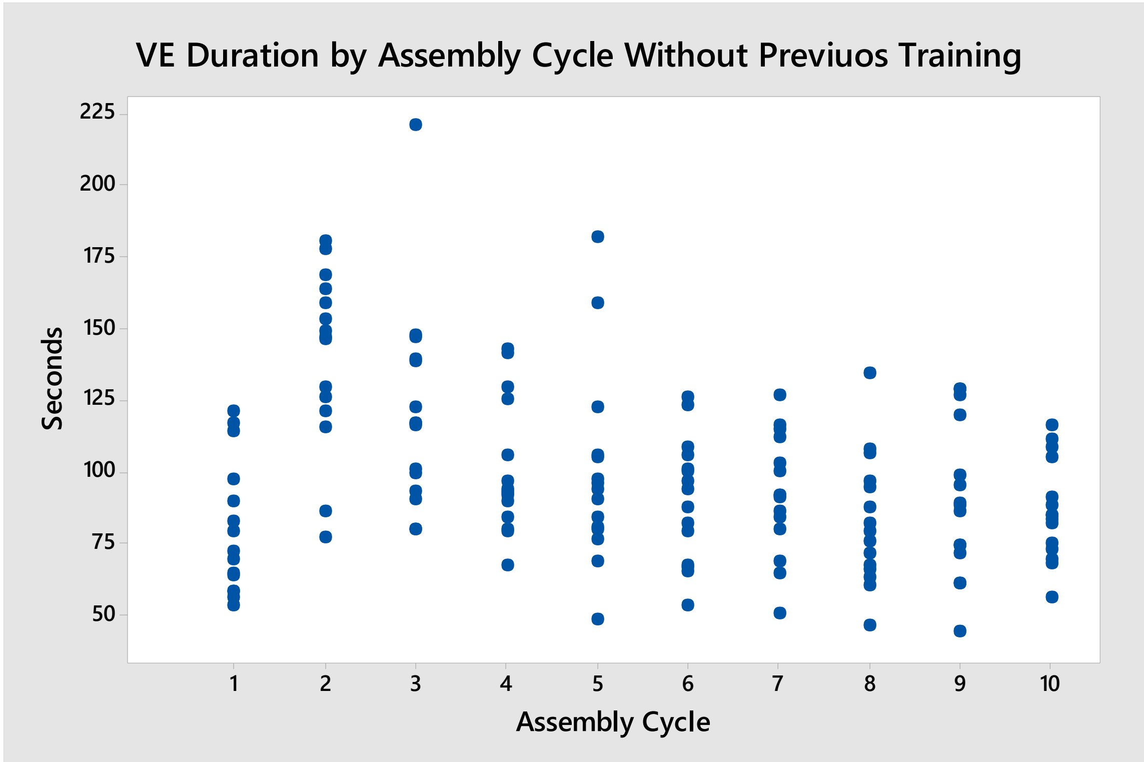

Figure 18: Assembly time by cycle for virtual environments with previous training.

As in the previous case, we observe in Figures 18 and 19 that as participants performed more cycles in virtual environments, they reduced their assembly time, and also that the assembly duration values had a lower dispersion.

On the other hand, it is interesting to note that the first cycle had a longer duration for participants with previous training in comparison to the first cycle of participants without previous training.

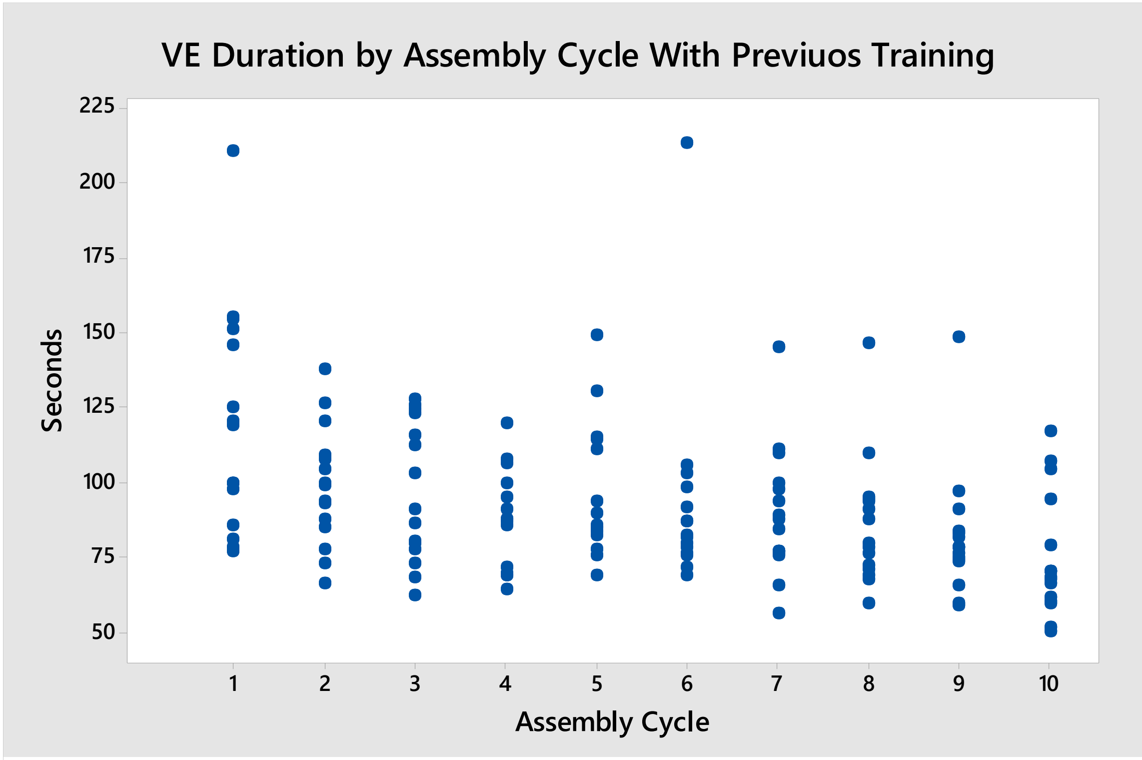

Figure 19: Assembly time by cycle for virtual environments without previous training.

In addition to the previous charts, Table 8 – 9 contain the descriptive statistics for the assembly duration in seconds by cycle for all participants with and without training for both environments.

Table 8: Assembly duration in seconds by assembly cycle for real environments with and without previous training.

Assembly duration in seconds by assembly cycle for RE with no training					
Cycle	N	Mean	SE	Mean	Std Dev
1	15	46.13	3.46	13.42	26.85
2	15	34.45	1.41	5.46	27.37
3	15	33.73	2.07	8.03	24.4
4	15	31.66	1.28	4.95	25.27
5	15	29.92	1.28	4.94	22.57
6	15	28.33	1.17	4.54	22.43
7	15	27.64	1.51	5.83	20.34
8	15	26.37	0.766	2.965	21.994
9	15	25.66	1.01	3.93	20.64
10	15	24.19	0.81	3.137	19.034

Assembly duration in seconds by assembly cycle for RE with training					
Cycle	N	Mean	SE	Mean	Std Dev
1	15	43.04	2.94	11.38	28.55
2	15	34.79	1.92	7.45	24.39
3	15	33.27	1.78	6.89	25.19
4	15	30.87	1.39	5.39	24.61
5	15	30.35	1.22	4.72	22.27
6	15	29.68	1.27	4.93	22.63
7	15	29.90	1.61	6.22	21.23
8	15	28.25	0.72	2.80	23.31
9	15	26.97	0.89	3.47	22.33
10	15	26.28	1.02	3.96	22.55

Table 9: Assembly duration in seconds by assembly cycle for virtual environments with and without previous training.

Assembly duration in seconds by assembly cycle for VE with no training					
Cycle	N	Mean	SE	Mean	Std Dev
1	15	79.47	5.98	23.14	53.13
2	15	139.62	7.95	30.8	76.64
3	15	120.41	9.27	35.92	79.66
4	15	100.58	6.02	23.31	66.78
5	15	99.00	8.77	33.96	48.09
6	15	90.03	5.53	21.43	52.85
7	15	91.74	5.42	21	50.19
8	15	82.23	5.83	22.57	45.99
9	15	89.56	6.89	26.7	44.01
10	15	86.18	4.51	17.49	55.86

Assembly duration in seconds by assembly cycle for VE with training					
Cycle	N	Mean	SE	Mean	Std Dev
1	15	119.68	9.72	37.65	76.56
2	15	98.35	5.12	19.81	66.30
3	15	96.34	5.98	23.15	61.77
4	15	87.16	4.28	16.57	64.02
5	15	96.11	5.84	22.63	68.52
6	15	93.04	9.02	34.95	68.9
7	15	89.46	5.60	21.67	55.86
8	15	85.88	5.51	21.35	59.47
9	15	81.47	5.47	21.19	58.46
10	15	74.92	5.38	20.85	49.75

4.4 Proportion of Fixation Duration by ROI

As observed in Tables 2 – 5, there is a reduction in the proportion of fixation duration for all ROIs between the first and tenth cycles. In addition, Tables 8 -9 and Figures 16 – 19 show a reduction in assembly duration per cycle for most cycles in both environments with and without previous training.

We now provide an analysis based on the proportion of fixation duration. These analyses consider the proportion of time that participants spent looking at each ROI by assembly cycle in both environments with and without training. In addition, we also use the proportion of fixation duration to understand the influence of previous training, and the Pearson correlation between the proportion of fixation duration with assembly cycle and assembly time.

4.4.1 The Influence of Previous Training

Figure 20 and Figure 21 show the proportion of fixation duration by assembly cycle for each ROI with and without previous training in both environments.

From the information from Figures 20 – 21, we observe that participants have a higher level of proportion of fixation duration for the assembly area and the blocks, in comparison to the time spent fixating on the distractor, the outside area, and the visual instruction pattern. In addition, a comparison of the fixation duration proportions for trained and untrained participants generated the following results. For real environments, VE training increased the proportion of fixation duration for the assembly area with a p-value < 0.001 and the outside area with a p-value < 0.000. On the contrary, the proportion of fixation duration was reduced for the blocks with a p-value < 0.010 and the visual instruction pattern with a p-value < 0.000. There were no significant changes in the distractor area.

On the other hand, for the VE the real environment training increased the proportion of fixation duration for the assembly area with a p-value < 0.000, but did not modify the proportion of fixation duration for the other ROIs.

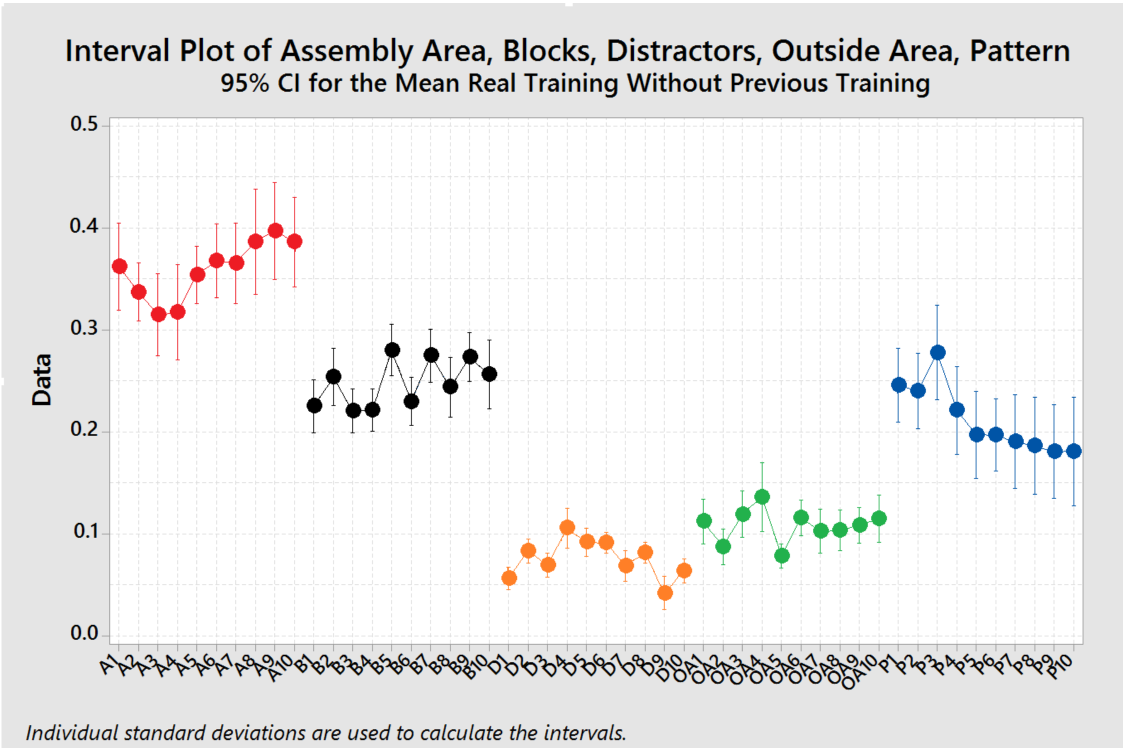

Figure 20: Proportion of fixation duration interval plot by ROIs for RE without previous training. Where A stands for assembly area, B stands for blocks, D stands for distractor, OA stands for outside area, and P stands for visual instructions.

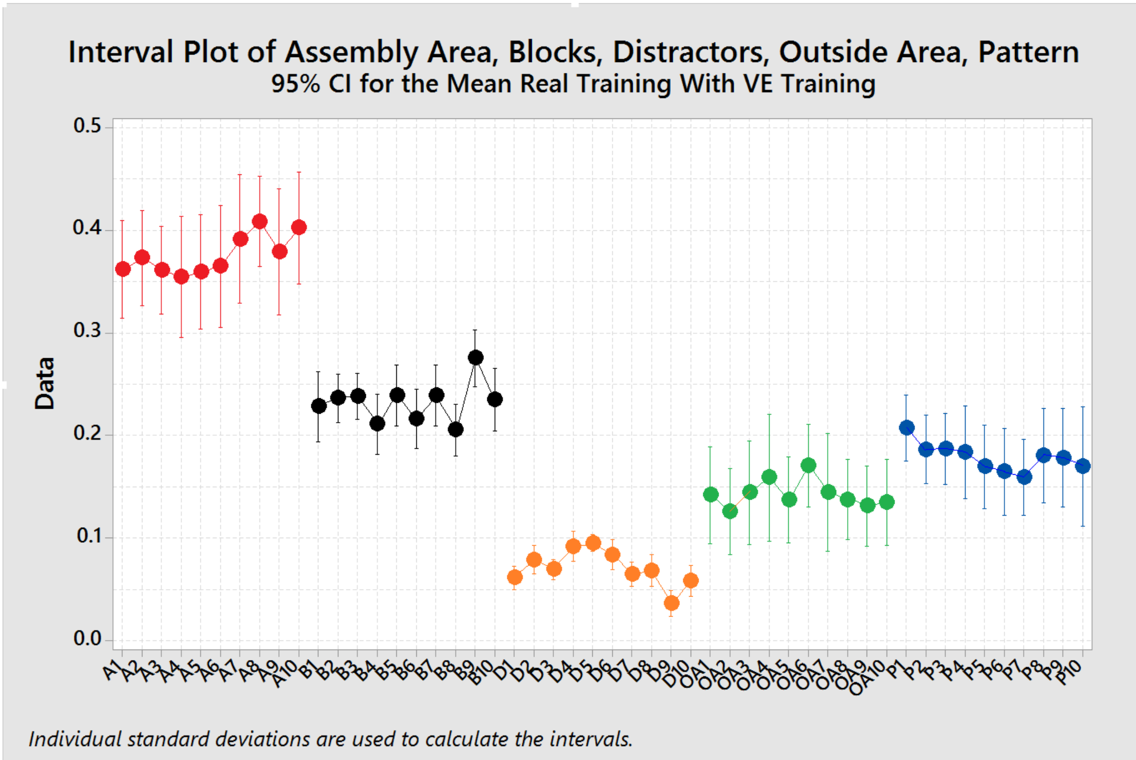

Figure 21: Proportion of fixation duration interval plot by ROIs for RE with previous training. Where A stands for assembly area, B stands for blocks, D stands for distractor, OA stands for outside area, and P stands for visual instructions.

4.4.2 Pearson Correlation Between the Proportion of Fixation Duration and Assembly Cycle

We also analyzed the influence of practice on the proportion of fixation duration for each ROI by calculating the Pearson correlation between the proportion of fixation duration for each ROI for both environments, with and without previous training.

A positive Pearson correlation value indicates that the proportion of fixation duration increased with practice. The correlation results for the real environment are described in Table 10.

Table 10: Pearson correlation between proportion of fixation duration and assembly cycle by ROI for RE and VE.

Pearson correlation between proportion of fixation duration and assembly cycle by ROI for RE				
ROIs	untrained		previously trained in VE	
	correlation	p-value	correlation	p-value
Assembly Area	0.262	0.001		statistically insignificant
Blocks	0.257	0.002		statistically insignificant
Distractor	-0.163	0.046	-0.263	0.001
Outside Area		statistically insignificant		statistically insignificant
Visual Instruction Pattern	-0.328	0.000		statistically insignificant

Pearson correlation between proportion of fixation duration and assembly cycle by ROI for VE				
ROIs	untrained		previously trained in RE	
	correlation	p-value	correlation	p-value
Assembly Area		statistically insignificant		statistically insignificant
Blocks	0.241	0.003	0.330	0.000
Distractor	-0.313	0.000		statistically insignificant
Outside Area	-0.209	0.010	-0.202	0.013
Visual Instruction Pattern		statistically insignificant		statistically insignificant

4.4.3 The Pearson Correlation Between the Proportion of Fixation Duration and Assembly Time

We also analyzed the influence of assembly time in the proportion of fixation duration for each ROI by calculating the Pearson correlation by ROI for each environment with and without previous training.

A positive Pearson correlation value indicates that the proportion of fixation duration increased for slower assemblies. Figure 22 shows an example of the scatterplot for the fixation duration proportion for the assembly area and the assembly time in RE without training.

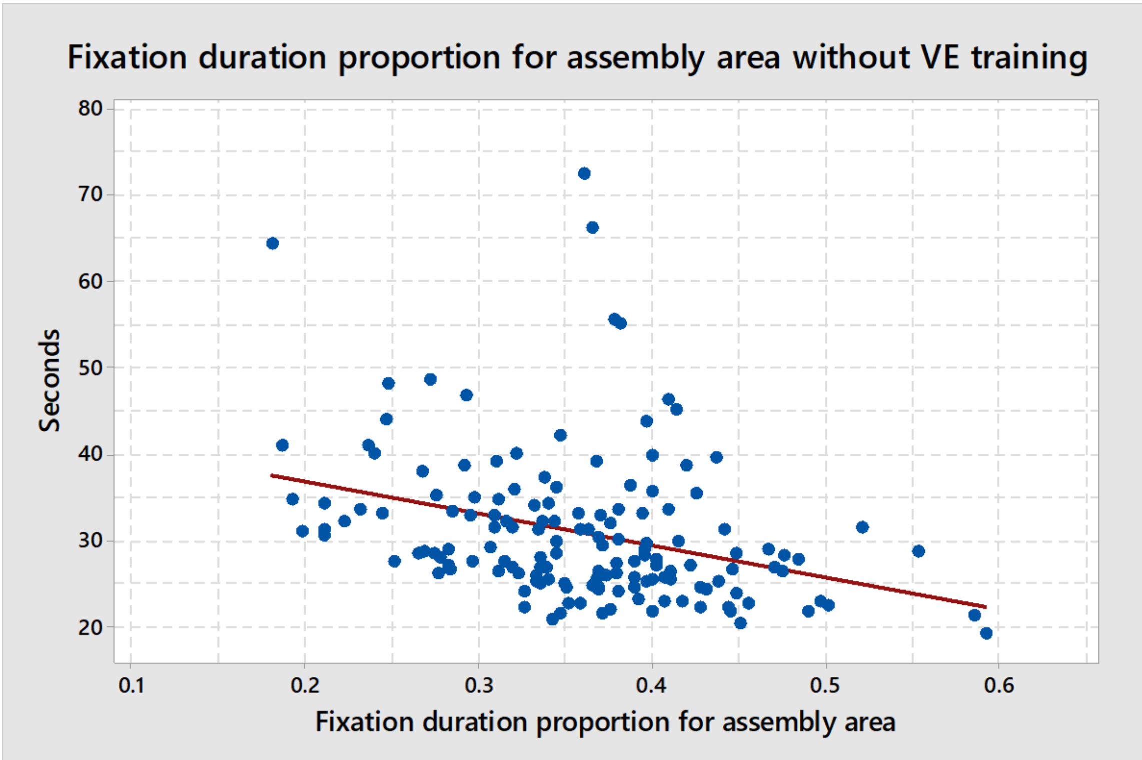

Figure 22: Scatterplot and correlation line between fixation duration proportion and assembly duration for assembly area in RE without VE training.

Table 11 shows the correlation results for fixation duration proportion and assembly duration for each ROI with and without previous training for both environments.

Table 11: Pearson correlation between proportion of fixation duration and assembly duration by ROI for RE and VE.

Pearson correlation between proportion of fixation duration and assembly duration by ROI for RE				
ROIs	untrained		Previously trained in VE	
	correlation	p-value	correlation	p-value
Assembly Area	-0.330	0.000	-0.276	0.001
Blocks		statistically insignificant		statistically insignificant
Distractor		statistically insignificant		statistically insignificant
Outside Area		statistically insignificant	0.262	0.001
Visual Instruction Pattern	0.289	0.000	0.164	0.046

Pearson correlation between proportion of fixation duration and assembly duration by ROI for VE				
ROIs	untrained		previously trained in RE	
	Correlation	p-value		p-value
Assembly Area	0.157	0.055*	0.158	0.053*
Blocks		statistically insignificant		statistically insignificant
Distractor		statistically insignificant		statistically insignificant
Outside Area		statistically insignificant		statistically insignificant
Visual Instruction Pattern	-0.181	0.027	-0.187	0.022
* statistically insignificant				

Figure 23 shows an example of the scatterplot for the fixation duration proportion for the visual instruction pattern and the assembly time in RE without training. This plot shows that participants had a higher fixation duration proportion for longer assembly cycles.

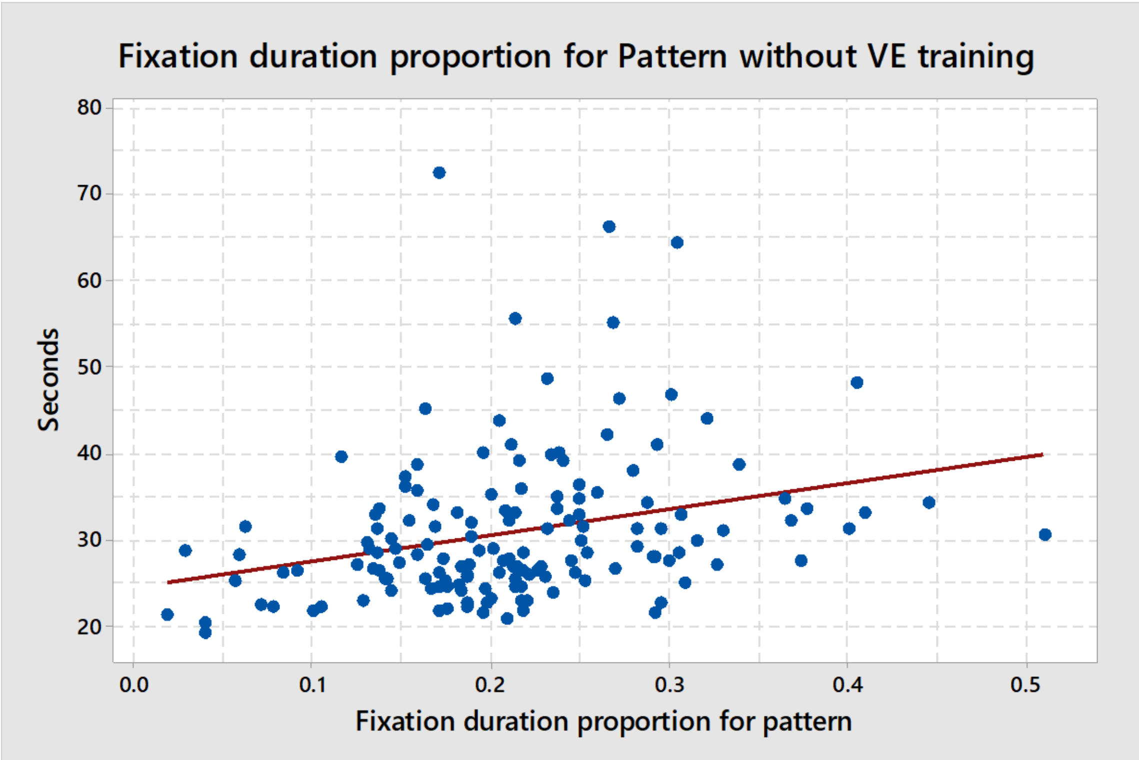

Figure 23: Scatterplot and correlation line between fixation duration proportion and assembly duration for visual instruction pattern in RE without VE training.

CHAPTER 5 – DISCUSSION

5.1 Key Visual Areas That Guide an Assembly Process

The main goal of this research was to identify the specific features that attract the observer's overt attention during an assembly task in real and virtual environments that would require an assembler to follow visual instructions in the presence of distracting objects. The results described in Table 6 and Figure 15 and in Section 4.4 show that the assembly area and the blocks received a higher proportion of fixation duration than did other regions of interest. On the other hand, these results also indicate that participants reduced the number of observations at the visual instruction pattern. Moreover, these results fluctuated with assembly practice. This finding indicates that during the initial cycles, participants considered two regions as more relevant for the task and that as they performed more assembly cycles, they reinforced their selection of these relevant areas. We consider that this was possible because assembly practice allowed participants to develop three abilities. First, participants developed the ability to recognize assembly blocks and distractors correctly. Second, participants learned some of the visual information displayed on the visual instruction pattern, resulting in less need to rely on visual instructions. Third, participants acclimated to the physical constraints of the assembly task allowing them to perform body motions faster and more accurately.

The ability to recognize relevant and irrelevant information such as assembly blocks and distractors has been described previously by several eye-tracking studies in radiology [44], [47], [48], [89], air traffic control [43] and chess [36]. This research shows that expertise allows observers to visually identify ROIs faster and more accurately. Selecting these key visual areas was very important for this task because eye gaze identified the key positions where the participant's hand had to be placed and moved as described by Johansson et al. [35].

On the other hand, we found that practice allowed participants to learn some of the visual information displayed on the visual instruction pattern. This finding does not contradict that of Ballard et al. [37], who demonstrated that participants tend to observe a visual instruction pattern rather than memorizing it. During the initial assembly cycles, participants more closely follow the visual instruction pattern, rather than memorizing it. However, as participants perform several assembly cycles, they learn some of the visual information contained in the visual instruction pattern, resulting in lower dependency on the information contained in the visual instruction pattern. This lower dependency can be observed in the heat map in Figure 13, particularly for the first and the last assembly pieces (vertical rectangles located at the top left and lower right corners).

We also described that assembly practice allowed participants to acquire physical learning; this finding agrees with the findings of Schmidt and Young [13], who describe that sequential movements serve as memory cues.

5.2 Differences Between Real and Virtual Environments

Table 6 shows that the observed proportion for the assembly area and the blocks are higher than the observed proportion for the distractors, the outside area, and the visual instruction pattern. Moreover, Table 6 also displays that the observed proportion for these key areas is greater for the virtual environment than for the real environment, and this difference is statistically significant with a p-value of < 0.001. This is a relevant finding because it shows how human behavior is affected by the interaction with virtual systems and how this interaction has to be considered when designing virtual environments.

We contend that the higher proportion in the observation of virtual environments is because of technical limitations. First, the hand tracking of the LEAP® motion controller would often encounter difficulties, especially when the hand was in motion. As a consequence, participants lost control of the virtual blocks, and they performed a new pinch gesture to acquire control of the virtual block.

Therefore, in order to prevent losing control of the virtual block, participants reduced the speed of their hand movements and also focused their vision on the assembly blocks to prevent losing control and also paid closer attention to the virtual assembly blocks in order to maintain control of them. Second, as participants used a computer monitor to interact with the virtual environment, they experienced difficulties matching their depth perception with the corresponding depth as portrayed on the computer monitor. As a result, they had to pay closer attention to the depth cues provided in our VE design, especially when picking up and placing the different blocks. This finding agrees with Carlson [63], who reported that the VE performance takes twice as long to perform in comparison to real environments.

Therefore, VE designers should consider that having visual cues will not only result in an increased mental workload [140], but will also modify how observers behave, and ultimately, how observers generate an understanding of the visual scene. This is relevant because the visual information acquired during their visual interpretation is also shared by other cognitive processes that use visual information such as memory or language [92], [93], and while visual cues could support human-computer interaction they could generate unrealistic visual interpretations that may interfere with a positive transfer of knowledge.

5.3 The Effect of Practice on Perceptual Learning and The Transfer of Knowledge

From Figures 16 – 19 and Tables 8 - 9 we learnt that participants reduced their assembly times for most cycles in both environments with and without previous training. However, knowing that participants improve their performance does not answer how their visual perception is affected by practice. The results in Section 4.4.2 show the proportion of fixation duration for each ROI.

Figures 20 -21 describe which areas receive a higher level of the proportion of fixation duration and how these proportion levels changed for each cycle. In addition, Table 10 shows cases of positive and negative correlations between the proportion of fixation duration and assembly cycle. A positive correlation indicates that as a result of practice, an ROI receives a higher level of attention. From these results, it is relevant to note that positive correlations only occur for key visual areas, while the negative correlations only happen for non-key visual areas. Therefore, the effect of practice is that participants learnt what key areas they need to observe to perform the task and they concentrated their attention in these areas.

Likewise, the correlation between the proportion of fixation duration and assembly time in Section 4.4.3 shows that for the real environment there is a positive correlation for the area most relevant to the task (the assembly area), and a negative correlation for the area that is receiving less attention as the assembly time decreases (the visual instruction pattern). This result is an effect of learning. As participants acquire more practice they reduce their need to observe the visual instruction pattern and concentrate more on what is relevant to the task.

On the other hand, it is interesting to note that this logic is reversed for the virtual environments where there is a positive correlation between the assembly area and a negative correlation between the visual instruction pattern. We consider that this result is an effect of the lack of reliability in the virtual environment. Participants had a higher proportion of fixation duration for the assembly area for assemblies that took longer to accomplish. Participants who experienced difficulty placing blocks in the virtual assembly area performed longer observations to place the blocks in the desired location.

With respect to the transfer of knowledge, it is interesting to note in Table 10 that there is a higher number of correlations between the proportion of fixation duration and assembly cycle for untrained individuals than for previously trained individuals in both environments. This means that previous training affects the learning rate of the environment that follows, as participants did not increase the proportion of their observations.

On the other hand, the lack of correlations does not mean that there is not an influence of previous training on the mean values of the proportion of fixation duration for each ROI. For the real environments, VE training increased the proportion of fixation duration for the assembly area and the outside area and produced a reduction in the proportion of fixation duration for the blocks and the visual instruction pattern. For the virtual environments with RE training, there was only an increase in the proportion of fixation duration for the assembly area. Therefore, previous training has a positive effect on the selection of key visual areas as well as on the learning rate for the second environment. As described by Carlson [63] and Hamblin [55], VE successfully transfer cognitive learning. However, this positive effect is not reflected as a statistical difference in assembly cycle times between participants with and without previous training; this might be due to the simplicity of the assembly task.

5.4 Study Limitations

This work had several technical limitations. First, as we previously described the hand tracking limitations of the LEAP® motion controller sensor affected the participants' behavior as they had to move their hand at a slower and more steady pace to prevent losing control of the virtual blocks. Second, performing an eye-tracking study on moving scenes is a challenging task because moving objects had to be mapped to specific snapshots and in some instances, the mapping was not accurate.Third, the lack of haptic information in the VE affected the participant's performance. As described by Carlson [63] and Adams[62], participants who perform assembly tasks in a VE without haptic information have a lower level of performance than participants who perform a VE task with haptic information.

On the other hand, our use of Lego® blocks instead of more complex assembly components could have prevented the generation of more complex knowledge. Having a more complex knowledge where participants must make calculate geometric calculations or that would require the user to generate a higher amount of top-down information, such as learning what the required angle for placing an object is could provide evidence of the positive transfer of knowledge between virtual and real environments.

On the other hand, our participant population was quite uniform, with all participants being male and of a similar age and technical background. This study will need to be extended to more diverse populations to generalize to produce more robust results and to provide external validity. Previous research has identified human-computer interaction differences between younger, and older participants and also between individuals with different visuospatial abilities. Therefore, it would be interesting to learn if these results would be modified for a more diverse population.

5.5 Study Implications

We accomplished the objective of developing an assembly task that was comparable in real and virtual environments. The results of this work identify the key visual areas needed for an assembly task in VEs and REs. Moreover, we also learned that the observation proportion of these key areas is affected by practice, and that as observers identify what is relevant, they spend more time looking at these visual cues.

In addition, we learned that although the visual key areas are the same in both environments, the reliability of the VE equipment modifies the way observers interact with and observe the VE. Furthermore, we also learned, that VE training helped participants to identify the key visual information needed to perform the assembly task. This is a significant contribution because the transfer of knowledge for the assembly task is mainly measured by the assembly duration time. However, we found that eye-scans can be used to learn if a participant knows what and where they have to observe.

Our study adds relevant information to the research field of visual cognition as it provides information about the relationship between visual information, assembly, and practice. Moreover, this study offers information about the similarities and differences between real and virtual environments that could be used for the development of VEs for assembly and training purposes.

CHAPTER 6 – CONCLUSION

The goal of this study was to identify the key visual areas for an assembly task in REs and VEs and to examine the relationship between these areas while taking into consideration pattern instructions, distractors, and learning curves. We successfully designed an assembly task for REs and VEs, with an appropriate level of interaction fidelity. We found that observers in both environments identified the same key visual areas and also displayed a clear decrease in their reliance on visual instructions. Performing several assembly cycles in both environments reduced the need to look at the visual instructions and increased the participants' ability to correctly discriminate between targets and distractors in both environments. Moreover, we learned that the interaction with the VE affects the way participants observed the VE, and that the VE required a higher proportion of observations for the key visual areas. We also learned that by comparing the fixation duration proportion for the different ROIs, we were able to detect if the participant knew which visual areas were needed to perform the task.

This work adds to our understanding of the relationship between the transformation of visual scans and the acquisition of knowledge. Furthermore, it shows how visual scans are affected by practice, visual instructions, and visual distractors. The results should be helpful to research on the use of visual scans as a source of training and training assessment. This is the first work in the field of visual cognition to study the development of visual learning curves for an assembly task in REs and VEs, particularly useful for the development of VE simulators, and reveals that VE developers should provide additional time for training in VEs in comparison to REs.

The study had two primary technical limitations. First, the LEAP® motion controller had some hand tracking problems that affected the behavior of the participants. Second, mapping a video feed to a series of snapshots produced some mapping errors that hindered data collection and its related analyses. Finally, our study only recruited a particular population sample; our participants were limited to male undergraduate and graduate engineering students.

In contrast, while the VE training provided information about which areas were key visually, having this knowledge did not represent an advantage in assembly time in comparison to participants who performed the assembly without training. Further research is needed to apply this knowledge to tasks where users can have a competitive advantage from knowing which areas are task-relevant, such as surgery or x-ray interpretation.

REFERENCES

[1] M. P. Groover, *Work Systems: The Methods, Measurement & Management of Work*, 1 edition. Upper Saddle River, NJ: Pearson, 2006.

[2] G. Smith, "Using microwave Doppler radar in automated manufacturing applications," *Retrosp. Theses Diss.*, Jan. 2004.

[3] J. Browne, D. Dubois, K. Rathmill, S. P. Sethi, K. E. Stecke, and others, "Classification of flexible manufacturing systems," *FMS Mag.*, vol. 2, no. 2, pp. 114–117, 1984.

[4] H. A. ElMaraghy, "Flexible and reconfigurable manufacturing systems paradigms," *Int. J. Flex. Manuf. Syst.*, vol. 17, no. 4, pp. 261–276, Oct. 2005.

[5] M. p. j. Pepper and T. a. Spedding, "The evolution of lean Six Sigma," *Int. J. Qual. Reliab. Manag.*, vol. 27, no. 2, pp. 138–155, Jan. 2010.

[6] J. P. Womack, D. T. Jones, and D. Roos, *The Machine That Changed the World: The Story of Lean Production-- Toyota's Secret Weapon in the Global Car Wars That Is Now Revolutionizing World Industry*, Reprint edition. Place of publication not identified: Free Press, 2007.

[7] N. N. Gindy and S. M. Saad, "Flexibility and responsiveness of machining environments," *Integr. Manuf. Syst.*, vol. 9, no. 4, pp. 218–227, Aug. 1998.

[8] M. P. Groover, *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems 5th Edition*, 5 edition. Hoboken, NJ: Wiley, 2012.

[9] S. J. Hu, X. Zhu, H. Wang, and Y. Koren, "Product variety and manufacturing complexity in assembly systems and supply chains," *CIRP Ann. - Manuf. Technol.*, vol. 57, no. 1, pp. 45–48, 2008.

[10] M. Yoshimura *et al.*, "A rapid analysis method for production line design," *Int. J. Prod. Res.*, vol. 44, no. 6, pp. 1171–1192, Mar. 2006.

[11] X. Wang, S. K. Ong, and A. Y. C. Nee, "A comprehensive survey of augmented reality assembly research," *Adv. Manuf.*, vol. 4, no. 1, pp. 1–22, Mar. 2016.

[12] S. Hoedt, A. Claeys, H. V. Landeghem, and J. Cottyn, "Evaluation Framework for Virtual Training within Mixed-Mo del Manual Assembly," *IFAC-Pap.*, vol. 49, no. 12, pp. 261–266, Jan. 2016.

[13] R. A. Schmidt and D. E. Young, "Transfer of Movement Control in Motor Skill Learning," Apr. 1986.

[14] S. Doane, Y. W. Sohn, and B. Schreiber, "The role of processing strategies in the acquisition and transfer of a cognitive skill," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 25, no. 5, pp. 1390–1410, 1999.

[15] J. Gold, P. J. Bennett, and A. B. Sekuler, "Signal but not noise changes with perceptual learning," *Nature*, vol. 402, no. 6758, pp. 176–178, Nov. 1999.

[16] D. Schuster, J. Rivera, B. C. Sellers, S. M. Fiore, and F. Jentsch, "Perceptual training for visual search," *Ergonomics*, vol. 0, no. 0, pp. 1–15.

[17] T. P. Wright, "Factors Affecting the Cost of Airplanes," *J. Aeronaut. Sci.*, vol. 3, no. 4, pp. 122–128, 1936.

[18] L. E. Yelle, "The Learning Curve: Historical Review and Comprehensive Survey," *Decis. Sci.*, vol. 10, no. 2, pp. 302–328, Apr. 1979.

[19] J. G. Carlson, "How Management Can Use the Improvement Phenomenon," *Calif. Manage. Rev.*, vol. 3, no. 2, pp. 83–94, Jan. 1961.

[20] P. G. Schyns and A. Oliva, "From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition," *Psychol. Sci.*, vol. 5, no. 4, pp. 195–200, Jul. 1994.
[21] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning Low-Level Vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, Oct. 2000.
[22] S. E. Palmer and J. L. Brooks, "Edge-region grouping in figure-ground organization and depth perception," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 34, no. 6, pp. 1353–1371, 2008.
[23] A. Oliva and A. Torralba, "Chapter 2 Building the gist of a scene: the role of global image features in recognition," in *Progress in Brain Research*, vol. Volume 155, Part B, S. L. M. S. Martinez-Conde, Ed. Elsevier, 2006, pp. 23–36.
[24] "High-Level Vision | MIT CogNet." [Online]. Available: http://cognet.mit.edu.proxy.lib.uiowa.edu/book/high-level-vision. [Accessed: 01-Nov-2016].
[25] A. Oliva, "Gist of the scene," *Neurobiol. Atten.*, vol. 696, p. 64, 2005.
[26] A. L. Yarbus, *Eye Movements and Vision*. 1973.
[27] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends Cogn. Sci.*, vol. 7, no. 11, pp. 498–504, Nov. 2003.
[28] M. Land, N. Mennie, and J. Rusted, "The Roles of Vision and Eye Movements in the Control of Activities of Daily Living," *Perception*, vol. 28, no. 11, pp. 1311–1328, Nov. 1999.
[29] "Where we look when we steer," *Nat. Lond.*, vol. 369, no. 6483, pp. 742–744, Jun. 1994.
[30] F. I. Kandil, A. Rotter, and M. Lappe, "Driving is smoother and more stable when using the tangent point," *J. Vis.*, vol. 9, no. 1, pp. 11–11, Jan. 2009.
[31] M. F. Land and M. Hayhoe, "In what ways do eye movements contribute to everyday activities?," *Vision Res.*, vol. 41, no. 25–26, pp. 3559–3565, Nov. 2001.
[32] K. A. Turano, D. R. Geruschat, and F. H. Baker, "Fixation behavior while walking: persons with central visual field loss," *Vision Res.*, vol. 42, no. 23, pp. 2635–2644, Oct. 2002.
[33] B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard, "Eye guidance in natural vision: Reinterpreting salience," *J. Vis.*, vol. 11, no. 5, pp. 5–5, May 2011.
[34] S. Thorpe, D. Fize, and C. Marlot, "Speed Of Processing in the Human Visual System," *Nature*, vol. 381, no. 6582, pp. 520–522, 1996.
[35] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, "Eye–Hand Coordination in Object Manipulation," *J. Neurosci.*, vol. 21, no. 17, pp. 6917–6932, Sep. 2001.
[36] E. M. Reingold and H. Sheridan, "Eye movements and visual expertise in chess and medicine," in *Oxford Handbook on Eye Movements*, Oxford University Press, 2011, pp. 528–550.
[37] D. H. Ballard, M. M. Hayhoe, and J. B. Pelz, "Memory Representations in Natural Tasks," *J. Cogn. Neurosci.*, vol. 7, no. 1, pp. 66–80, Jan. 1995.
[38] D. Litchfield, L. J. Ball, T. Donovan, D. J. Manning, and T. Crawford, "Learning from others: effects of viewing another person's eye movements while searching for chest nodules," pp. 691715–691715, Mar. 2008.

[39] L. Dempere-Marco, X.-P. Hu, S. L. S. MacDonald, S. M. Ellis, D. M. Hansell, and G.-Z. Yang, “The use of visual search for knowledge gathering in image decision support,” *IEEE Trans. Med. Imaging*, vol. 21, no. 7, pp. 741–754, Jul. 2002.
[40] S. Sadasivan, J. S. Greenstein, A. K. Gramopadhye, and A. T. Duchowski, “Use of Eye Movements As Feedforward Training for a Synthetic Aircraft Inspection Task,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2005, pp. 141–149.
[41] Z. Kang and S. J. Landry, “Using Scanpaths as a Learning Method for a Conflict Detection Task of Multiple Target Tracking,” *Hum. Factors*, vol. 56, no. 6, pp. 1150–1162, Sep. 2014.
[42] Z. Kang and S. J. Landry, “Capturing and Analyzing Visual Groupings of Multiple Moving Targets in an Aircraft Conflict Detection Task Using Eye Movements,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 54, no. 23, pp. 1906–1910, Sep. 2010.
[43] Z. K. Ph.D and S. J. Landry, “Top-Down Approach for a Linguistic Fuzzy Logic Model,” *Cybern. Syst.*, vol. 45, no. 1, pp. 39–55, Jan. 2014.
[44] E. M. Kok, A. B. H. de Bruin, S. G. F. Robben, and J. J. G. van Merriënboer, “Looking in the Same Manner but Seeing it Differently: Bottom-up and Expertise Effects in Radiology,” *Appl. Cogn. Psychol.*, vol. 26, no. 6, pp. 854–862, 2012.
[45] M. Pusic, M. Pecaric, and K. Boutis, “How Much Practice Is Enough? Using Learning Curves to Assess the Deliberate Practice of Radiograph Interpretation,” *Acad. Med.*, vol. 86, no. 6, pp. 731–736, Jun. 2011.
[46] E. A. Krupinski, “Visual scanning patterns of radiologists searching mammograms,” *Acad. Radiol.*, vol. 3, no. 2, pp. 137–144, Feb. 1996.
[47] E. A. Krupinski, W. G. Berger, W. J. Dallas, and H. Roehrig, “Searching for nodules: What features attract attention and influence detection?,” *Acad. Radiol.*, vol. 10, no. 8, pp. 861–868, Aug. 2003.
[48] D. Manning, S. Ethell, T. Donovan, and T. Crawford, “How do radiologists do it? The influence of experience and training on searching for chest nodules,” *Radiography*, vol. 12, no. 2, pp. 134–142, May 2006.
[49] E. A. Krupinski, “Visual Search of Mammographic Images: Influence of Lesion Subtlety1,” *Acad. Radiol.*, vol. 12, no. 8, pp. 965–969, Aug. 2005.
[50] E. D. Reichle, K. Rayner, and A. Pollatsek, “The E-Z Reader model of eye-movement control in reading: Comparisons to other models,” *Behav. Brain Sci. N. Y.*, vol. 26, no. 4, pp. 445-76; discussion 477-526, Aug. 2003.
[51] K. Rayner, “Eye movements in reading and information processing: 20 years of research,” *Psychol. Bull.*, vol. 124, no. 3, pp. 372–422, 1998.
[52] H. E. Blanchard, G. W. McConkie, D. Zola, and G. S. Wolverton, “Time course of visual information utilization during fixations in reading,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 10, no. 1, pp. 75–89, 1984.
[53] M. Slawsky, “Transitioning from Student to Teacher in the Master-Apprentice Model of Piano Pedagogy: An Exploratory Study of Challenges, Solutions, Resources, Reflections, and Suggestions for the Future,” *Grad. Theses Diss.*, Jan. 2011.

[54] "Virtual environment | Define Virtual environment at Dictionary.com." [Online]. Available: http://www.dictionary.com/browse/virtual-environment. [Accessed: 08-Jan-2017].

[55] C. J. Hamblin, "Transfer of training from virtual reality environments," Wichita State University, 2005.

[56] R. B. Loftin and P. Kenney, "Training the Hubble space telescope flight team," *IEEE Comput. Graph. Appl.*, vol. 15, no. 5, pp. 31–37, Sep. 1995.

[57] A. G. Gallagher, N. E. Seymour, J.-A. Jordan-Black, B. P. Bunting, K. McGlade, and R. M. Satava, "Prospective, Randomized Assessment of Transfer of Training (ToT) and Transfer Effectiveness Ratio (TER) of Virtual Reality Simulation Training for Laparoscopic Skill Acquisition:," *Ann. Surg.*, vol. 257, no. 6, pp. 1025–1031, Jun. 2013.

[58] A. P. Byrd, R. T. Stone, R. G. Anderson, and K. Woltjer, "The Use of Virtual Welding Simulators to Evaluate Experienced Welders," *Weld. J.*, vol. 94, no. 12, p. 389S–395S, 2015.

[59] H. Alm and L. Nilsson, "Changes in driver behaviour as a function of handsfree mobile phones— A simulator study," *Accid. Anal. Prev.*, vol. 26, no. 4, pp. 441–451, Aug. 1994.

[60] J. P. Bliss, P. D. Tidwell, and M. A. Guest, "The Effectiveness of Virtual Reality for Administering Spatial Navigation Training to Firefighters," *Presence Teleoperators Virtual Environ.*, vol. 6, no. 1, pp. 73–86, Feb. 1997.

[61] D. Waller, E. Hunt, and D. Knapp, "The Transfer of Spatial Knowledge in Virtual Environment Training," *Presence*, vol. 7, no. 2, pp. 129–143, Apr. 1998.

[62] R. J. Adams, D. Klowden, and B. Hannaford, "Virtual Training for a Manual Assembly Task," Oct. 2001.

[63] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse, "Virtual Training: Learning Transfer of Assembly Tasks," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 6, pp. 770–782, Jun. 2015.

[64] K. Mania, T. Troscianko, R. Hawkes, and A. Chalmers, "Fidelity Metrics for Virtual Environment Simulations Based on Spatial Memory Awareness States," *Presence Teleoperators Virtual Environ.*, vol. 12, no. 3, pp. 296–310, Jun. 2003.

[65] L. Argote and P. Ingram, "Knowledge Transfer: A Basis for Competitive Advantage in Firms," *Organ. Behav. Hum. Decis. Process.*, vol. 82, no. 1, pp. 150–169, May 2000.

[66] K. Nemire and S. R. Ellis, "Calibration and evaluation of virtual environment displays," in *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*, 1993, pp. 33–40.

[67] N. Mourkoussis, F. Rivera, T. Troscianko, R. Hawkes, P. L. Watten, and K. Mania, "Towards perceptual fidelity: Slant perception in real and interactive virtual environments," *Int. J. Hum.-Comput. Stud.*, vol. 70, no. 11, pp. 781–793, Nov. 2012.

[68] D. Waller, D. Knapp, and E. Hunt, "Spatial Representations of Virtual Mazes: The Role of Visual Fidelity and Individual Differences," *Hum. Factors*, vol. 43, no. 1, pp. 147–158, Mar. 2001.

[69] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe, "Task and context determine where you look," *J. Vis.*, vol. 7, no. 14, pp. 16–16, Nov. 2007.

[70] J. Triesch, D. H. Ballard, M. M. Hayhoe, and B. T. Sullivan, “What you see is what you need,” *J. Vis.*, vol. 3, no. 1, pp. 9–9, Jan. 2003.
[71] S. Jim, *Field Guide to Visual and Ophthalmic Optics | Publications: SPIE*, vol. FG04. Bellingham, WA USA: SPIE, 2004.
[72] M. Millodot, *Dictionary of Optometry and Visual Science*. Elsevier Health Sciences, 2014.
[73] R. L. Van Metter, J. Beutel, and H. L. Kundel, Eds., *Handbook of Medical Imaging, Volume 1. Physics and Psychophysics*. 1000 20th Street, Bellingham, WA 98227-0010 USA: SPIE, 2000.
[74] H. Strasburger, I. Rentschler, and M. Jüttner, “Peripheral vision and pattern recognition: A review,” *J. Vis.*, vol. 11, no. 5, pp. 13–13, May 2011.
[75] E. A. B. Over, “Quantitative analysis of saccadic search strategy.,” Universiteit van Utrecht, Utrecht, The Netherlands, 2007.
[76] H. Collewijn, C. J. Erkelens, and R. M. Steinman, “Binocular co-ordination of human horizontal saccadic eye movements.,” *J. Physiol.*, vol. 404, no. 1, pp. 157–182, Oct. 1988.
[77] S. Thorpe, D. Fize, and C. Marlot, “Speed of processing in the human visual system,” *Nature*, vol. 381, no. 6582, pp. 520–522, Jun. 1996.
[78] M. C. Potter, “Short-term conceptual memory for pictures,” *J. Exp. Psychol. [Hum. Learn.]*, vol. 2, no. 5, pp. 509–522, 1976.
[79] I. Biederman and P. C. Gerhardstein, “Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 19, no. 6, pp. 1162–1182, 1993.
[80] D. Navon, “Forest before trees: The precedence of global features in visual perception,” *Cognit. Psychol.*, vol. 9, no. 3, pp. 353–383, Jul. 1977.
[81] S. P. Liversedge and J. M. Findlay, “Saccadic eye movements and cognition,” *Trends Cogn. Sci.*, vol. 4, no. 1, pp. 6–14, Jan. 2000.
[82] E. A. B. Over, I. T. C. Hooge, B. N. S. Vlaskamp, and C. J. Erkelens, “Coarse-to-fine eye movement strategy in visual search,” *Vision Res.*, vol. 47, no. 17, pp. 2272–2280, Aug. 2007.
[83] M. Alzubaidi, J. Black, J.A., A. Patel, and S. Panchanathan, “Conscious vs. subconscious perception, as a function of radiological expertise,” in *22nd IEEE International Symposium on Computer-Based Medical Systems, 2009. CBMS 2009*, 2009, pp. 1–8.
[84] J. H. Goldberg and X. P. Kotval, “Computer interface evaluation using eye movements: methods and constructs,” *Int. J. Ind. Ergon.*, vol. 24, no. 6, pp. 631–645, Oct. 1999.
[85] “Eye movement indices of mental workload.” [Online]. Available: http://www.sciencedirect.com/science/article/pii/000169189090067P. [Accessed: 11-Feb-2017].
[86] K. F. Van Orden, W. Limbert, S. Makeig, and T.-P. Jung, “Eye Activity Correlates of Workload during a Visuospatial Memory Task,” *Hum. Factors*, vol. 43, no. 1, pp. 111–121, Mar. 2001.
[87] Z. Kang, E. J. Bass, and D. W. Lee, “Air traffic controllers’ visual scanning, aircraft selection, and comparison strategies in support of conflict detection,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 58, no. 1, pp. 77–81, Sep. 2014.

[88] P. R. Montague, S. E. Hyman, and J. D. Cohen, "Computational roles for dopamine in behavioural control," *Nature*, vol. 431, no. 7010, pp. 760–767, Oct. 2004.
[89] D. Manning, "Cognitive factors in reading medical images a survey of cognitive factors and models of medical image interpretation," in *The Handbook of Medical Image Perception and Techniques*, E. Samei and E. Krupinski, Eds. Cambridge University Press, 2009.
[90] B. J. Jennings and J. Martinovic, "Luminance and color inputs to mid-level and high-level vision," *J. Vis.*, vol. 14, no. 2, pp. 9–9, Feb. 2014.
[91] "Vision and Visual Attention." [Online]. Available: http://www.cep.ucsb.edu/topics/vision.htm. [Accessed: 01-Jan-2017].
[92] B. J. Baars, *A Cognitive Theory of Consciousness*. Cambridge University Press, 1993.
[93] P. Cavanagh, "Visual cognition," *Vision Res.*, vol. 51, no. 13, pp. 1538–1551, Jul. 2011.
[94] W. G. Hayward, "After the viewpoint debate: where next in object recognition?," *Trends Cogn. Sci.*, vol. 7, no. 10, pp. 425–427, Oct. 2003.
[95] D. Marr and H. K. Nishihara, "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 200, no. 1140, pp. 269–294, 1978.
[96] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychol. Rev.*, vol. 94, no. 2, pp. 115–147, 1987.
[97] T. Poggio and S. Edelman, "A Network That Learns to Recognize Three-Dimensional Objects," *Nat. Lond.*, vol. 343, no. 6255, pp. 263–6, Jan. 1990.
[98] S. Ullman and R. Basri, "Recognition by linear combinations of models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 10, pp. 992–1006, 1991.
[99] H. H. Buelthoff, S. Y. Edelman, and M. J. Tarr, "How are Three-Deminsional Objects Represented in the Brain?," Apr. 1994.
[100] S. Ullman, "Three-dimensional object recognition based on the combination of views," *Cognition*, vol. 67, no. 1–2, pp. 21–44, Jul. 1998.
[101] S. Edelman, T. Poggio, and D. W. Thompson, *Representations in High-Level Vision: Reassessing the Inverse Optics Paradigm*. 1989.
[102] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends Cogn. Sci.*, vol. 11, no. 12, pp. 520–527, Dec. 2007.
[103] M. Bar, "A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition," *J. Cogn. Neurosci.*, vol. 15, no. 4, pp. 600–609, May 2003.
[104] R. Patterson, L. Fournier, B. J. Pierce, M. D. Winterbottom, and L. M. Tripp, "System Dynamics Modeling of the Time Course of the Recognition-Primed Decision Model," *J. Cogn. Eng. Decis. Mak.*, vol. 3, no. 3, pp. 253–279, Sep. 2009.
[105] S. M. Kosslyn, "You can play 20 questions with nature and win: Categorical versus coordinate spatial relations as a case study," *Neuropsychologia*, vol. 44, no. 9, pp. 1519–1523, 2006.
[106] S. M. Kosslyn and S. M. Kosslyn, *Image and Brain: The Resolution of the Imagery Debate*. MIT Press, 1996.
[107] D. Marr, *Vision: a computational investigation into the human representation and processing of visual information*, 1ST ed. San Francisco, CA: W H Freeman Limited, 1982.

[108] J. E. Hummel and B. J. Stankiewicz, “Two Roles for Attention in Shape Perception: A Structural Description Model of Visual Scrutiny,” *Vis. Cogn.*, vol. 5, no. 1–2, pp. 49–79, Mar. 1998.
[109] D. H. Foster and S. J. Gilson, “Recognizing novel three–dimensional objects by summing signals from parts and views,” *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 269, no. 1503, pp. 1939–1947, Sep. 2002.
[110] A.-M. Brouwer, V. H. Franz, and K. R. Gegenfurtner, “Differences in fixations between grasping and viewing objects,” *J. Vis.*, vol. 9, no. 1, pp. 18–18, Jan. 2009.
[111] E. H. Cohen, B. S. Schnitzer, T. M. Gersch, M. Singh, and E. Kowler, “The relationship between spatial pooling and attention in saccadic and perceptual tasks,” *Vision Res.*, vol. 47, no. 14, pp. 1907–1923, Jun. 2007.
[112] J. M. Findlay and V. Brown, “Eye scanning of multi-element displays: I. Scanpath planning,” *Vision Res.*, vol. 46, no. 1–2, pp. 179–195, Jan. 2006.
[113] P. He and E. Kowler, “Saccadic localization of eccentric forms,” *JOSA A*, vol. 8, no. 2, pp. 440–449, Feb. 1991.
[114] F. D. Rose, E. A. Attree, B. M. Brooks, D. M. Parslow, and P. R. Penn, “Training in virtual environments: transfer to real world tasks and equivalence to real task training,” *Ergonomics*, vol. 43, no. 4, pp. 494–511, Apr. 2000.
[115] G. Lintern, S. N. Roscoe, J. M. Koonce, and L. D. Segal, “Transfer of Landing Skills in Beginning Flight Training,” *Hum. Factors*, vol. 32, no. 3, pp. 319–327, Jun. 1990.
[116] B. Johns, “The creation and validation of an augmented reality orthopaedic drilling simulator for surgical training,” *Theses Diss.*, May 2014.
[117] A. Seth, J. M. Vance, and J. H. Oliver, “Virtual reality for assembly methods prototyping: a review,” *Virtual Real.*, vol. 15, no. 1, pp. 5–20, Mar. 2011.
[118] S. Jayaram, H. I. Connacher, and K. W. Lyons, “Virtual assembly using virtual reality techniques,” *Comput.-Aided Des.*, vol. 29, no. 8, pp. 575–584, 1997.
[119] S. Aukstakalnis, *Practical augmented reality*, First. Crawfordsville, Indiana: Addison-Wesley, 2017.
[120] M. J. Piller and M. M. Sebrechts, “Spatial Learning in Transparent Virtual Environments,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 47, no. 20, pp. 2133–2136, Oct. 2003.
[121] F. P. Brooks, “What’s real about virtual reality?,” *IEEE Comput. Graph. Appl.*, vol. 19, no. 6, pp. 16–27, Nov. 1999.
[122] R. P. Darken and J. L. Sibert, “Wayfinding Strategies and Behaviors in Large Virtual Worlds,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1996, pp. 142–149.
[123] J. Psotka, “Immersive training systems: Virtual reality and education and training,” *Instr. Sci.*, vol. 23, no. 5–6, pp. 405–431, Nov. 1995.
[124] S. M. Cormier and J. D. Hagman, *Transfer of Learning: Contemporary Research and Applications*. Academic Press, 2014.
[125] W. Barfield and T. A. Furness, *Virtual Environments and Advanced Interface Design*. Oxford University Press, USA, 1995.
[126] M. K. Young, G. B. Gaylor, S. M. Andrus, and B. Bodenheimer, “A Comparison of Two Cost-differentiated Virtual Reality Systems for Perception and Action

Tasks," in *Proceedings of the ACM Symposium on Applied Perception*, New York, NY, USA, 2014, pp. 83–90.
[127] B. Li, R. Zhang, A. Nordman, and S. A. Kuhl, "The Effects of Minification and Display Field of View on Distance Judgments in Real and HMD-based Environments," in *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, New York, NY, USA, 2015, pp. 55–58.
[128] D. R. Begault and L. J. Trejo, "3-D Sound for Virtual Reality and Multimedia," Aug. 2000.
[129] M. H. Yun, D. Cannon, A. Freivalds, and G. Thomas, "An instrumented glove for grasp specification in virtual-reality-based point-and-direct telerobotics," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 27, no. 5, pp. 835–846, Oct. 1997.
[130] R. J. Pleban, M. D. Matthews, M. S. Salter, and D. E. Eakin, "Training and Assessing Complex Decision-Making in a Virtual Environment," *Percept. Mot. Skills*, vol. 94, no. 3, pp. 871–882, Jun. 2002.
[131] E. Patrick, D. Cosgrove, A. Slavkovic, J. A. Rode, T. Verratti, and G. Chiselko, "Using a Large Projection Screen As an Alternative to Head-mounted Displays for Virtual Environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2000, pp. 478–485.
[132] R. A. Ruddle, S. J. Payne, and D. M. Jones, "Navigating Large-Scale Virtual Environments: What Differences Occur Between Helmet-Mounted and Desk-Top Displays?," *Presence*, vol. 8, no. 2, pp. 157–168, Apr. 1999.
[133] S. Tachi and K. Yasuda, "Evaluation Experiments of a Teleexistence Manipulation System," *Presence Teleoperators Virtual Environ.*, vol. 3, no. 1, pp. 35–44, Jan. 1994.
[134] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, "Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 4, pp. 626–633, Apr. 2012.
[135] A. Gomes de Sá and G. Zachmann, "Virtual reality as a tool for verification of assembly and maintenance processes," *Comput. Graph.*, vol. 23, no. 3, pp. 389–403, Jun. 1999.
[136] J. Aleotti, S. Caselli, and M. Reggiani, "Evaluation of virtual fixtures for a robot programming by demonstration interface," *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 35, no. 4, pp. 536–545, Jul. 2005.
[137] L.-T. Cheng, R. Kazman, and J. Robinson, "Vibrotactile Feedback in Delicate Virtual Reality Operations," in *Proceedings of the Fourth ACM International Conference on Multimedia*, New York, NY, USA, 1996, pp. 243–251.
[138] A. Lécuyer *et al.*, "The effect of haptic, visual and auditory feedback on an insertion task on a 2-screen workbench," in *Proceedings of the Immersive Projection Technology Symposium*, 2002.
[139] B. Petzold *et al.*, "A study on visual, auditory, and haptic feedback for assembly tasks," *Presence Teleoperators Virtual Environ.*, vol. 13, no. 1, pp. 16–21, 2004.
[140] S. Volkov and J. M. Vance, "Effectiveness of Haptic Sensation for the Evaluation of Virtual Prototypes," *J. Comput. Inf. Sci. Eng.*, vol. 1, no. 2, pp. 123–128, Apr. 2001.
[141] M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz, "Visual memory and motor planning in a natural task," *J. Vis.*, vol. 3, no. 1, pp. 6–6, Jan. 2003.

[142] J. B. Carroll, *Human cognitive abilities: a survey of factor-analytic studies*. Cambridge ; New York: Cambridge University Press, 1993.
[143] T. Vecchi, L. H. Phillips, and C. Cornoldi, "Individual differences in visuo-spatial working memory," in *Imagery, language and visuo-spatial thinking*, M. Denis, R. H. Logie, C. Cornoldi, M. de, and J. Engelkamp, Eds. New York, NY, US: Psychology Press, 2001, pp. 29–58.
[144] M. Peters, P. Chisholm, and B. Laeng, "Spatial Ability, Student Gender, and Academic Performance," *J. Eng. Educ.*, vol. 84, no. 1, pp. 69–73, Jan. 1995.
[145] D. Lubinski, "Spatial ability and STEM: A sleeping giant for talent identification and development," *Personal. Individ. Differ.*, vol. 49, no. 4, pp. 344–351, Sep. 2010.
[146] J. Wai, D. Lubinski, and C. P. Benbow, "Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance," *J. Educ. Psychol.*, vol. 101, no. 4, pp. 817–835, 2009.
[147] M. Jr, B. Lc, F. Sg, and W. Mh, "Can a simple test of visual-motor skill predict the performance of microsurgeons?," *J. R. Coll. Surg. Edinb.*, vol. 39, no. 3, pp. 150–152, Jun. 1994.
[148] R. R. Ranney, M. B. Wilson, and R. B. Bennett, "Evaluation of Applicants to Predoctoral Dental Education Programs: Review of the Literature," *J. Dent. Educ.*, vol. 69, no. 10, pp. 1095–1106, Oct. 2005.
[149] K. R. Wanzel, S. J. Hamstra, M. F. Caminiti, D. J. Anastakis, E. D. Grober, and R. K. Reznick, "Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance," *Surgery*, vol. 134, no. 5, pp. 750–757, Nov. 2003.
[150] J.-M. Luursema, W. B. Verwey, P. A. M. Kommers, R. H. Geelkerken, and H. J. Vos, "Optimizing conditions for computer-assisted anatomical learning," *Interact. Comput.*, vol. 18, no. 5, pp. 1123–1138, Sep. 2006.
[151] F. Meijer and E. L. van den Broek, "Representing 3D virtual objects: Interaction between visuo-spatial ability and type of exploration," *Vision Res.*, vol. 50, no. 6, pp. 630–635, Mar. 2010.
[152] T. Huk, "Who benefits from learning with 3D models? the case of spatial ability," *J. Comput. Assist. Learn.*, vol. 22, no. 6, pp. 392–404, 2006.
[153] D. Waller, "Individual differences in spatial learning from computer-simulated environments," *J. Exp. Psychol. Appl.*, vol. 6, no. 4, pp. 307–321, 2000.
[154] M. Sjölinder, K. Höök, L.-G. Nilsson, and G. Andersson, "Age differences and the acquisition of spatial knowledge in a three-dimensional environment: Evaluating the use of an overview map as a navigation aid," *Int. J. Hum.-Comput. Stud.*, vol. 63, no. 6, pp. 537–564, Dec. 2005.
[155] C. D. Wickens, S. E. Gordon, and Y. Liu, *An Introduction to Human Factors Engineering*. Pearson Prentice Hall, 2004.
[156] "Experimental Design | Simply Psychology." [Online]. Available: http://www.simplypsychology.org/experimental-designs.html. [Accessed: 22-Jan-2017].
[157] "Essilor.com," *Essilor Group*. [Online]. Available: https://www.essilor.com/en/vision-tests/. [Accessed: 19-Jan-2017].
[158] "Normal Color Vision | EnChroma." .
[159] *Leap Motion Controller for Mac or PC*. .

[160] "tobii-pro-glasses-2-product-description.pdf." [Online]. Available: http://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-glasses-2-product-description.pdf/?v=1.0.8. [Accessed: 17-Feb-2017].

[161] G. Wallis and H. Bülthoff, "Learning to recognize objects," *Trends Cogn. Sci.*, vol. 3, no. 1, pp. 22–31, Jan. 1999.

[162] M. J. Tarr and H. H. Bülthoff, "Image-based object recognition in man, monkey and machine," *Cognition*, vol. 67, no. 1–2, pp. 1–20, Jul. 1998.

[163] "Unity - Game Engine," *Unity*. [Online]. Available: https://unity3d.com. [Accessed: 21-Feb-2017].

[164] D. Mizell, S. Jones, P. Jackson, and D. Pickett, "Is VR Better than a Workstation? A Report on Human Performance Experiments in Progress," in *Virtual Environments '95*, D. M. Göbel, Ed. Springer Vienna, 1995, pp. 1–7.

[165] "Motion-controlled Servos with Leap Motion & Raspberry Pi," *PubNub*. [Online]. Available: https://www.pubnub.com/blog/2015-08-19-motion-controlled-servos-with-leap-motion-raspberry-pi/. [Accessed: 17-Feb-2017].

[166] Tobii, "Eye Tracking Metrics," 22-Mar-2017. [Online]. Available: https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/interpret/eye-tracking-metrics/. [Accessed: 27-Aug-2017].