

Predicting User Performance for the Evaluation of User Interfaces in Immersive Augmented Reality

Majid Pourmemar

**A Thesis
in
The Department
of
Computer Science and Software Engineering**

**Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Computer Science) at
Concordia University
Montréal, Québec, Canada**

January 2024

© Majid Pourmemar, 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Majid Pourmemar**

Entitled: **Predicting User Performance for the Evaluation of User Interfaces in
Immersive Augmented Reality**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to
originality and quality. Signed by the Final Examining Committee:

Dr. Mustafa K. Mehmet Ali Chair

Dr. Paul Hungler External Examiner

Dr. Aaron Johnson External to Program

Dr. Leila Kosseim Examiner

Dr. Marta Kersten Examiner

Dr. Charalambos Poullis Supervisor

Approved by

Dr. Joey Paquet, Chair
Department of Computer Science and Software Engineering

2024

Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Predicting User Performance for the Evaluation of User Interfaces in Immersive Augmented Reality

Majid Pourmemar, Ph.D.

Concordia University, 2024

Augmented Reality (AR) is defined as the enhancement of the user's view with interactive, computer-generated content, and arises in several forms, including Mobile AR, Spatial AR, and notably, Immersive AR. Among these, immersive AR has piqued considerable attention from both academia and industry. It enables users to immerse themselves in virtual content registered to real-world objects through head-mounted displays (HMDs) and optics technology. Providing robust interaction with virtual objects has necessitated immersive headset designers and producers to develop a broad array of interaction techniques and input modalities, enabling users to experience both a natural interaction with virtual objects and a profound sense of presence. Well-known companies in this domain, such as Microsoft, Meta, Apple, and Magic Leap, have pioneered various interactions and input modalities such as hand gestures, head pointing, eye tracking, and voice commands. The current state of the art indicates that specific input modalities, like hand gestures and head pointing, can increase both the physical and mental workload on users, subsequently increasing the error margins in object selection and manipulation within immersive environments. In response, this research adeptly applies statistical and contemporary machine learning techniques to establish guidelines for immersive AR environment designers and developers, aiming to mitigate the effects of physical and mental workload on users, particularly during hierarchical menu selection tasks, a commonplace activity in various computer applications, including immersive AR. This thesis, embodied by four research papers, extends guidelines and recommendations for designers and developers of immersive AR headsets, endeavoring to alleviate workload and error rates induced by

natural interactions such as hand gestures and head pointing while performing hierarchical menu selection. The initial research focused on identifying the most efficient combination of hierarchical menu types, like radial and drop-down menus, and input modalities, such as hand gestures and head pointing in terms of workload and performance measures. The subsequent study deployed a machine learning approach, leveraging semantic encoders and the standard cognitive performance test, WAIS-IV, to predict human performance in an immersive AR environment during hierarchical menu selection tasks. The third research introduces an analysis and index for error rate in hierarchical menu selection, utilizing both subjective and objective data derived from the users. The final research is an analysis of head pointing and hand gesture path data during menu selection tasks, elucidating an index and its relationship with both subjective and objective methods for calculating workload and error rate in immersive AR.

Acknowledgments

Embarking on this journey, I am enveloped with immense gratitude towards those who have been pivotal in the completion of this thesis. First and foremost, I extend my deepest appreciation to my supervisor at Concordia University, Dr. Charalambos Poullis, whose unwavering support and insightful guidance have been instrumental in navigating through the multifaceted dimensions of this research. Additionally, I would like to thank the examining committee: Dr. Paul Hungler from Queen's University, Dr. Aaron Johnson, Dr. Leila Kosseim, and Dr. Marta Kersten, all from Concordia University. Their insightful feedback and valuable suggestions have significantly contributed to the refinement of this thesis.

To my parents, whose constant love, prayers, and sacrifices have been my enduring strength. Your belief in me, even in moments when I wavered, has been the unyielding pillar that fostered resilience and determination throughout this endeavor. I am profoundly thankful for the roots and wings you have provided me; this achievement is as much yours as it is mine.

Lastly, I express my gratitude to friends, colleagues, and everyone who has contributed, either directly or indirectly, towards the fruition of this work. Your role, no matter how subtle, has been essential and highly appreciated.

Thank you.

Contents

List of Figures	x
List of Tables	xii
List of Acronyms	xiii
1 Introduction	1
1.1 Virtual World and Mixed Realities	1
1.2 Augmented Reality, Definition and its Applications	2
1.3 Hardware and Software Supporting AR	4
1.4 Interactions in immersive AR	6
1.5 Motivation of this Thesis	7
1.6 Thesis Structure	8
1.7 Contributions of the Thesis	9
2 Background Theory	11
2.1 Input Modalities in Immersive AR	11
2.1.1 Head Pointing (head-gaze)	11
2.1.2 Hand Gesture	12
2.1.3 Voice Recognition	14
2.1.4 Eye Tracking (eye-gaze)	14
2.1.5 Hand Touch	15
2.2 Menus as Graphical User Interfaces in Immersive AR	16

2.3	Performance Factors for User Interfaces in Immersive AR	17
2.4	Measuring workload in Immersive AR	18
2.5	Analyzing User Interface Efficiency: GOMS and Keystroke-Level Models	19
2.6	Background Theory in Performance Prediction	20
2.7	Objective and Challenges	21
3	Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality	22
3.1	Introduction	23
3.2	Related Work	26
3.2.1	Hierarchical Menus in AR	26
3.2.2	Selection Techniques	27
3.3	Application	28
3.4	Methodology	30
3.4.1	Experimental Setting	30
3.4.2	Pilot Test	32
3.4.3	Participant Demographics	33
3.4.4	Analysis	33
3.5	Evaluation	33
3.5.1	Pre-test Questionnaire	33
3.5.2	Quantitative Results	36
3.5.3	Post-test Questionnaire	37
3.6	Discussion	39
3.7	Conclusion and Future Work	41
4	Predicting Human Performance in Vertical Hierarchical Menu Selection in Immersive AR Using Hand-gesture and Head-gaze	42
4.1	Introduction	43
4.1.1	Human Performance on Hierarchical Menu Selection	45
4.1.2	Contributions	46
4.2	Literature Review	47

4.2.1	Traditional Methods	47
4.2.2	Machine Learning-based Methods	48
4.3	Methodology	49
4.3.1	Application	49
4.3.2	Measures	50
4.3.3	User Study Procedure	52
4.3.4	Participant Demographics	53
4.4	Model Design and Learning	53
4.4.1	Data Prepossessing	54
4.4.2	Neural Network Model	54
4.4.3	Training	54
4.5	Discussion and Analysis of the Results	55
4.5.1	Ablation	58
4.6	Conclusion and Future Work	59
5	Analysis of Error Rate in Hierarchical Menu Selection in Immersive Augmented Reality	61
5.1	Introduction	62
5.1.1	User-related factors	63
5.1.2	GUI-related features	63
5.2	Related Work	64
5.2.1	Pointing time and task completion time	64
5.2.2	Physical and mental workload	64
5.3	Methodology	65
5.3.1	Metrics	65
5.3.2	WAIS IV	67
5.3.3	The Proposed Index	67
5.4	User Study	70
5.4.1	Immersive AR Application	71

5.5	Discussion and Analysis of the Results	71
5.6	Conclusion and future works	72
6	Analysis of Hand Movement and Head Orientation in Hierarchical Menu Selection in Immersive AR	74
6.1	Introduction	75
6.2	Related Work	78
6.2.1	Hand gesture and head pointing in immersive AR	78
6.2.2	Coordination of hand gesture and head pointing	79
6.3	Application	80
6.4	Methodology	81
6.4.1	Metrics	81
6.4.2	Analysis	84
6.5	Evaluation	85
6.6	Discussion	87
6.7	Conclusion and Future Work	88
7	Conclusion and Future Work	89
7.1	Limitations	91
7.2	Future Perspective	91
	Appendix A	93
A.1	WAIS IV Test	93
A.2	Pre-test Questionnaire	102
A.3	Post-test Questionnaire	105
	Bibliography	109

List of Figures

Figure 1.1	Mixed Reality spectrum [99]	2
Figure 1.2	Several display types for AR [42]	2
Figure 1.3	The reconstructed scene in a immersive AR application [13]	3
Figure 1.4	Mobile AR application of IKEA furnishing [3]	5
Figure 1.5	Immersive AR [8]	6
Figure 2.1	Air tap gesture for MS HoloLens [8]	13
Figure 2.2	Bloom gesture for MS HoloLens [8]	13
Figure 2.3	The hierarchical menu structures, radial and drop-down [98]	17
Figure 2.4	Marking Menus [65]	17
Figure 2.5	Human arm model used to calculate Consumed Endurance [59]	19
Figure 2.6	Consumed Endurance (CE) setup	19
Figure 3.1	Closeups of menus and the paired input modality in user study 1	25
Figure 3.2	Closeups of a sample task in user study 1	31
Figure 3.3	The experimental setting, user study 1	32
Figure 3.4	Visualization of Pre-test questionnaire responses in user study 1	34
Figure 3.5	The mean pointing time per (menu, input modality) in user study 1	37
Figure 3.6	Mean responses of the participants to physical demand questionnaire	38
Figure 3.7	Comparing physical demand and Frustration with average pointing time	39
Figure 3.8	Mean ratings of (menu, input modality) pairs	39
Figure 4.1	An overview of tasks in user study 2	44
Figure 4.2	Data acquisition procedure in user study 2	49

Figure 4.3	The experimental setting in user study 2	52
Figure 4.4	An overview of menu selection tasks in user study 2	53
Figure 4.5	The sentence encoder applied to capture the semantics of tasks	55
Figure 4.6	An overview of data preprocessing and machine learning pipeline	56
Figure 4.7	An overview of predicted CE/PT versus the actual CE/PT	57
Figure 5.1	A sample hierarchical menu task presented to the users	63
Figure 5.2	The values of normalized pointing times	66
Figure 5.3	The WAIS-IV test (Symbol Search and Symbol Coding)	68
Figure 5.4	Example of a three-level hierarchical menu structure.	69
Figure 5.5	Participants' Demographics	70
Figure 5.6	The experimental setting	70
Figure 5.7	Evaluation of the index of error	72
Figure 6.1	A user executing menu selection tasks during the user study	77
Figure 6.2	User executing tasks as the MS Kinect captures hand movements	82
Figure 6.3	Capturing user's head and hand displacement	84

List of Tables

Table 3.1	Participants' responses for 'Rate your experience using AR	35
Table 3.2	Pairwise comparisons of (menu, input modality)	35
Table 3.3	The two types of errors, selection and recognition	36
Table 4.1	Results of Average R^2 and MSE for PT and CE	58
Table 4.2	Results of Ablation study to determine the importance of each feature	58
Table 4.3	Pearson's correlation analysis on pointing time and Consumed Endurance . .	59
Table 5.1	Pearson's correlation analysis on Learning Rate and Error Rate	72
Table 6.1	The proposed index for combinations of (GUI, input modality)	86
Table 6.2	The proposed index for different attempts in user study 2	87
Table 6.3	Pearson's correlation analysis on the proposed index	87

List of Acronyms

AE	AR Experience
AG	Age
AHM	Accumulated Hand Movements
AHO	Accumulated Head Orientations
AI	Artificial Intelligence
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
API	Application Programming Interface
AR	Augmented Reality
AV	Augmented Virtuality
BERT	Bidirectional Encoder Representations from Transformers
CAVE	Cave Automatic Virtual Environment
CE	Consumed Endurance
DOF	Degrees of Freedom
ER	Error Rate
GOMS	Goals, Operators, Methods, and Selection
GPT	Generative Pre-Trained Transformers
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HHR	Hand-Head usage Ratio
HMD	Head Mounted Display

IQ	Intelligence Quotient
KLM	Keystroke-Level Model
LiDAR	Light Detection and Ranging
LSTM	Long Short-Term Memory
MF	Menu Factor
MR	Mixed Reality
MRTK	Mixed Reality Toolkit
MSE	Mean Squared Error
PCC	Pearson's Correlation Coefficient
PIQ	Performance IQ
PSI	Processing Speed Index
PT	Pointing Time
RE	Real Environment
SAR	Spatial Augmented Reality
SC	Symbol Coding
SD	Standard Deviation
SPD	Subjective Physical Demand
SS	Symbol Search
TLX	Task Load Index
UI	User Interface
USE	Universal Sentence Encoder
UX	User Experience
VE	Virtual Environment
VIQ	Verbal IQ
VR	Virtual Reality
WAIS	Wechsler Adult Intelligence Scale
XR	Extended Reality

Chapter 1

Introduction

1.1 Virtual World and Mixed Realities

Virtual Reality (VR) is characterized as a computer-generated, three-dimensional environment, replete with scenes and objects that not only appear real but also engender a deeply immersive and interactive user experience. Various forms of VR exist, including head-mounted VR (using immersive headsets such as Oculus Quest [7]) and VR CAVE systems [75] (cube-shaped rooms having displays on the walls). The environment is interactive and the user has a sense of presence meaning that the user feels to be in a virtual location [103]. In a VR experience, users are solely exposed to the virtual world, devoid of any real-world indicators. A hybrid experience, known as Mixed Reality (MR), combines both virtual and real-world objects to forge a novel interactive landscape. Figure 1.1 delineates several types of MR. While MR is a mixture of the real and virtual world, the whole experience including AR, and VR, MR is called extended Reality (XR).

As it is indicated in Figure 1.1 Augmented Reality (AR) and Augmented Virtuality (AV) weave virtual and real-world elements into a unified user interface. Their distinction lies in the locus of user interaction: AR involves interaction in the real world, whereas AV entails interaction within a virtual space [99]. The extent of MR is ascertained by the quantity and impact of virtual content on the user's perspective [42].

The focus of this thesis pertains to the exploration of interactions and user interfaces within the realm of AR, which will be elucidated in the subsequent sections.

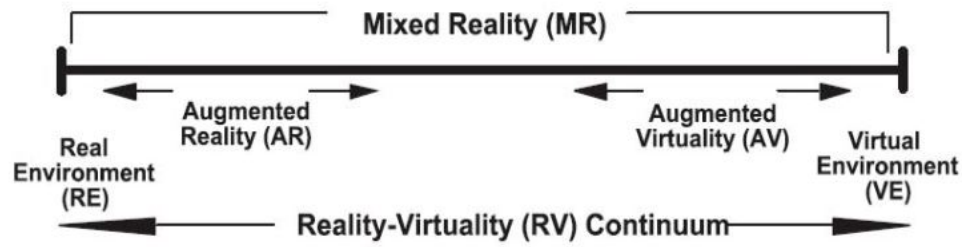


Figure 1.1: Mixed Reality spectrum [99]

1.2 Augmented Reality, Definition and its Applications

Among different types of MR, Augmented Reality (AR) has garnered significant popularity and interest from both industry and academia. AR enhances users' perception of the real environment by overlaying computer-generated content through mobile devices, subsequently sparking extensive research and development in user interaction and interface designs. AR is defined as a combination of computer vision/graphics based software and hardware that has the following characteristics.

- (1) Combines real and virtual objects in real environment (augmentation),
- (2) Is interactive in real-time (interaction)
- (3) Registers and aligns real and virtual objects with each other (spatial mapping)
- (4) Runs and/or displays the augmented on a mobile device (visualization)



Figure 1.2: Several display types for AR [42]

In order to augment the view of the mobile user with computer-generated content, the system has to have knowledge of the geometry of the environment. The surfaces, the walls, and locations of the objects must be extracted by the device. To this end, most AR applications have a pre-processing scanning procedure to extract the geometry of the environment and load a 3D reconstruction model to the device's memory. After loading the 3D map of the environment in the device's memory (Figure 1.3) the AR application will be able to place the virtual objects on the real surfaces. Computer vision and machine learning algorithms are technically employed to accomplish this goal.

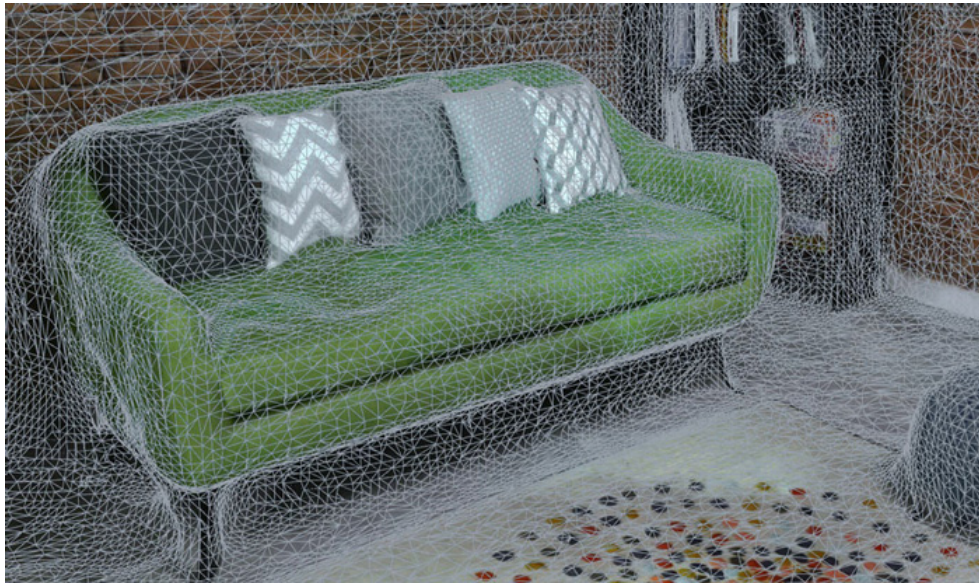


Figure 1.3: The reconstructed scene in a immersive AR application [13]

Regarding applications of AR, the technology currently has many applications in various fields in both industry and academia. Some of these applications will be discussed in the following.

- **Computer Games:** Games like "Pokémon GO" [9] where users can catch virtual creatures in the real world.
- **Education and Training:** Application such as HoloAnatomy [2] that is used to teach the anatomy of the human body to medical students. A full review of applications AR in education and training can be found at [76].
- **Retail and Shopping [79]:** As an example Shopify company [12] has integrated AR capabilities into its platform, allowing merchants to offer 3D product models for customers to view

in their real-world environment using mobile devices. This AR feature enhances the online shopping experience by providing a realistic product perspective, potentially increasing engagement and reducing return rates. The success of this feature largely depends on the quality of the 3D models and is especially beneficial for products where size and spatial placement are crucial.

- **Healthcare:** There are various applications of AR in surgeries and medicine. As an example [70] and [23] suggest a system for real-time AR video enhancement on mobile devices specifically for image-guided neurosurgery.
- **Navigation and Travel:** The most famous application of this kind is Google Maps which uses AR views to assist people in finding their way [35].
- **Realstate and Interior Design:** For example Magicplan [6] and Roomplan [11] allow users to create detailed floor plans using AR, simply by scanning the room with a mobile device, making property assessments and listings more comprehensive.

1.3 Hardware and Software Supporting AR

There are several devices such as smartphones, tablets, and glasses which have been designed to support AR. The AR application receives its input from the device's embedded sensors such as camera, depth camera, LiDAR sensors, Gyroscope, accelerometer, GPS, and microphone, and processes it in real-time in order to overlay the virtual content on top of the current view of the user, on the screen of the mobile device.

Since AR devices have to be mobile, they must be small and light such that the user is able to freely move and see the environment by looking through the device's display. Below we provide a brief overview of existing AR devices.

There are two kinds of devices which support AR namely mobile AR and immersive AR.

Mobile AR: Smartphones and tablets recently have been equipped with powerful processors and different types of built-in sensors such as cameras and gyroscopes. This has increased their ability to process the environment in real-time and therefore support AR. Although smartphones are

highly portable and more acceptable to people, there are still shortcomings such as small field-of-view and low computing capability. In addition to this, using mobile AR forces the user to hold the device in front of them and move it from place to place [42]. The second type of AR which is immersive AR has emerged to address many of these limitations. Figure 1.4 shows an example of a mobile AR application in the domain of interior design.



Figure 1.4: Mobile AR application of IKEA furnishing [3]

Immersive AR: These are wearable devices or headsets such as HoloLens2 [8] or Apple Vision Pro [1] which provide an increased sense of presence with higher Degree of Freedom (DOF) and more interactions available for the user [22]. Smart glasses play a more effective role in ubiquitous computing when compared with mobile AR devices since they provide more interactions. In contrast to mobile AR where the users have to hold the device with their hands and have to move around the room, in immersive AR the user's hands are free which allows for more types of interactions to be supported. The interaction techniques and user interfaces in this field of AR is the main focus of this thesis.

On the software side, there are several libraries available for developing AR applications. Since all AR applications deal with 3D content and spatial perception, 3D game engines such as Unity3D [15] or Unreal [16] engine are the best choices for developing AR applications. Some of the



Figure 1.5: Immersive AR [8]

main frameworks for developing AR applications are MRTK (by Microsoft), ARCore (by Google), ARKit, and RealityKit (by Apple).

Spatial AR: Spatial AR (SAR) [26] utilizes projectors to overlay digital graphics directly onto physical surfaces, transforming objects into interactive displays without the need for wearable devices. Unlike other kinds of AR, which rely on screens or headsets, SAR projects information onto real-world objects, allowing for dynamic visual interactions. This technology has applications ranging from entertainment and art to education, industrial design, and medical visualization. Challenges include controlled lighting conditions and accurate calibration, but SAR offers a unique, immersive augmented experience by converting the physical environment into a digital canvas.

1.4 Interactions in immersive AR

Immersive AR allows the users to interact with 3D virtual content that has been positioned in the real world surfaces. There are three types of interactions in AR namely selection, manipulation, and navigation [29]. The user can use both natural input modalities such as hand gestures, voice recognition, body movements etc, or manual inputs such as clickers and controllers.

Selection: Selection in AR refers to identifying and picking a virtual 3D model for manipulation. The selection action itself has two parts. The first part is pointing i.e. changing the location

of cursor from the current location to the desired location. It can be done by head pointing, gesture, gaze, or other methods. The second part is to pick up the desired object and receive the feedback from the object. Picking by clicker hand gesture or dwell-time are the frequent ways of performing this section.

Manipulation: After selecting an object the next step is to manipulate that 3D model with the existing input modalities and change the features of that object. Controllers and hand gestures are most commonly used.

Navigation: The navigation is essential for virtual environments where the view of the user is occluded by the virtual objects. To navigate in the virtual world the user can use eye gaze, hand gestures controllers, or body movements.

1.5 Motivation of this Thesis

Applications in AR are by default interactive. Different types of interactions have already been proposed ranging from simple passive actions such as the user looking through the HMD to see virtual objects positioned on the real surface [111], to more complex-active actions where the user can create, edit, and manipulate the virtual objects [32], [107].

Until recently, AR applications have typically included a small set of available actions. The user was informed about these actions either before using the application, or through an introduction/tutorial at the beginning. In cases where the number of available actions/options is small then this approach is adequate, however as this number increases the complexity of pre-usage instructions and tutorials also increases. Consider for example a 3D authoring software such as Autodesk Maya, 3DS Max, etc, where hundreds of actions are available at any given point in time. Designing such software in AR means that the user has to be aware of and remember all these actions, which is not feasible in most cases. An obvious solution to the aforementioned problem is to visually present these actions to the user within the AR application in a similar way to the hierarchical menus in desktop applications. However, this raises the question of how to visualize these actions within the context of AR. Some systems from the literature [60] replicate what has already worked for desktop applications to AR e.g. drop-down menus. However, when dealing with an application with a large

number of actions this approach will result in the user's view being cluttered with menus; in fact, to date, there are no design guidelines as to how to visualize actions and options in AR. In this thesis, we address the aforementioned problems of visualizing actions within the context of AR applications and will present our research on predicting the performance of user interfaces in immersive AR applications.

Interactions in immersive AR are done using natural input modalities such as hand gestures, head pointing, eye tracking, touch, and other types of input modalities. Research in the area have shown that although natural interactions in immersive AR are preferred by many users, there are some problems that make it difficult for many users. Having a high physical/mental workload, high error rate in selections, and coordination between these input modalities and the 3D user interface (UI) is an important factor in dealing with immersive AR applications. As menu selection is one of the most important interactions in any computer application we focused on it and evaluated hierarchical menu selections in immersive AR through various metrics and methods that will be discussed in the following chapters.

In this thesis, firstly we focused on different combinations of input modalities and graphical user interfaces in immersive AR. Next, we focused on using machine learning and mathematical methods for predicting human performance while doing hierarchical menu selection tasks using both user-related and GUI-related aspects. We applied performance metrics such as physical workload, task completion time, and error rate in our evaluations.

1.6 Thesis Structure

This thesis is organized into seven chapters, encompassing four published research papers focused on user interfaces and interaction techniques in immersive AR. In Chapter 2, a review of current background theory in the domain is provided. The paper, presented in Chapter 3, as it is published in [92], features a user study evaluating the efficacy of various GUI and input modality pairings, both in terms of task completion time and user subjective ratings. Chapter 4, published as [91], introduces a predictive model, developed using machine learning techniques, which accounts for both user-related and GUI-related factors to predict human performance during menu

selection in immersive AR. Further, Chapter 5, published as [93], delves into the analysis of error rates and incorrect menu selections in immersive AR. Meanwhile, Chapter 6 explores the dynamics of hand gestures and head orientations during menu selection in immersive AR, emphasizing their correlations with workload and error rates. The final chapter offers a conclusion and outlines potential future research directions in the field.

1.7 Contributions of the Thesis

The contributions of this thesis are as follows:

- **Evaluation of GUI, input modality pairs in immersive AR:** We provided an analysis for different pairs of GUI and input modalities in immersive AR. We introduced and analyzed the efficiency and performance of Marking Menus [66] in immersive AR using input modalities such as hand gestures and head pointing. Marking Menus are widely used in 3D authoring applications and since immersive AR is in 3D it is vital to know about its performance and efficiency.
- **Predictive models in immersive AR:** To the best of our knowledge, this thesis is the first work on predicting human performance of vertical hierarchical menu selection in immersive AR incorporating both user-related and GUI-related factors.
- **Semantics of menus:** This work is the first work that considers the semantic relationship of the targets at different levels in the hierarchy. Previously, Li et al. [71], focused only on the semantics of the words in vertical menus. We demonstrate that when it comes to a hierarchical menu selection, the semantic relationship between the targets in each menu level is an important factor to consider.
- **User-related factors and WAIS Test** Since the performance of human actions in immersive AR is widely dependent on the cognitive capabilities of users, we applied the results of a cognitive test named WAIS-IV [105] from our users to improve the final model of human performance and workload in immersive AR. The test is a standard one that measures the performance of the human brain during cognitive tasks.

- **Menu selection error rate assessment:** We provided a metric for assessing the error rate in menu selection in immersive AR using both user-related and GUI-related factors. The index can be used by designers and developers of immersive experiences to optimize their design.
- **Relation of hand and head usage in immersive AR Menu selection:** Since interaction in immersive AR is widely dependent on hand and head movements in the last work we provided an analysis of head and hand usage in immersive AR while performing menu selection tasks with respect to both subjective and objective methods evaluating workload.

Chapter 2

Background Theory

In this chapter, we provide an overview and a background theory on interactions and input modalities used in immersive settings as well as graphical user interfaces. The performance factors in immersive settings are also discussed.

2.1 Input Modalities in Immersive AR

Input modalities are ways that the user can interact with the immersive HMDs and therefore virtual environments. Most of the HMDs support manual input modalities such as clickers and controllers. Other input methods for natural interactions have been proposed and will be reviewed in this chapter including head pointing [8], hand gesture [44], eye tracking [68], [21], voice input [28].

2.1.1 Head Pointing (head-gaze)

While using head-mounted displays the orientation of the user's head is recorded in real time using built-in gyroscope and accelerometer sensors. A ray-cast from the user's head with the specified head orientation is used to select virtual/real objects. The point of the intersection between the ray-cast and the real/virtual object is called as cursor. As Kyto et al. [68] indicated, head pointing provides more targeting accuracy when compared with other methods such as gaze or gesture; but it is slower than others. Blattgerste et al. [27] compared the head gaze method with the eye-gaze method for dwell-time based UI interaction using the HoloLens, with respect to speed, task load,

and user preference. The HMD has been equipped with an eye tracker device to extract eye gaze while interacting. The results show that eye gaze outperformed head gaze in terms of speed, task load, and error ratio.

2.1.2 Hand Gesture

Due to the increased Degrees of Freedom, in immersive AR the mid-air gestures such as air-tap and bloom were proposed. Rather than relying on desktop methods such as controllers and clickers, direct manipulation of virtual objects is performed to allow the users to interact with the virtual objects in 3D space. Hand gesture tracking for augmented reality environments was proposed by Dorfmüller et al. [44] in order to interact with virtual content and perform actions such as rotation, translation and scale in a 3D environment. The user had to wear a certain kind of marked glove for better optical hand gesture and finger recognition and tracking. Much later with improvements in computer vision techniques the leap motion hand tracker [4], a USB peripheral device, was designed to be placed in front of the user [64] or on top of the immersive HMD [88] and allow the user to perform a wide range of hand gestures for interacting with virtual contents. MS Kinect is another device that is widely used to detect and recognize hand gestures. The technology behind these devices is a combination of regular cameras and infrared depth cameras used to find the depth in recorded scenes. The MS HoloLens [8] is another AR device that uses Kinect technology to detect hands and recognize hand gestures. There are three main hand gestures defined for a HoloLens user: Bloom, Air-tap, and Tap-and-hold however developers can define new hand gestures for the MS HoloLens. Figure 2.1 and Figure 2.2 show the two most common hand gestures of the MS HoloLens.

Although the hand gesture for immersive AR is an appropriate way to increase the DOF, holding the hands up for a long time can cause user fatigue known as Gorilla Arm. Reducing arm fatigue is an open research topic in the area of mid-air hand gestures. Hincapié et al. [59] have proposed a method to characterize arm fatigue while using hand gesture in AR named Consumed Endurance (CE). Using CE researchers do not need to ask people about their subjective opinions on feeling and fatigue while performing mid-air gesture in user studies. There are various ways of addressing the Gorilla Arm problem. One way is to optimize the graphical user interface in AR in such a way

that it decreases the hand use duration in the application [78]. Using other complementary input modalities such as voice recognition or gaze [55] is one such way to reduce arm fatigue. Xiao et al. [107] proposed a device-free touch input method for immersive AR displays which reduces arm fatigue by adding touch input to head-mounted AR. More information on their method is in section 2.1.5.

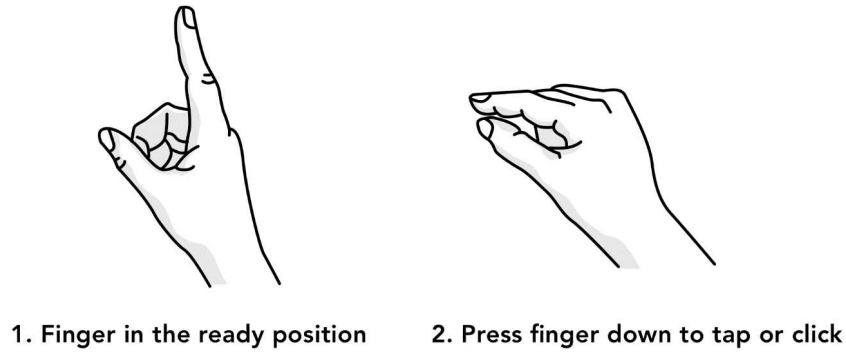


Figure 2.1: Air tap gesture for MS HoloLens [8]

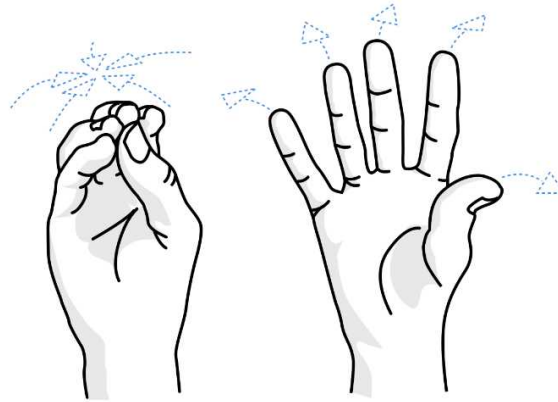


Figure 2.2: Bloom gesture for MS HoloLens [8]

Ubii (Ubiquitous interface and interaction) is another novel interaction method for immersive environments based on hand gesture input modality. It allows the users to interact with smart devices in a room remotely using hand gestures. Beforehand, in order to work with smart devices like computers, laptops physical touch was needed. The physical touch has prevented the users to interact with out of reach devices or interact with multiple devices at the same time. Ubii has

been implemented on an immersive AR HMD and using hand gestures and a user friendly graphical user interface allow the users to work with smart devices from long distance in a room and create a network of the smart devices to interact with all at the same time.

2.1.3 Voice Recognition

With improvements in signal processing and natural language processing algorithms, more and more devices support voice recognition and speech processing as an input modality. Siri (Apple), Cortana (Microsoft), and Alexa (Amazon) are examples of speech-processing platforms. AR head-mounted displays such as Microsoft HoloLens also support voice recognition as an interaction method in addition to head pointing and gesture. The results of the user study conducted by Pick et al. [89] shows that voice recognition can increase the task execution speed in menu selection in immersive AR. But still, there are problems with the voice recognition input modality in immersive HMDs. Firstly the results of the user study in [45] show that most users do not like to speak with a device in public due to privacy issues (i.e. speaking loud in a library). Secondly, voice recognition is not always an accurate way to communicate with a device and it highly depends on environment conditions because of noises in the environment. The third problem is that there are some tasks while manipulating 3D content that are not feasible with voice commands such as drag and drop a virtual object to a certain location, which can easily be done by hand gesture rather than voice command.

2.1.4 Eye Tracking (eye-gaze)

Eye tracking methods require using certain wearable devices to be used with head-mounted displays. Tobii [14] and PupilLabs [10] are two eye tracker devices which are compatible with immersive headsets.

Grinshpoon et al [55], proposed a hands-free user interface for the MS HoloLens headset. The new user interface was used by surgeons who wanted to annotate the patient with virtual information during the surgery. In this scenario the surgeon is unable to use her hands to interact with the device therefore eye gaze and voice commands were used. Similar head based interactions and interfaces for AR have also been proposed in [85], [96]. In [68] the authors performed a comparison

between the pointing time and the accuracy of different methods of different interaction methods for immersive AR. In addition to the available headset sensors, they equipped the headset with an eye tracker. The reported results show that eye gaze alone is faster than head pointing, but head pointing allows a greater targeting accuracy. The paper also showed that performing gestures to interact with objects needs considerably more effort when compared to other methods. This was also confirmed later by Hincapi-Ramos et al in [59] where they showed that using hand gestures to interact with immersive environments for a long time can lead to arm fatigue; a problem known as 'Gorilla Arm'.

A novel wearable input method for augmented reality called ubiGaze which uses gaze input was proposed by Bace et al. [21]. The method enables the user to augment any real world object with invisible messages. Users can read and write invisible messages on real world objects using eye tracking and hand gestures. The hand gesture is enabled by a smartwatch on the user's hand. To detect the user's gaze the authors have used the Pupil Labs eye tracker. As it is mentioned in [21], an important challenge in dealing with all eye tracker systems is the calibration and registration of the device. Problems in calibration lead to errors and inaccurate gaze recognition by the device.

2.1.5 Hand Touch

The current generation of augmented reality headsets such as MS HoloLens 2 or MagicLeap support interactions such as mid-air gestures, gaze, and voice commands. Although these methods have their advantages such as increasing the DOF, they are not precise or rapid [33]. Moreover, long-time usage of methods such as gaze or hand gesture can lead to fatigue problems such as the 'Gorilla Arm' effect [59] as discussed in section 2.2.2. To resolve the problems some AR/VR systems are equipped with hand/finger-based controllers and trackers. However, adding an extra device to an AR system may impede the free hand manipulations of the user. Touch input can be applied in VR/AR environment to manipulate the virtual content more accurately, faster, and more naturally.

To resolve the aforementioned problems, Xiao et al. [107] have added touch input to the head-mounted AR named MRTouch. They used the MS HoloLens 1 as the immersive AR device and real-world flat surfaces such as table/wall as the touchable surface.

The input touch does not need any other peripheral devices or extra touch screens. According

to the paper, the system enables the users to manipulate virtual objects on real-world surfaces using a novel touch input solution which lets the user act with surfaces as real touch screens. A robust finger-tracking application has been added to the existing hand-tracking module of MS HoloLens to track the fingers while moving on a flat surface in front of the user and a user study with 17 participants supports the results of the paper.

As a solution to improve the accuracy of the manipulation while using hand gesture, DualCAD, presented by Millette et al. [81], has mixed the AR environment with a Desktop environment and a touch-based device i.e. a smartphone. Through the system, the user can switch between AR mode and Desktop mode using a touchable device. The desktop mode provides the user with precise and stable input of the mouse and through the AR mode the user can benefit from the depth cues provided by the HMD. Moreover, to improve the AR mode, the authors have provided access to multitouch smartphones.

2.2 Menus as Graphical User Interfaces in Immersive AR

Since in the AR environment, the workspace is a boundless 3D area around the user, multiple varieties of graphical user interfaces coupled with different input modalities have been proposed to manipulate 3D objects and interact with them. Perhaps, the most ubiquitous structure for a graphical user interface is a hierarchical one (Figure 2.3). A hierarchical menu is a multi-level arrangement of options, organized to allow users to find the desired action or information easily than they could in an unstructured presentation. From simple 2D drop-down menus, taken from desktop environments [60] to more interactive 2D menus such as radial/ ring/ tile menus [53], [43], [90] and more specifically 3D menus which can not be represented in 2D areas [32] are all examples of hierarchical menus in immersive AR.

Marking menu [65] is a type of hierarchical radial menu that is proposed by Kurtenbach et al. to reduce the pointing time and improve the performance of menu selection in a desktop environment. Through this kind of menu, the user uses the dwell-time method to select an option. Consequently, radial menus will appear and the user will select sub-menus from the newly appeared radial menu. Autodesk Maya, a 3D authoring application, widely uses marking menus to support the hierarchical

structure of actions and options.

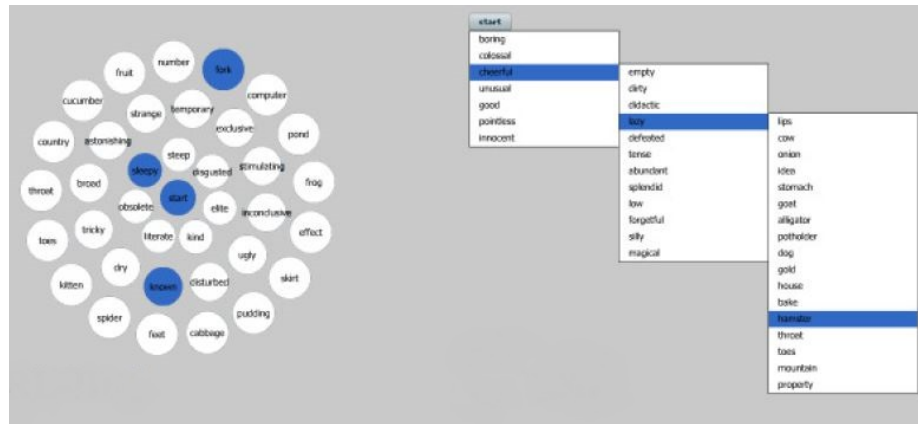


Figure 2.3: The hierarchical menu structures, radial and drop-down [98]

2.3 Performance Factors for User Interfaces in Immersive AR

According to recent studies [49], [71], the performance of a user on menu selection in desktop configurations is dependent on the visual search time and pointing time. Further studies specific to immersive AR applications [59] have shown that the user's performance on menu selection is dependent on additional factors such as workload and error rate. The workload is the amount of physical and mental effort of user while doing menu selection tasks and the error rate is defined as the number of wrong menu selections divided to the number of total tasks. Experiments have shown that this is due to the fact that users are exerted from the physical and mental activity while

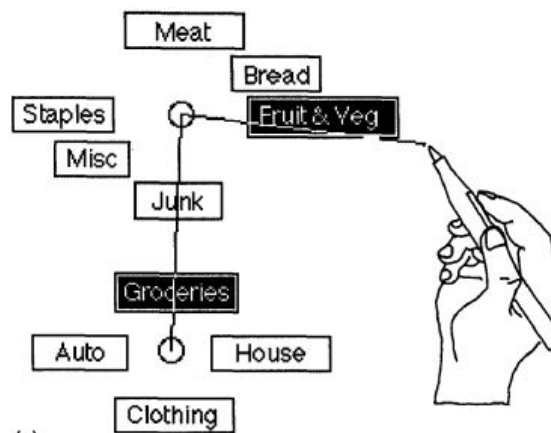


Figure 2.4: Marking Menus [65]

performing the menu selection in immersive leading to arm fatigue known as 'Gorilla arms'. In this work, the user's performance on a target GUI is predicted with respect to the expected pointing time, search time, and workload corresponding to selecting a menu item from a hierarchical drop-down menu through a sequence of menu selection actions.

Below we list the definitions of the important factors we will consider:

- **Visual Search Time:** the time it takes to find an item in the list
- **Pointing Time:** the time it takes to navigate to an item
- **Workload:** the mental/physical effort's of the user to point the cursor to a certain menu option
- **Error Rate:** the number of failed/wrong menu selections divided by the number of all menu selections

2.4 Measuring workload in Immersive AR

Gestural user interfaces and head orientation tracking are widely used in most AR/VR applications. Therefore, they require the users to do physical activity to perform the tasks. As the research shows [59] [92], this might cause fatigue after a while using them. There are various ways to evaluate the user's fatigue in performing gestural tasks. From completely subjective methods such as NASA-TLX [56] to objective methods such as Consumed Endurance. Hincapi'e-Ramos et al. [59] proposed a metric for evaluating arm fatigue in mid-air gestures in immersive AR/VR named Consumed Endurance (CE). CE only requires the tracking system (i.e. Microsoft Kinect) used to interact with the mid-air hand gesture itself, and thus it is a real-time, objective, non-invasive, and non-obtrusive approach to assess arm fatigue. The following Formulas, 1 and 2, indicate the calculation of CE based on the physical features of the hands and the angles of the shoulders while working with the immersive HMD.

$$E(T_{shoulder}) = \frac{1236.5}{\left(\frac{T_{shoulder}}{T_{max}} * 100 - 15\right)^{0.618}} - 72.5 \quad (1)$$

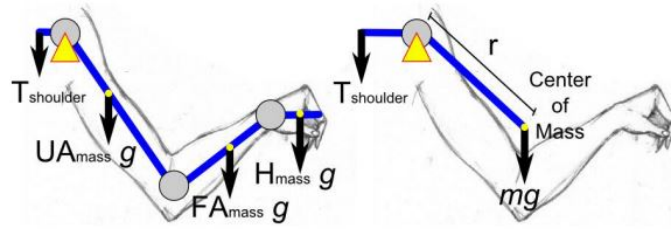


Figure 2.5: Human arm model used to calculate Consumed Endurance [59]

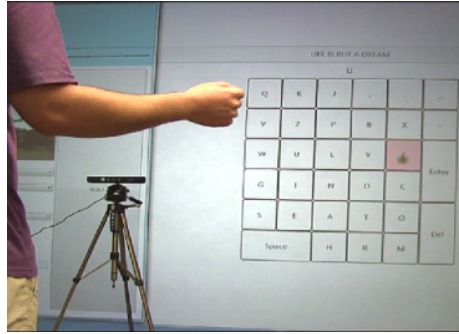


Figure 2.6: Consumed Endurance, the Microsoft Kinect will be in front of the user tracking the user's hand [59]

$$CE(T, TotalTime) = \frac{TotalTime}{E(T)} * 100 \quad (2)$$

2.5 Analyzing User Interface Efficiency: GOMS and Keystroke-Level Models

GOMS (Goals, Operators, Methods, and Selection rules) [37], [36] is a widely used theoretical model in human-computer interaction (HCI) and cognitive psychology. Developed by Card, Moran, and Newell in the early 1980s, GOMS provides a framework for understanding and modeling user interactions with a computer system. The model breaks down tasks into Goals (what the user wants to achieve), Operators (basic actions performed by the user, like clicking a mouse or pressing a key), Methods (procedures to achieve goals), and Selection rules (decisions made to choose among methods). GOMS is valuable for predicting how users will interact with a system and for guiding the design of user interfaces to be more efficient and user-friendly. It allows researchers and designers

to analyze the cognitive processes involved in human-computer interaction and to estimate task completion times, identifying potential areas for improvement.

The Keystroke-Level Model (KLM) [38], a derivative of the GOMS model, is specifically focused on quantifying the time it takes for users to complete routine tasks without errors. Developed by Stuart Card, Thomas P. Moran, and Allen Newell, KLM breaks down user actions into seven operators: keystrokes, pointing with a mouse, homing between keyboard and mouse, drawing, system response time, mental preparation, and speaking. By assigning standard time values to these operators, KLM can be used to predict the time necessary to complete a task. This model is particularly useful in the early stages of design, where it can help in comparing the efficiency of different interface layouts or workflows. By applying KLM, designers and researchers can identify and minimize the time-consuming aspects of a task, thereby improving the overall usability and efficiency of a computer system.

2.6 Background Theory in Performance Prediction

Several works have been offered in the field of Human-Computer Interactions (HCI) to predict human performance in performing UI tasks. From basic models for predicting pointing time in HCI tasks such as Fitts' Law [50] and Hick's Law [58] for predicting visual search time to complicated models using machine learning algorithms [71], [24].

The predictive models have to consider all human factors that are related to the performance of the user interacting with the UI. There are factors such as pointing and visual search time [24], learning effect rate [71], the semantic meaning of menu options [72], the appearance of the menu and workload [59] which all can affect the performance of the menu selection. We assume that the user interface here consists of hundreds of menu options which are gathered together using a hierarchical menu design. When the user aims at an object in 3D, he/she needs to go through the menu design and select the desired option using the available input modalities. There are some popular predictive models on this topic such as Fitt's law (pointing time) and Hick's law (decision-making time). These models have been extracted from information theory and experimental psychology and cover several factors that come into play in the performance of a real user interface such as the pointing time or

the distance between two items in the screen.

2.7 Objective and Challenges

Evaluating user interfaces in HCI requires testing them with real users. To this end, user studies with a certain number of users will be conducted and performance factors such as time, error rate, and workload will be gathered. Especially in AR applications the users need to have Head Mounted Displays to test the application. One of the main objectives of this research is to find a way to evaluate a proposed user interface using Machine Learning. Therefore, the designers and developers can use our model to create more perfect user interfaces in terms of usability and human performance. Machine learning methods are proper ways to find predictive functions.

Chapter 3

Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality

This chapter is the paper published and orally presented at the 17th ACM International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI), Brisbane, Queensland, Australia [92]. The paper was also awarded the Diversity and Inclusion scholarship from the VRCAI committee.

Abstract

Graphical User Interfaces (GUIs) have long been used as a way to inform the user of the large number of available actions and options. GUIs in desktop applications traditionally appear in the form of two-dimensional hierarchical menus due to the limited screen real estate, the spatial restrictions imposed by the hardware e.g. 2D, and the available input modalities e.g. mouse/keyboard point-and-click, touch, dwell-time, etc. In immersive Augmented Reality (AR), there are no such restrictions and the available input modalities are different (i.e. hand gestures, head pointing, or voice recognition), yet the majority of the applications in AR still use the same type of GUIs as with desktop applications.

In this paper we focus on identifying the most efficient combination of (*hierarchical menu type, input modality*) to use in immersive applications using AR headsets. We report on the results of a within-subjects study with 25 participants who performed a number of tasks using four combinations of the most popular hierarchical menu types with the most popular input modalities in AR, namely: (*drop-down menu, hand gestures*), (*drop-down menu, voice*), (*radial menu, hand gestures*), and (*radial menu, head pointing*). Results show that the majority of the participants (60%, 15) achieved a faster performance using the hierarchical radial menu with head pointing control. Furthermore, the participants clearly indicated the radial menu with head pointing control as the most preferred interaction technique due to the limited physical demand as opposed to the current de facto interaction technique in AR i.e. hand gestures, which after prolonged use becomes physically demanding leading to arm fatigue known as 'Gorilla arms'.

3.1 Introduction

Applications in AR are by default interactive. Different types of interactions have already been proposed ranging from simple passive actions such as the user looking through the Head-mounted Display (HMD) to see virtual objects positioned on the real surface [111], to more complex-active actions where the user can create, edit, and manipulate the virtual objects [32], [107].

Until recently, AR applications have typically included a small set of available actions. The user was informed about these actions either before using the application, or through an introduction/tutorial at the beginning. In cases where the number of available actions/options is small then this approach is adequate, however, as this number increases the complexity of pre-usage instructions and tutorials also increases. Consider for example a 3D authoring software such as Autodesk Maya, 3DS Max, etc, where hundreds of actions are available at any given point in time. Designing such software in AR means that the user has to be aware of and remember all these actions, which is not feasible in most cases.

An obvious solution to the aforementioned problem is to visually present these actions to the user within the AR application in a similar way as with the hierarchical menus in desktop applications. However, this raises the question as to how to visualize these actions within the context of

AR. Some systems from the literature [60] replicate what has already worked for desktop applications to AR e.g. drop-down menus. However, when dealing with an application with a large number of actions this approach will result in the user’s view being cluttered with menus; in fact, to date, there are no design guidelines as to how to visualize actions and options in AR. More importantly, though, the constraints imposed by the two-dimensional window of the desktop application (e.g. limited real-estate) and available interaction methods (e.g. pointing devices), are transferred over to the three-dimensional space of the AR application.

In this paper we address the aforementioned problems of visualizing actions within the context of AR applications. We posit that in order to determine the best menu to use in AR one has to consider the fact that it is tightly coupled with the choice of the input modality used. In the context of immersive AR the input modalities such as head pointing, hand gesture or voice recognition are the popular choices. We specifically focus on 3D authoring applications which typically have hundreds of actions and options available at any given time and conduct a within-subjects user-study with 25 participants. Using a custom designed AR application, the participants are asked to perform a number of tasks involving basic operations on a 3D object i.e. rotate, translate, scale, change color. Our application implements four combinations of (menu type, input modality) pairs, namely (drop-down menu, hand gestures), (drop-down menu, voice), (radial menu, hand gestures), and (radial menu, head pointing). The participants are asked to perform the same four operations using all four (menu type, input modality) pairs. A number of subjective and objective measures are recorded and analyzed. The results show that the majority of the participants 60% achieved a faster performance using the radial menu with head pointing control therefore confirming our hypothesis. Furthermore, out of the three interaction methods presented e.g. hand gestures, head pointing, and voice, hand gestures were found to be the least preferred input modality for menu selection due to the increased physical activities required for prolonged use.

Paper organization. The paper is organized as follows: Section 3.2 provides a brief overview of the state-of-the-art in the area. In Section 3.3 we describe the AR application we developed and used to conduct the user-study. The methodology including the pilot test, demographics on the participants, is described in Section 3.4. Section 3.5 describes the evaluation of user study data including objective and subjective measures. A discussion is presented in Section 3.6 followed by

the conclusion and future work in Section 3.7.



(a)



(b)



(c)

Figure 3.1: (a) A user interacting with the application from the observer's perspective. The tasks involve manipulating the animated bird flying in front of the buildings. (b) Closeup of the hierarchical drop-down menu from the user's perspective. (c) Closeup of the radial menu from the user's perspective. The green line is controlled by the user (either with hand gestures or head pointing) and the selection is made after a dwell-time of 1.5s.

3.2 Related Work

A plethora of work has already been proposed in the literature addressing menu types and input modalities in AR, although not so much work has been conducted in assessing menu types in combination with the input modality used. Below we provide a brief overview of the most relevant work with the proposed.

3.2.1 Hierarchical Menus in AR

Since in the AR environment, the workspace is a boundless 3D area around the user, multiple varieties of GUIs coupled with different interaction methods have been proposed to manipulate 3D objects in the view of the user and interact with them. From simple 2D drop-down menus [60], to more interactive 2D menus such as radial/ring menus [53], [43], [54] and more specifically 3D menus which can not be represented in 2D areas [20], [30].

As previously mentioned, hierarchical menus have become ubiquitous. Ever since their introduction, researchers have been trying to improve on the way a large number of actions and options are visually presented to the user in the form of a hierarchical menu. Along this line, in [78] the authors present a way to optimize the existing structure of a hierarchical menu (drop-down menu) such that the average selection time is reduced. Although their experiments show an increase in the selection speed of about 40% the proposed technique requires a set of user-defined thresholds which can be difficult to determine and tweak.

Another solution for reducing the pointing time and improving the performance is to visualize the menus in alternative forms such as radials [98] or squares [17]. Radial menus since their introduction by Callahan et al. [34] were shown to reduce the pointing time when compared to a traditional drop-down menu. At the time of the study in [34] the authors only considered single-level drop-down menus. It was much later that Samp et al [98] extensively compared the linear hierarchical menu (e.g. multi-level drop-down menu) with a radial menu in a desktop configuration, in terms of the visual search time as well as the pointing time. The results showed that by using radial menus the participants performed 34% faster in the pointing time. The results presented in the paper also showed that the drop-down menus perform better in the visual search time because

of the more intuitive categorization and visualization of the multi-level menus.

In [53] the authors presented a hierarchical radial menu specially designed for immersive virtual environments where the interaction method is based on hand movements recorded by a 6-DoF interaction device (rotation, projection, or picking a ray). The authors show that participants performed best using this pick-ray method with respect to the pointing time. The proposed method requires that the user points at a selection and a ray is cast from the user's hand (the controller) to that selection. The paper shows that visualizing a ray from the user's hand to the menu selection results in positive feedback when considering the pointing time in immersive 3D environments.

Another variation of radial menus has been presented by Davis et al in [43]. This variant of radial menu is implemented based on different depths. The different levels or hierarchies have been placed in the different layers of the menu. The selection process is still based on hand gestures recorded by a Leap Motion tracker (a vision based tracking system especially for gestures). The authors conducted a user study with 10 participants and the results emphasize the fact that this type of radial menu is easy to learn and capable of supporting a large number of menu actions and options.

Having a comparison between speech-based and radial-menu-based interaction metaphors in dealing with hierarchical menus in immersive AR, Pick et al. [89] confirmed that the speech-based interaction metaphor is faster while radial menu based interaction metaphor has less error ratio. The user study has been conducted through a 3D authoring application and different users asked to perform tasks using both radial-menu-based and speech-menu-based metaphors.

3.2.2 Selection Techniques

The interaction methods in AR are mostly via natural user interfaces such as hand gestures, head pointing, eye tracking and voice recognition. The manual inputs such as clickers or controllers are also used for selecting or manipulating the virtual objects [62].

Skiun et al [72] presented a new interaction method for immersive AR named Ubii (short for Ubiquitous Interface and Interaction) for system manipulation between physical and virtual world. Ubii enables the user of interacting with multiple smart devices such as laptop and desktop computers in a room (i.e. transferring files between them) using an AR immersive headset (Google Glass).

In their work, the hand gestures provide the primary input method for selection and manipulation while radial menus are used as the graphical user interface for the device.

Grinshpoon et al [55], proposed a hands-free user interface for the MS HoloLens headset. The user interface was used by surgeons who wanted to annotate the patient with virtual information during the surgery. In this scenario the surgeon is unable to use her hands for interacting with the device. Therefore head pointing and voice commands were used in order to interact with the AR headset. Similar head based interactions and interfaces for AR have also been proposed in [85], [96].

In [68] and [27] the authors performed a comparison between different input modalities (hand gesture, head pointing, eye tracking) for target selection in immersive AR with respect to pointing time and accuracy. In addition to the available headset sensors they equipped the headset with an eye tracker. The reported results of both show that eye tracking alone is faster than head pointing, but head pointing allows a greater targeting accuracy.

Moreover, authors in [68] also have shown that performing gestures in order to select virtual objects and manipulate them will need considerably more effort (physical and mental demand) compared to eye tracking and head pointing. This was also confirmed later by Hincapie-Ramos et al in [59] where they showed that using hand gestures to interact with immersive environments for a long time can lead to arm fatigue; a problem known as 'Gorilla arm'.

3.3 Application

We developed a simplified 3D authoring application in AR (consisting of a total of 20 menu items) that includes a number of tasks which involve actions to be performed on a 3D object. These tasks are presented in four different (menu type, input modality) combinations all of which are in the form of hierarchical menus: a drop-down menu controlled with hand gestures, a radial menu controlled with hand gestures, a drop-down menu controlled with voice and a radial menu controlled with head pointing. As previously mentioned several studies have been conducted over the years [34], [98] and have shown that radial menus reduce pointing time because of the short distances when compared to drop-down menus. Since this has been established already we opted out of including the (drop-down, head pointing) pair in our application. Similarly, the (radial, voice) pair

was also not included because when using voice for interaction the user does not interact with the actual menu per se.

The application presents the four tasks relating to basic operations on a 3D object, in sequence. For each task, the application starts with a pop-up window containing an explanation of which action the user should select, and the absolute path to the action e.g. Edit → Transform → Rotate. The information on the exact menu location of the action eliminates the possible bias which may be introduced in cases the user is familiar with 3D authoring software and already knows under which top-level menus the action is located. Upon reading the explanation, the user has to complete four tasks, namely: rotation, scale, move, and change of color). The task is considered as *completed* (a) if the user successfully selects the indicated action, or (b) mistakenly selects another action. The user must complete all four tasks in the sequence they appear in: rotate, scale, move, change color. Once all the tasks are completed the user is prompted with the next (menu, method) pair and is asked to repeat the same tasks. Thus, a total of sixteen tasks have to be completed.

The first menu type is a drop-down menu containing multiple levels of sub-menus and actions e.g. File, Edit, Geometry, Rendering and etc. The menu is represented as a 2D plane floating in 3D space and located next to the object of interest as shown in Figure 3.1b. We have implemented two interaction methods for this menu: hand gestures (which include tracking for pointing, and air-tapping for clicking), and voice control.

The second type of menu we have designed is a radial menu which presents all sub-menus and/or actions as circle-buttons spread out along the circumferences of a set of multi-level co-centric circles centered at the object of interest. An example is shown in Figures 3.1a, 3.1c. Due to the fact that a radial menu has a fixed center point and all options are guaranteed to lie on the circumference of one of the co-centric circles, we allow the user to move and rotate a line originating from the center in order to make a selection, using the two interaction methods which are most appropriate for this case: head pointing control, and hand gestures (which includes tracking for orienting and positioning the line, and 1.5s dwell-time for selecting). Menus using a rotating line for making a selection are known as marking menus and are extensively used in 3D authoring software. In the post-questionnaire we have used the terms 'line guided radial menu with hand gestures' and 'line guided radial menu with head pointing control' for simplicity and clarity purposes. Figures 3.2a,

3.2b, 3.2c show the application from the user's perspective.

Objective measures regarding the participants' interactions with the application were recorded throughout the duration of the experiment on a per participant, per task basis:

- start time: the timer begins as soon as the application starts
- completion time: the time is recorded when each task is completed. The task is considered as *completed* (a) if the user successfully selects the indicated action, or (b) mistakenly selects another action.
- cursor location: continuous recording at 0.1s intervals
- menu selections made
- wrong menu selection number for each participant
- video of what the user is seeing through the headset

From an implementation standpoint, the application was developed in Unity3D with Microsoft's HoloLens headset without using any peripheral devices.

There are no knowledge prerequisites to using the application however during the design and development cycles we noticed that the learning curve is quite steep for the first-time AR users. In particular, first-time AR users have difficulty familiarizing themselves with the hand gestures recognized by the device. For this reason, the users are first asked to follow an instructional tutorial on the recognized hand gestures and usage of the AR headset. After introducing the tutorial, the effective learning time for a first-time AR user drastically reduced to a few seconds as also confirmed by the pilot test.

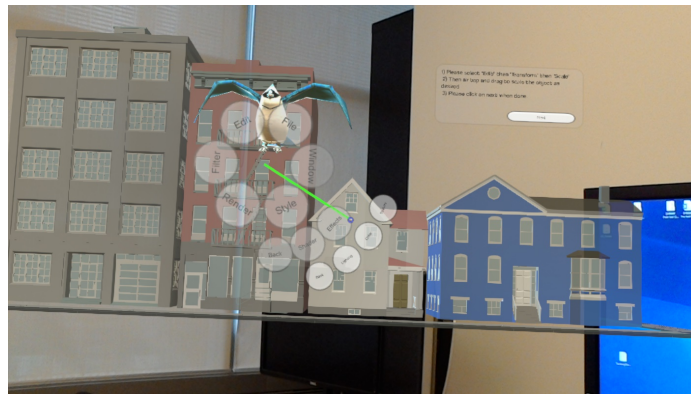
3.4 Methodology

3.4.1 Experimental Setting

The experiments received approval by the University Research Ethics Committee and were conducted over the period of 14 days and involved 25 participants. Prior to the study, all participants



(a)



(b)



(c)

Figure 3.2: (a) The application as seen from the user's perspective. The tasks include manipulating animated 3D objects i.e. bird flying in front of the buildings. The text in the dialog box (top right corner) asks the user to change the color of the animated bird using menu options. (b) A closeup of the hierarchical radial menu as seen by the user and (c) the animated bird after performing the color change. The green line is controlled by the user (either with hand gestures or head pointing) and the selection is made after a dwell-time of 1.5s.

were informed of the purpose of the study and procedures, after which they gave informed consent. The experiment duration for each participant ranged from 35 to 45 minutes.

The participants were first asked to complete a pre-test questionnaire containing questions relating to demographics and background information on their experience in using AR headsets/ technologies. Following the pre-test questionnaire, the participants had to complete a 5 minute calibration process of the headset and a 10 minute instructional tutorial on the recognized hand gestures such as air-tap and bloom, head pointing, voice commands, and general usage of the AR headset. Next, the participants had to complete all the tasks in the 3D authoring application in AR. Upon completion of the tasks all participants were asked to complete a post-test questionnaire containing specific questions on the menu types and interaction methods presented in the application. Throughout the experiment a member of the research team was present for observation (e.g. the contents from the user’s point-of-view displayed on a monitor connected to the headset) and to ensure that the procedure is followed correctly. Figure 3.3 summarizes the various steps of the experiments.

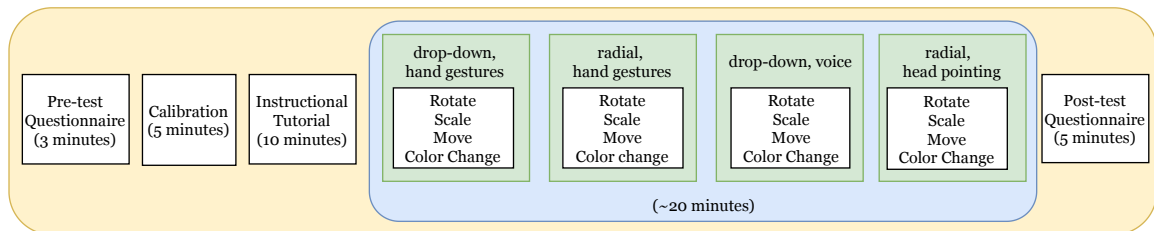


Figure 3.3: The experimental setting. All participants had to fill out a pre-test questionnaire followed by a 5-minute system calibration to adjust the headset, and a 10-minute instructional tutorial to familiarize them with the hand gestures used in AR e.g. air-tapping, bloom, etc. Next, the participants had to perform four tasks with four (menu, input modality) pairs, after which they had to fill out the post-test questionnaire.

3.4.2 Pilot Test

A pilot test of the application was conducted with 3 participants. The purpose of this test was to gather feedback and make adjustments and possible updates to improve the application as well as the procedure. The participants were asked to complete all steps of the process: complete a pre-test questionnaire, follow an instructional tutorial on the recognized hand gestures and usage of the AR headset, perform the tasks, and complete a post-test questionnaire. During the pilot test,

participants reported that they found it difficult to complete the tasks within the allotted time hence we have increased the time from 15 minutes to 20 minutes. Furthermore, based on the feedback provided by the participants, we introduced a 30 second break-interval when switching between the two menu types.

3.4.3 Participant Demographics

A sample size estimation with an effect size of 0.3 showed that a total of 24 participants were required for the experiments [47]. A total of 25 participants [32% female, 68% male] were recruited for the experiments. The participants ranged from 18-44 years old with the majority 52% within the age group 25-34, and 24% for each age group 18-24 and 35-44, respectively.

3.4.4 Analysis

In this paper, we employed a one-way ANOVA with repeated measures ($\alpha = 0.05$) for analyzing the pointing time i.e. the time it takes to navigate to an item and select it [98] using the four pairs of (hierarchical menu, input modality). We performed a descriptive post-hoc analysis on the data using Bonferroni confidence interval adjustment for pairwise comparison. In addition to that we have gathered the number of failed and successful attempts during the experiments for each user and for each task and have provided a comparison of these with respect to all four interaction techniques. After completing the tasks in the AR application the participants were asked to fill out a post-test questionnaire about the physical demand, frustration, and overall rating for each type of interaction. We have applied both quantitative and qualitative analyses to support our findings.

3.5 Evaluation

3.5.1 Pre-test Questionnaire

In the pre-test questionnaire, the majority 64% indicated that they had no to little experience using AR devices and selected 1 (48%) and 2 (16%) on a Likert scale of 5 [1 - 'I have never used an AR device', to 5 - 'I have used an AR device and I am comfortable in using this technology'] when asked to "Rate your experience using Augmented Reality (AR) devices". The remaining

36% indicated that they had some experience using AR devices and selected 3, 4, and 5 (16%, 4%, 16% respectively). In retrospect, considering the participants' responses, the decision to include an instructional tutorial was clearly justified; the tutorial served the purpose of familiarizing the participants with AR devices prior to the experiment, and therefore removing potential bias introduced from participants with prior experience using AR devices.

Figure 3.4 shows the responses for the question 'Rate your experience using X' where X is each of [Augmented Reality (AR) devices, hand gestures for interaction, 3D authoring software, voice input for interaction]. In summary, 32% selected 4 for 'voice input for interaction', 72% selected 1 for '3D authoring software...', and 44% selected 2 for 'hand gestures for interaction'.

Table 3.1 shows a comparison between the participants' experience in using AR devices versus experience in using hand gestures. As it is evident, of high statistical significance is the fact that 86% of the participants who indicated that they had no experience using hand gestures (selected 1) they also indicated that they had no experience using AR devices (selected 1). This seems to imply that they must have used hand gestures outside the context of AR e.g. using Leap motion, Kinect, etc, and were unfamiliar with the hand gestures used for pointing and clicking in AR.

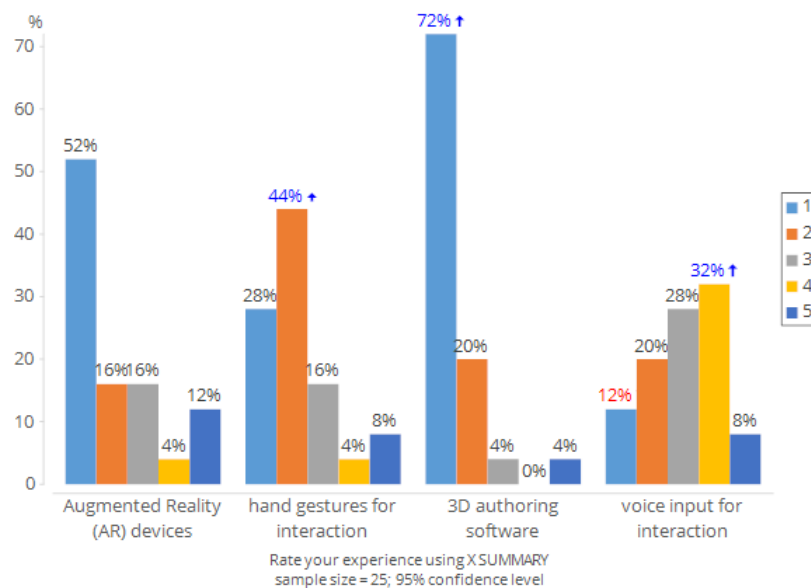


Figure 3.4: Pre-test responses for the question 'Rate your experience using X' where X is each of [Augmented Reality (AR) devices, hand gestures for interaction, 3D authoring software, voice input for interaction]

Table 3.1: Participants' responses for 'Rate your experience using Augmented Reality (AR) devices' versus 'Rate your experience using hand gestures', sample size 25, confidence level 95%

Col. %	1	2	3	4	5	NET
1	86%	36%	50%	0%	50%	52%
2	14%	27%	0%	0%	0%	16%
3	0%	27%	25%	0%	0%	16%
4	0%	0%	0%	100%	0%	4%
5	0%	9%	25%	0%	50%	12%
NET	100%	100%	100%	100%	100%	100%

Table 3.2: Pairwise comparisons of (menu, input modality). Significant differences appear between (*, hand gestures) with (drop-down, voice) and (radial, head pointing) pairs. In addition, there is a significant difference between (drop-down, voice) and (radial, head pointing) with respect to the pointing time.

(Menu, Input modality)	(Menu, Input modality)	Mean Difference	Std. Error	Sig.	Lower Bound	Upper Bound
(drop-down, hand gestures)	(radial, hand gestures)	-.746	1.002	1.000	-3.444	1.953
	(drop-down, voice)	2.030	.739	.043	.040	4.021
	(radial, head pointing)	3.560	.703	.000	1.669	5.452
(radial, hand gestures)	(drop-down, hand gestures)	.746	1.002	1.000	-1.953	3.444
	(drop-down, voice)	2.776	.736	.002	.794	4.758
	(radial, hand gestures)	4.306	.782	.000	2.200	6.412
(drop-down, voice)	(drop-down, hand gestures)	-	.739	.043	-	-.040
	(radial, hand gesture)	2.030	.736	.002	4.021	-.794
	(radial, head pointing)	2.776	.316	.000	4.758	2.381
	(radial, head pointing)	1.530	.316	.000	.679	2.381
(radial, head pointing)	(drop-down, hand gestures)	-	.703	.000	-	-
	(radial, hand gestures)	3.560	.782	.000	5.452	1.669
	(drop-down, voice)	-	.782	.000	-	-
	(drop-down, voice)	4.306	.316	.000	6.412	2.200
		-	.316	.000	-	-.679
		1.530			2.381	

Table 3.3: The two types of errors: selection and recognition error. The % shown is for the total 100 tasks performed by all participants i.e. 25x4 (rotation, move, scale and change color) for each (menu, input modality) pair.

(Menu, Input modality)	Number of experiments	Selection error (%)	Recognition error (%)
(drop-down, hand gestures)	100	10	31
(radial, hand gestures)	100	25	21
(drop-down, voice)	100	0	13
(radial, head pointing)	100	10	0

3.5.2 Quantitative Results

A quantitative evaluation was performed taking into account the two dependent conditions: pointing time, and the error. Pointing time refers to the elapsed time it takes for the user to point/move to a menu item and select it. The error is defined in terms of (a) the number of mistaken selections made (Selection Error), and (b) the number of failed interactions e.g. hand gesture not recognized, voice command not recognized (Recognition Error). The participants' total error rates are shown in Table 3.3. As it is evident the Recognition Error for the pairs (drop-down, hand gestures) and (radial, hand-gestures) is 31 and 21 respectively, which is higher than the pairs not involving hand-gestures. The table also shows that Selection Error for the (radial, hand gestures) pair is the highest in the four different combinations of menu type and input modality. This confirms the fact that using hand gesture for menu selection in immersive AR can lead to high error rate.

An analysis of the pointing times shows that there is a significant difference between (radial, head pointing) and both menus with hand gestures (the drop-down and the radial one). In particular, (drop-down, voice) and (radial, head pointing) resulted in the fastest pointing times when performing the four tasks ($F(3, 297) = 13.298, p < 0.05$). The pairwise comparisons are summarized in Table 3.2. Figure 3.5 shows the mean pointing times grouped by the tasks; (radial, head pointing) is the fastest with ($M = 7.06s, SD = 2.20$) with the second fastest being the (drop-down, voice) with ($M = 8.59s, SD = 2.96$).

Further analysis of the mean pointing times per participant shows that out of 25 participants 60% have performed their best performance using (radial, head pointing), 20% using (drop-down,

gesture), respectively 19.6% using (drop-down, voice) as the fastest, and and 0.4% using (radial, gesture).

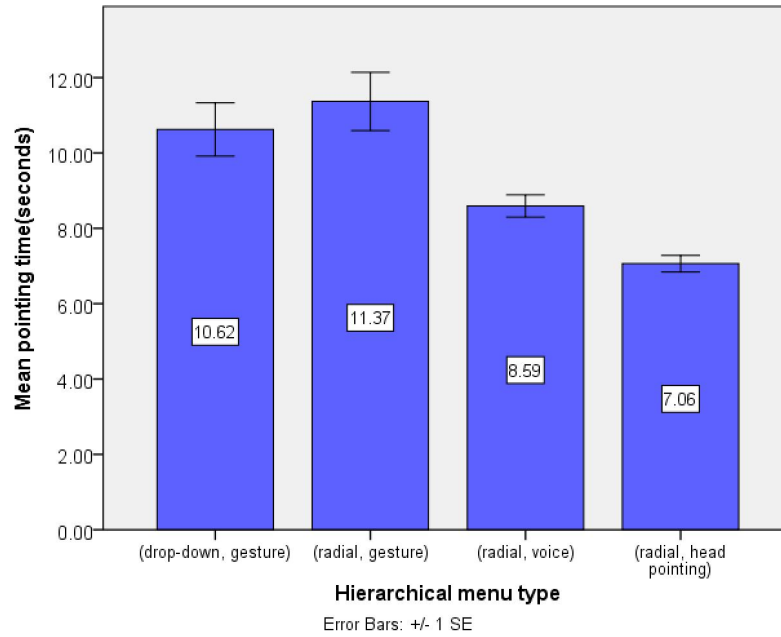


Figure 3.5: The mean pointing time per (menu, input modality) pair grouped by the tasks, and by the participants.

3.5.3 Post-test Questionnaire

Upon completing the test the participants were asked to fill-out a post-test questionnaire pertaining to the performed tasks with each (menu, input modality) pair.

The first section of the post-test questionnaire contained one question about the 'level of physical demand' and 'level of frustration' for each (menu, input modality) pair respectively on a Likert scale of 5. Figure 3.6 summarizes the participants' responses. Our analysis shows that there is a significant difference between the physical activities required when using the (radial, hand gestures) and the (drop-down, hand gestures). The results show that the (drop-down, voice) pair was the least demanding in terms of physical activities $F(3, 72) = 32.471, p < 0.05$. With respect to the 'level of frustration', the (drop-down, voice) pair was also shown to cause the least frustration amongst the participants with a close second being the (radial, head pointing) pair $F(3, 72) = 70.795, p < 0.05$.

In the second section of the post-test questionnaire, the participants were asked to provide an

overall rating for each pair on a Likert scale of 5. Figure 3.8 summarizes the results which clearly show a preference towards the (radial, head pointing) pair. In addition to the rating, we asked the participants to explain their choice. Fast object selection and less fatigue were the significant differences between the (radial, head pointing) and (radial, hand gestures). Furthermore, participants indicated that they did not prefer to use voice for interaction. Figure 3.7 illustrates the relationship between physical Demand/frustration (Y-axis), and average pointing time (X-axis). Each scatter plot represents a unique pairing of columns from the analyzed data: blue dots indicate the relationship between average pointing times and physical demand/frustration in (drop-down, hand gesture) pairs, green dots represent the relationship between average pointing times and physical demand/frustration in (radial, hand gesture) pair, red dots show the correlation between average pointing time and physical demand/frustration in (drop-down, voice) pair, and purple dots depict the association between average pointing time and physical demand/frustration in (radial, head pointing) pair.

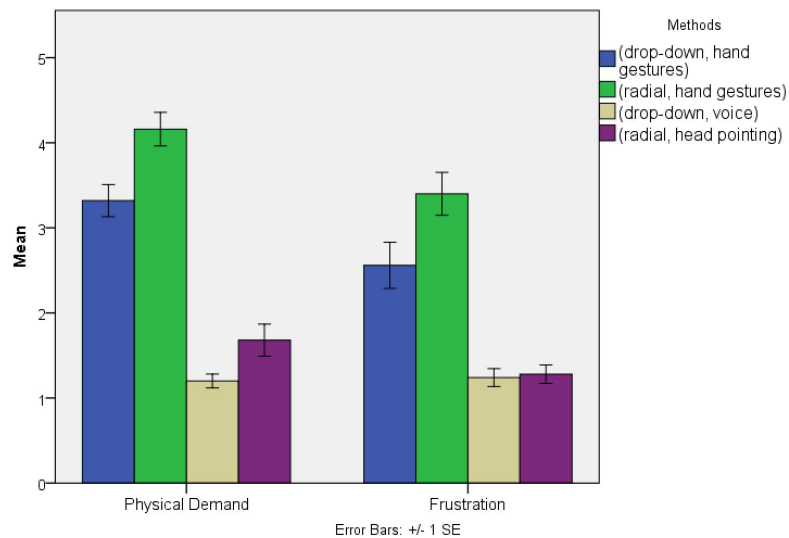


Figure 3.6: The mean responses from the users for physical demand and frustration when they were asked to rate the level of physical demand and frustration scaling from 1 to 5.

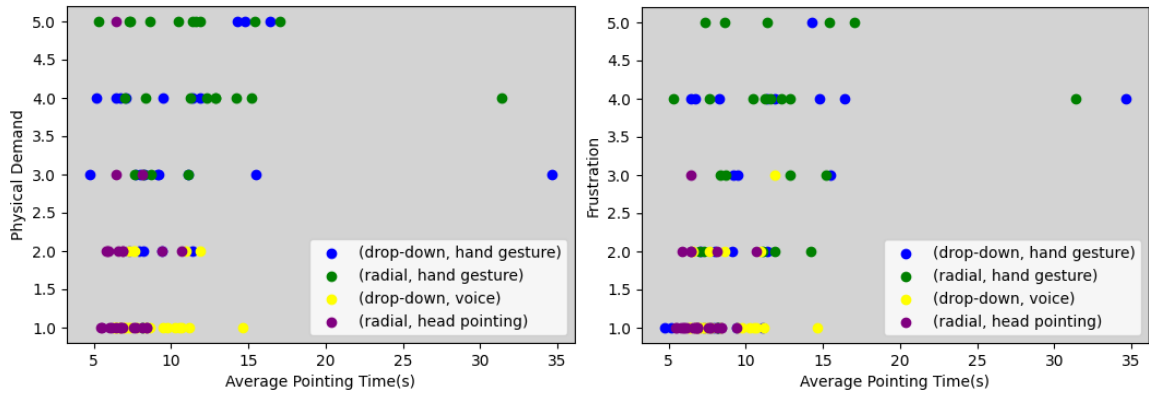


Figure 3.7: Scatter plots comparing physical demand and frustration (subjective rating by users) with average pointing times

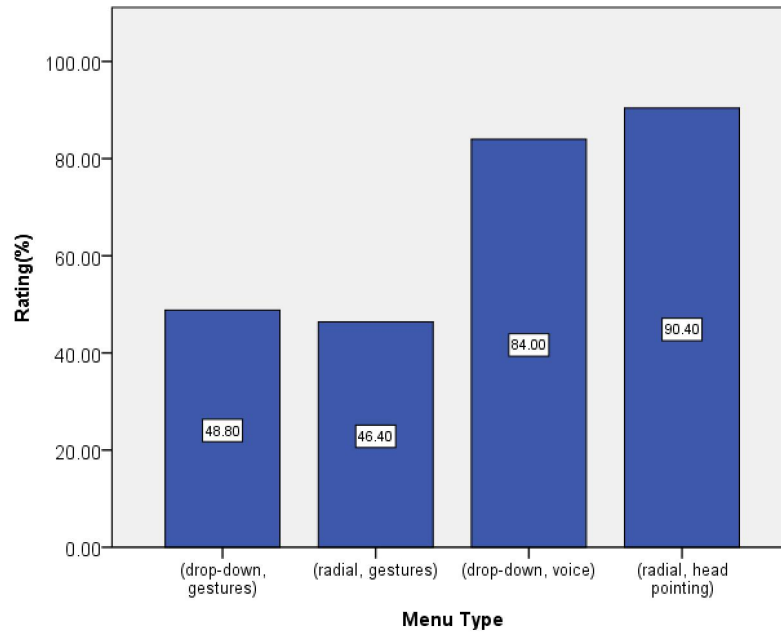


Figure 3.8: Rating of (menu, input modality) pairs. Each pair was rated on 5-point Likert scale.

3.6 Discussion

The analysis of the results shows that although the majority (85%) of the participants indicated familiarity with the drop-down menus in desktop applications, the majority (90%) gave the (radial, head pointing) the highest rating compared to the others. This can be attributed to the fact that since the early beginning of windowed applications drop-down menus have been used for visualizing the actions/options and are nowadays ubiquitous. Despite the fact, that in our application we have

implemented the drop-down menu in an object-centric manner - i.e. actions/options vary depending on the object of interest and are positioned in close proximity to the object in 3D space - the participants still found that the drop-down menu is not the best. This was also expressed by some participants in the 'additional comments' questions in their post-test questionnaire: 'Much easier to use than a traditional drop-down menu because there's more space between options. In drop down menus, options are crammed into a rectangle and there is a lot of room for error when selecting an option.'

Analysis of the objective measures showed that using (radial, head pointing) the participants were able to complete all tasks in a shorter time compared to the other (menu, input modality) pairs. Figure 3.5 shows the mean pointing times for all the tasks and all participants. In the post-test questionnaire the participants indicated that it was 'easy to use' and resulted in 'less fatigue'.

Furthermore, the fact that the (radial, head pointing) was given the second lowest rating in regards to the 'level of physical demand', and 'level of frustration' yet it was chosen as the preferred (menu, input modality) pair, can be attributed to the fact that people do not want to use voice as an input modality. The reason as to why they wouldn't use voice is either because of privacy concerns i.e. '...Moreover, I wouldn't want to stand out in public by talking out loud to a device.', or because they believe using voice in a noisy environment will cause recognition problems i.e. 'voice control in a crowded environment may not work very well due to noise'.

The study also has confirmed another important shortcoming of all (*, hand gestures) pairs, namely Gorilla arms. The main experiment (tasks) only lasted for a few minutes yet the participants reported high physical demand and fatigue (84%), which can be attributed to the fact that the user has to keep her hands raised when interacting with the application i.e. 'Hand pointing is way too tiring'.

The analysis of the error shows that the (radial, head pointing) and (drop-down, voice) pairs have the lowest error ratios across all tasks performed by the users. Two types of possible errors in the context of selection techniques in immersive devices [84], [108] have been evaluated in this analysis. The results show that each of the (drop-down, voice) and (radial, head pointing) pairs performs well in reducing the one type of error. In total, adding the two types of errors together the (radial, head pointing) pair performs better (13 and 10). This was also confirmed by the feedback

provided by the users where they indicated that voice control in a crowded environment may not work very well due to noise. The noise can cause problems in recognizing the words in voice recognition.

Another interesting outcome is the fact that (radial, hand gestures) was rated the lowest of the four pairs. As indicated in the post-test questionnaire some participants had some trouble controlling the line i.e. '...The line gets sometimes out of scope....'. We have observed that right-handed participants trying to make a selection on the left side of the radial (and vice-versa) had difficulties controlling the line.

3.7 Conclusion and Future Work

Graphical user interfaces in augmented/virtual reality are mostly inspired by 2D desktop environments. In this paper we present the results of a within-subjects user study comparing combinations of popular hierarchical graphical user interfaces coupled with popular input modalities for use in 3D authoring software in AR headsets. The results show that all (*, hand gestures) pairs were rated as the least preferred because of the physical demand involved. This confirms the 'Gorilla arm' problem already indicated in the literature. The (radial, head pointing) pair was quantitatively shown to be the fastest. Furthermore, when using the (radial, hand gestures) pair some participants indicated difficulties in controlling the line. This primarily occurred when right-handed participants were trying to rotate the line to make a selection on the left side of the radial, and vice-versa. We consider this to be a very interesting direction and we would like to explore how machine learning techniques can be used to re-arrange the menu items according to the frequency of their selection and the user's physical characteristics e.g. left-handed or right-handed.

Acknowledgements

This research is based upon work supported by the Natural Sciences and Engineering Research Council of Canada Grant DG-N01670 (Discovery Grant).

Chapter 4

Predicting Human Performance in Vertical Hierarchical Menu Selection in Immersive AR Using Hand-gesture and Head-gaze

This chapter is the paper published and orally presented at the 15th IEEE International Conference on Human System Interaction (HSI), Melbourne, Victoria, Australia [\[91\]](#).

Abstract

There are currently limited guidelines on designing user interfaces (UI) for immersive augmented reality (AR) applications. Designers must reflect on their experience designing UI for desktop and mobile applications and conjecture how a UI will influence AR users' performance. In this work, we introduce a predictive model for determining users' performance for a target UI without the subsequent involvement of participants in user studies. The model is trained on participants' responses to objective performance measures such as consumed endurance (CE) and pointing time (PT) using hierarchical drop-down menus. Large variability in the depth and context of the menus is ensured by randomly and dynamically creating the hierarchical drop-down menus and associated

user tasks from words contained in the lexical database WordNet. Subjective performance bias is reduced by incorporating the users' non-verbal standard performance WAIS-IV during the model training. The semantic information of the menu is encoded using the Universal Sentence Encoder. We present the results of a user study that demonstrates that the proposed predictive model achieves high accuracy in predicting the CE on hierarchical menus of users with various cognitive abilities. To the best of our knowledge, this is the first work on predicting CE in designing UI for immersive AR applications.

4.1 Introduction

In recent years, Augmented Reality (AR) has emerged as one of the most promising technologies with applications in gaming, education, and medicine [19], [41], [95]. Currently, AR hardware is categorized in terms of: (a) mobile AR, where the AR content is shown through hand-held tablets or smartphones, (b) spatial AR, which is based on projecting digital content on physical objects, and (c) immersive AR, where see-through head-mounted displays (HMDs) are used to augment the user's field-of-view. Each category of devices allows users to experience the real world overlaid with virtual, computer-generated content. One of the most important characteristics of AR is the interactions with virtual content [42]. In immersive AR, this feature is supported through graphical User Interfaces (GUI) and input modalities relying on natural interactions such as hand-gestures, head-gaze, eye-gaze, and voice recognition. Thus, the design of GUIs and the choice of input modality is of paramount importance for immersive AR applications, and a determining factor for their usability. Moreover, in immersive AR, the interaction with the virtual content involves considerable physical or mental activity [92], [22]. Therefore, it is important to design these interactions in a way that reduces the amount of overall workload on the users. This will be more important for prolonged usage of an immersive headset, i.e. performing a sequence of interactions to reach a final goal. As the applications become more complicated, containing numerous actions and options, the popular choice amongst designers and developers is to combine the aforementioned menu options into groups and hierarchies (i.e. hierarchical drop-down menus as shown in Figure 4.1).



Figure 4.1: We present an AI model for predicting human performance (pointing time and consumed endurance) in hierarchical menu selection in immersive Augmented Reality applications.

Although drop-down menus in immersive AR can be replaced with more user-friendly menus like radial [53] or tile menus [32], they are still widely used due to their familiar nature. This is especially the case for developers who transfer the two dimensional desktop menu layouts to three dimensional immersive AR. This means that optimizing the placement of the menu items in hierarchical drop-down menus can lead to improved interactions in immersive AR. To achieve this, one has to evaluate the performance of users of different cognitive capacities on the menu.

User studies must be conducted during and after the menu design process to provide insights into the usability concerning the effectiveness, efficiency, and overall satisfaction of the end-users. However, conducting user studies for immersive AR applications is a time-consuming and costly process. The participants must have access to an AR-capable HMD, e.g. HoloLens, Magic Leap, etc, which considering the current cost of HMDs is very unlikely. This restricts the type of user studies to on-site, face-to-face usability testing. This can be very time-consuming and expensive since participants are often compensated for their time. This is especially the case when an iterative design process is followed.

To overcome these challenges, we propose a predictive model that determines the users' performance for a target GUI without any involvement of participants during the user studies. The model

is trained on data acquired during a user study involving a large number of hierarchical drop-down menus, and the responses of participants to objective performance measures.

4.1.1 Human Performance on Hierarchical Menu Selection

Hierarchical menus are widely used in a variety of applications ranging from simple web pages such as portals and websites, to highly professional desktop software like 3D authoring applications (i.e., Autodesk Maya and Unity3D). Each phase of the software contains hundreds of actions and options in a hierarchical menu format. According to the related literature, [92,98] selecting a desired menu item, especially when it appears deep in the hierarchy, can be challenging. Based on several studies [49, 71], a user's performance on menu selection in a desktop configuration, depends on the pointing time (PT). The PT is defined as the time it takes for the user to find the desired menu option and point to it, i.e. the time duration between the first and the last click. This is different for immersive AR/VR, since the input modalities for interacting with the GUIs are different. VR mainly uses controllers to interact with the GUI, while AR relies on head-gaze or hand-gestures. Studies specific to immersive AR/VR applications [59, 92] have shown that the user's performance for menu selection also depends on additional factors like workload and error rate. The workload is defined as physical or mental effort required to point the cursor at the desired menu item and select it. Since errors are more likely to occur in immersive AR than in the desktop environment [22], the error rate is also an important measure for evaluating the performance. Increased error rates are attributed to the fact that most interactions in immersive AR, are performed using hand-gestures and pointing the head-gaze. This increases the workload and amplifies the possibility of an error during a selection.

In this work, the user's performance on a hierarchical GUI is predicted considering the factors like expected PT and physical workload corresponding to selecting a menu item from a hierarchical drop-down menu through a sequence of menu selection tasks in immersive AR. Inspired by related work in the area, we consider the semantics and organization of the menu items (i.e. number of levels, number of items in each level, and number of characters for each target in the every level) and the users' cognitive ability in performing visual tasks (WAIS-IV test).

Previous works on performance prediction in HCI tasks categorize the measures into two sub-groups, user-related and GUI-related factors. A brief description of such factors is presented in this section.

User-related Factors

The user-related factors depend on the capabilities of the users performing selection tasks. Their learning ability, cognitive performance, and fatigue are important factors that affect the performance of menu selection. These factors are recorded during the user study and considered in the final neural network for predicting the human performance. We applied a standard cognitive approach, WAIS-IV test [105], to estimate each user’s mind Processing Speed Index(PSI).

GUI-related Factors

The appearance of a menu is an essential factor in the user performance during a selection task. Shape of the menu (i.e., vertical list, radial) [92], saliency of items within, their semantic meaning, and the organization in general (i.e., number of levels) [71] are the most important factors. In this work, we focus on the vertical lists such as hierarchical menus with the maximum depth of 3 and consider the saliency, semantics, and organization of menu items as a part of the machine learning model input.

4.1.2 Contributions

The contributions of this work are as follows:

- To the best of our knowledge, this paper is the first work on predicting human performance of vertical hierarchical menu selection in immersive AR incorporating both user-related and GUI-related factors.
- This work is the first work that considers the semantic relationship of the targets at different levels in the hierarchy. Previously, Li et al. [71], focused only on the semantics of the words in vertical menus. We demonstrate that when it comes to a hierarchical menu selection, the semantic relationship between the targets in each menu level is an important factor to consider.

4.2 Literature Review

Human performance is an important factor in designing and developing user interfaces(UIs). The performance of a UI can be calculated using statistical methods and user studies with a certain number of participants [31]. Although the performance of a UI can be evaluated using user studies, predicting the performance using extensive mathematical and machine learning models without testing it with real users will be less time-consuming and less expensive. The related work in predicting human performance in HCI is divided into two broad categories, namely the traditional methods that mostly use statistical and mathematical approaches such as Fitts' law [50] and Hick's law [58] and the novel machine learning methods such as [71, 101, 109] which use deep neural networks to consider all user-related and GUI-related factors.

4.2.1 Traditional Methods

There are a plethora of mathematical models presented to predict human performance in HCI. Fitts' [50], and Hick's law [58] are the two fundamental laws for predicting task completion time and form the basis for numerous prediction methods. The complex models such as [24] in the literature, predict visual search time based on Fitts' law. However, the model [24] predicts visual search time in vertical menus without considering the saliency of the items (i.e., number of characters in a word), which is one of the important factors in vertical menu selection. The work also does not consider the semantic meaning of each item as an important factor.

Considering the visual appearance of GUIs, the authors in [63] presented a computational model of visual search on graphical layouts. They gave several examples and user studies indicating how users learn GUIs and how changes in visual design will affect the performance of the UI. Bailly et al. [24] and Li et al. [71] call this measure as a learning effect and consider it in vertical menu selection.

Regarding the immersive environments, in [46], the authors proposed a mathematical model for predicting task execution time for mid-air and touch-based human hand-gestures. Although, it predicts the task completion time with $R^2 = 0.936$ based on their own dataset gathered in a user study, they do not consider predicting the performance in a sequence of tasks with factors such as

arm fatigue [59], learning effect [71] and GUI-related factors i.e. semantics and menu appearance.

4.2.2 Machine Learning-based Methods

Recent advances in machine learning and deep learning motivated HCI researchers to provide machine learning models for HCI applications. Therefore, predictive models of human performance are of significant interest to many researchers. Although traditional models perform very well in predicting human performance in simple cases with the condition of removing factors such as learning effect, semantics, task load and user's cognitive performance, they do not work well if all possible effective factors are considered. In such situations, the machine learning methods appear to be resolving the issues [110].

The authors in [71] presented a novel deep recurrent neural network to capture semantics, learning effect and saliency of menu items to predict visual search time in vertical menu selection. The method outperformed previous traditional works [24] in terms of accuracy and goodness of fit (R^2). However, the method does not explain the performance of hierarchical menu selection which is most common in many real use-cases. Another work that uses recurrent neural networks, Wei et al. [106] presented a deep model to predict the mouse-click position in a sequence of interactions. The LSTM model was trained on previous paths of the mouse and could predict the future mouse-click area with a better accuracy and error rate compared to previous works including [69] and [86].

The aforementioned methods perform well in desktop environments. However, there is no indication that using the current 2D based models can also predict the performance in menu selection in 3D immersive environments using hand-gestures and head-gaze as the input modalities. In such environments the workload of the tasks is higher compared to desktop environments [22] and it can affect the performance of the menu selections by user.

In conclusion, since in the immersive AR environment the possibility of noise and error and workload are higher, more complicated tools have to be applied to predict the performance of user. In the following sections, we present our method for predicting human performance in vertical hierarchical menu selection in immersive AR.

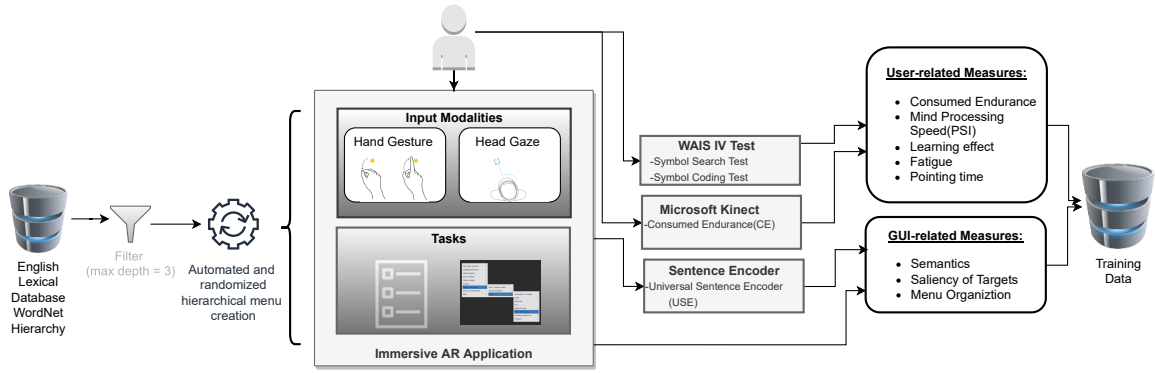


Figure 4.2: The diagram shows the data acquisition procedure during the user studies. An integrated AR application is developed on MS HoloLens I with hundreds of hierarchical drop-down menu selection tasks. The menu hierarchies were extracted from WordNet [48], a lexical database of semantic relations between the words in English. The users performed the menu selection tasks through two separate user studies using hand-gestures and head-gaze with MS HoloLes I. A Microsoft Kinect I recorded the user’s hand movements during the user studies. Finally, the data from the whole procedure is used to train and validate our machine learning model.

4.3 Methodology

This research aims to provide a validated predictive model of CE and PT for hierarchical menu selection in immersive AR. Two user studies were conducted to gather the required training data, and evaluate the machine learning model. The proposed model was trained on the data from the first user study and evaluated through a set of unseen data collected from the second user study. In the following sections, we present details for these user studies. The diagram in Figure 4.2 summarizes the components of data collection procedure during the user studies.

4.3.1 Application

We developed an immersive AR application on MS HoloLens I, coupled with a Microsoft Kinect I for data acquisition. The integrated application is developed with Unity3D and includes two separate modes for training and evaluation. Microsoft’s Mixed Reality Toolkit (MRTK) is employed to develop immersive AR features and interactions. The applied input modality was the combination of hand-gestures (for menu selection) and head-gaze (for moving the cursor), as they are most common in immersive AR headsets. The GUI consists of a regular hierarchical drop-down menu as this is widely used in AR/VR applications. The hierarchical menu contains hundreds of drop-down

menu selections for different tasks in an immersive AR environment. Since the usual menu depths in any typical software varies between 2 to 3 (i.e., in 3D authoring applications such as Autodesk Maya or 3D Studio Max, or Unity3D), we have selected 3 as the maximum menu depth in our application. Each user selection is randomized and the task depth varies from 2 to 3. A display prompt on the top showed the item to be selected from the hierarchy, and the users were instructed to perform that selection using either hand-gestures or head-gaze. The PT and user's hand movements for each selection were recorded using the built-in application's procedure with Microsoft Kinect. In order to have a complete hierarchy of words in our menu, we used the taxonomy of WordNet [80], which is an extensive lexical database of English nouns, verbs, and adjectives. For example, "sport → cycling → dune cycling" is showed on the display prompt, asking the user to select dune cycling from the hierarchy, as a single task. The semantics of the words on each level and the relation between the words are applied in the performance prediction pipeline.

4.3.2 Measures

In addition to measuring the pointing time (PT) for each menu selection task, we also measure CE for evaluating physical workload, WAIS-IV test for assessing the participants' cognitive performance, Semantics, and Organization of menus for assessing the GUI-related aspects of menu selection. This section discusses these measures.

Consumed Endurance (CE)

The authors in [59] proposed a measure for evaluating workload in tasks with mid-air hand-gestures. This method can be used in immersive AR/VR applications that rely on mid-air hand-gestures as input modalities. CE requires a Microsoft Kinect to track the user's hand movements in real-time and assigns a workload value to the 3D manipulation task. The model is derived from the biomechanical structure of the upper arm for both male and female users, and it is widely used as a metric for evaluating the Gorilla Arm effect. Our proposed model predicts CE for hierarchical menu selection tasks.

WAIS-IV Test

Prior to performing the menu selection tasks, the participants were asked to complete a paper-based subtest of WAIS-IV [104] to measure their cognitive ability to perform visual menu tasks. This data lets the machine learning model discriminate between users in performing menu selection tasks and provide the designers and developers with insights on their design's personalized performance. The complete WAIS-IV test is a standard test that covers all aspects of human intelligence, including Verbal IQ (VIQ) and Performance IQ (PIQ). For our experiments, the Processing Speed Index (PSI) is measured as part of the PIQ. The PSI test consists of two subtests: Symbol Search and Symbol Coding. The symbol search test includes 63 symbol search matching questions printed on seven one-sided A4 papers. The users were asked to find similar symbols in each line and mark them during the standard 120 seconds to complete the test. On the symbol coding test, the user had 120 seconds to copy symbols paired with the numbers. The output will be two integer numbers, symbol search results between 0 and 63 and symbol coding between 0 and 135.

Semantics and Menu organization

Based on the results of [71] presented by Li et al., the semantics of menus affect the performance of vertical menu selection. Therefore, including the semantics of menu items forming a list improves the accuracy of the predictions. In our work, we also extract the semantics of the menu items using sentence encoding techniques. A novel encoding approach named Universal Sentence Encoder (USE) [40] is applied on the menu items to encode the varied length menu lists to a fixed length one. USE takes as input a sentence and produces a 512-dimensional vector representing the items' semantic meaning.

The hierarchical menu organization is also considered as one of the inputs for the machine learning model. For each target in each task, the menu item's position in the list coupled with the number of whole menu items in that list is considered. The saliency of the targets is also considered and is calculated as the number of characters for each word.

4.3.3 User Study Procedure

This research received approval from the University’s Research Ethics Committee. Two user studies were conducted to gather training data and evaluate the machine learning model. The participants were assigned through a random process to do either the first or second user study. The participants were first asked to complete a pre-test questionnaire containing questions relating to their demographics and background experience in using immersive AR technologies through each user study. Then they were asked to complete the WAIS-IV test in the researcher’s presence. A 10-minute calibration of HoloLens I and a 10-minute instructional tutorial were the following steps before starting the main hierarchical menu selection attempts. Participants were asked to perform menu selection tasks with MS HoloLens I during the main test. This part included five menu selection attempts consisting of seven menu selection tasks. The number of attempts and tasks resulted from a pilot study conducted before the main study considering the user’s subjective feedback of the test. Considering the high physical demand of immersive AR tasks using hand-gestures, the users were given a 2-minute break between performing each menu selection attempt. The menu selection tasks were presented in a single line of instruction at the top-right corner in the user’s view, informing the user which menu item should be selected. A Microsoft Kinect recorded the participant’s hand movements during the activity, and the CE was calculated per task. Figure 5.6 shows the procedure followed and Figure 4.4 shows the setup of environment during the user studies.

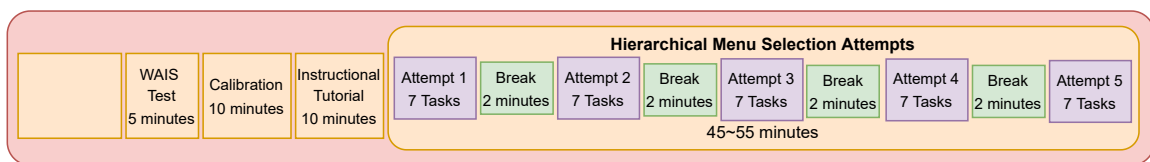


Figure 4.3: The participants had to fill out a pre-test questionnaire followed by a 5 minutes standard WAIS-IV Test and a 10-minute calibration to adjust the headset. A 10-minute instructional tutorial to familiarize the participants with hand-gestures used in immersive AR, e.g., air-tapping, was presented as the next step. Then, the participants had to do the main part of the test: the menu selection tasks. The section consists of five menu selection attempts, including seven hierarchical menu selection tasks. To avoid users’ fatigue, each attempt was followed by a 2-minute break.

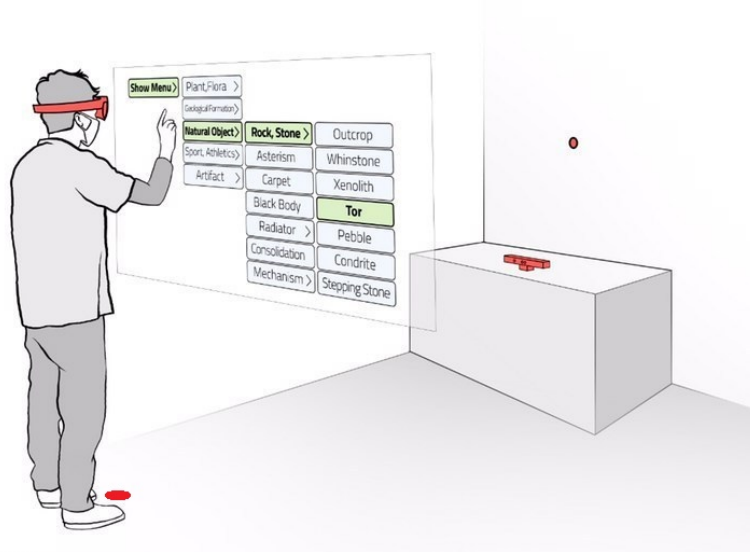


Figure 4.4: The participant is performing menu selection tasks while the MS Kinect records the hand motions. The red dots on the ground and the wall in front of the user are used for having similar initiation configurations for all users.

4.3.4 Participant Demographics

A total number of 26 users (12 female, 14 male) participated in our user study for both training and evaluation parts. The participants were randomly assigned to the training/evaluation user study. The participants ranged from 24-39 years old with the majority of 73% within the age group 27 to 35 and 11% within group of 35 to 39 and 16% within the age group of 23 to 27.

4.4 Model Design and Learning

We trained a recurrent neural network on user-related and GUI-related measures extracted from the user studies to predict the PT and CE of hierarchical menu selection tasks in a sequence of interactions. The input to the model consists of three parts, including normalized WAIS-IV test data representing the user-related measures, organization of menu items, and semantic vectors as it is shown in the Figure 4.6. The model's output is the desired CE and PT through a sequence of tasks. We took advantage of a recurrent neural network to assess the user's activity through a series of tasks and consider the effects of the learning effect and the user's fatigue while performing the tasks. In this section, first, we will explain the data preprocessing for preparing the data for feeding to the machine learning model, Section 4.4.1, second the architecture of the machine learning model

in Section 4.4.2 and finally, the training of the model in Section 4.4.3.

4.4.1 Data Preprocessing

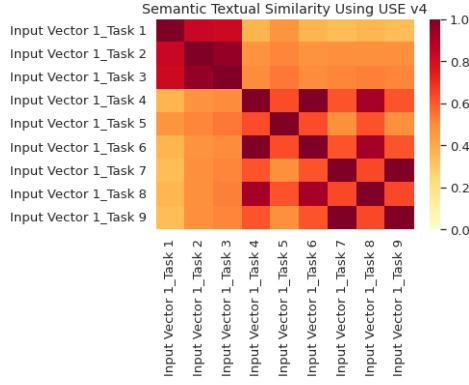
Menus and hierarchies have variable length and depth. To address this problem, we converted the sequence of menu items to fixed-length vectors using word embedding techniques as discussed in [71] and employed the USE technique [39]. The preprocessing was applied to the inputs as it is indicated in Figure 4.6. Since the USE receives sequences of words, we created comma-separated sentences from menu items, including all menus from the start to the end of the list, namely Input Vector 1. Another sentence was also produced from targets in each level, separated by a single comma to emphasize each level’s target, emphasizing semantic relation between the targets, namely Input Vector 2. The word vectors were later fed to USE to be converted to fixed-length vectors. Figure 4.5 shows an example of the similarity of the fixed-length embedding vectors for nine different tasks.

4.4.2 Neural Network Model

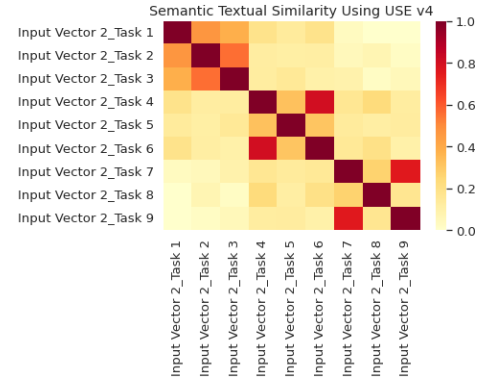
The machine learning model consists of the encoder and the prediction networks, as shown in Figure 4.6. To encode the varied length hierarchies, we used USE [39] which is a transformer-based model designed to convert the varied-length paragraph to fixed-length float value vectors. After converting the tasks to fixed-length vectors, the prediction network is trained on the user’s performance on a sequence of tasks to predict the PT and CE. We used an LSTM model with one hidden layer, an input layer size of 523, and an output layer size of 2 (CE and PT). The training window size for considering the previous tasks was 15, meaning that the prediction of performance for each menu selection task is based on the previous 15 tasks performed by each user. Therefore, the recurrent model was able to predict the learning effect through the tasks.

4.4.3 Training

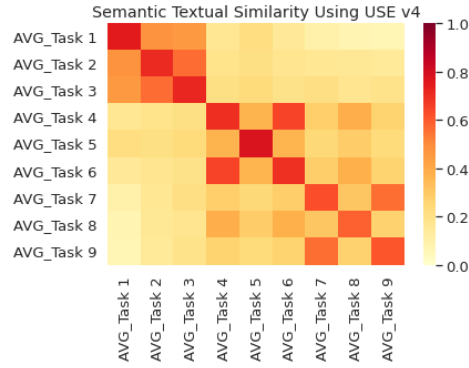
We trained our model on a workstation with NVIDIA Geforce RTX 2060 GPU using the PyTorch library. The training data size is 840, and the test data size is 140, including the menu selection tasks done by users. Since this is a regression problem, we optimized the Mean Squared



(a) Semantic similarity map for input vector 1 for 9 selected tasks



(b) Semantic similarity map for input vector 2 for 9 selected tasks



(c) Semantic similarity map for average value of input vectors 1 & 2

Task 1 "sport, skating, skateboarding"
Task 2 "sport, skiing, cross-country"
Task 3 "sport, rowing row, sculling"
Task 4 "natural, rock stone, pebble"
Task 5 "geological, ice mass, ice floe"
Task 6 "natural, rock stone, wall rock"
Task 7 "artifact, Ready-made"
Task 8 "natural, Consolidation"
Task 9 "artifact, button"

(d) The targets for each menu selection task

Figure 4.5: USE [39] is applied to the tasks to capture the semantic similarity by converting them to fixed-length vectors. The colors along with the diagonal show semantic similarity between the tasks.

Error (MSE) during the training. This is equivalent to optimizing the R^2 , which is a measure for determining the goodness of fit. We use the Adam optimizer with a learning rate of 0.00002. In addition, we found that the dropout rate of 0.3 reduces over-fitting.

4.5 Discussion and Analysis of the Results

We have evaluated our model using test data extracted from the evaluation user study. The model was trained on optimizing R^2 as the goodness of fit. In other words, the aim is to regress the PT and CE for a sequence of unseen tasks from the test data gathered from the user study with four random participants having presented to them completely different menu hierarchies than the ones in the training data. Therefore, the model was trained to capture different aspects that affect

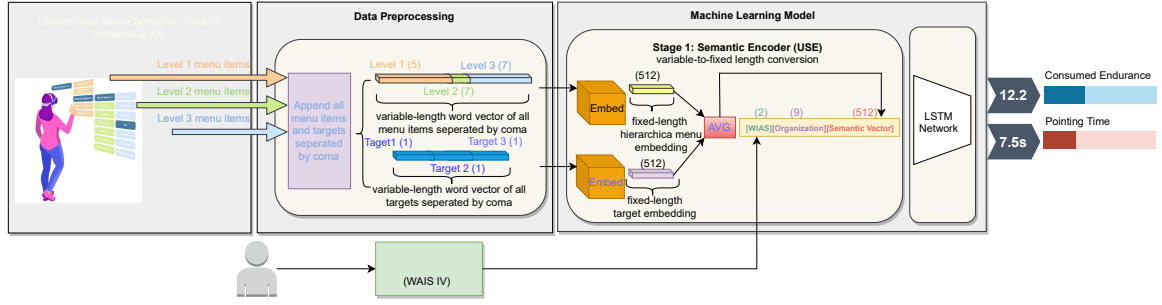


Figure 4.6: The data preprocessing and machine learning pipeline. In order to convert varied length menu items to the fixed length vectors firstly, a sequence of all items starting from the beginning of the first-level to the end of last level d ($d=2, 3$) separated by single comma between each two words created, Input Vector 1. Secondly a sequence consisting of all target items in each level, separated by a single comma is created, Input Vector 2. The two sequences were passed to the USE network [39] to be converted to a fixed-length 512-dimensional vectors. We applied the pre-trained models of USE which is available on TensorFlow Hub. Finally, we found that the element-wise average of two vectors lead to better results for the validation set. The WAIS-IV data and the organization of the tasks are also concatenated to the result of the semantic vector.

performance including semantics, menu organization, and users' cognitive performance. Figure 4.7 shows the predictions and actual values for PT and CE for one of the users from the evaluation user study. Table 4.1 shows the results of the predictions for each of the users who participated in the evaluation user study. Each table row consists of the R^2 and MSE for each user. The results show that our model effectively predicts the CE and PT in a sequence of hierarchical menu tasks in immersive AR. Our investigations also show that the model performs better after learning the users behaviors in selecting the menu options i.e. after several tasks performed by the user. We confirmed this by evaluating the average error for the first five tasks and the last five tasks of each user.

Although we achieved lower R^2 s compared to the previous works such as [71], our work has some specifications that make it different from other works. Firstly we trained our model on the data gathered in an immersive AR application. Interacting with virtual content in an AR application includes a lot of inconsistency and noise since the input modalities are head-gaze and hand-gesture as the work [92] shows that the error rate in menu selection in such an environment is high. On top of that, most of the users who participated in our user study had almost no previous experience in working with immersive AR/VR headsets. As we asked from users in the pretest questionnaire, only 7% of them claimed they had good experience using AR headsets. That was why we added an instructional tutorial for each participant before starting the menu selection tasks. Secondly, in

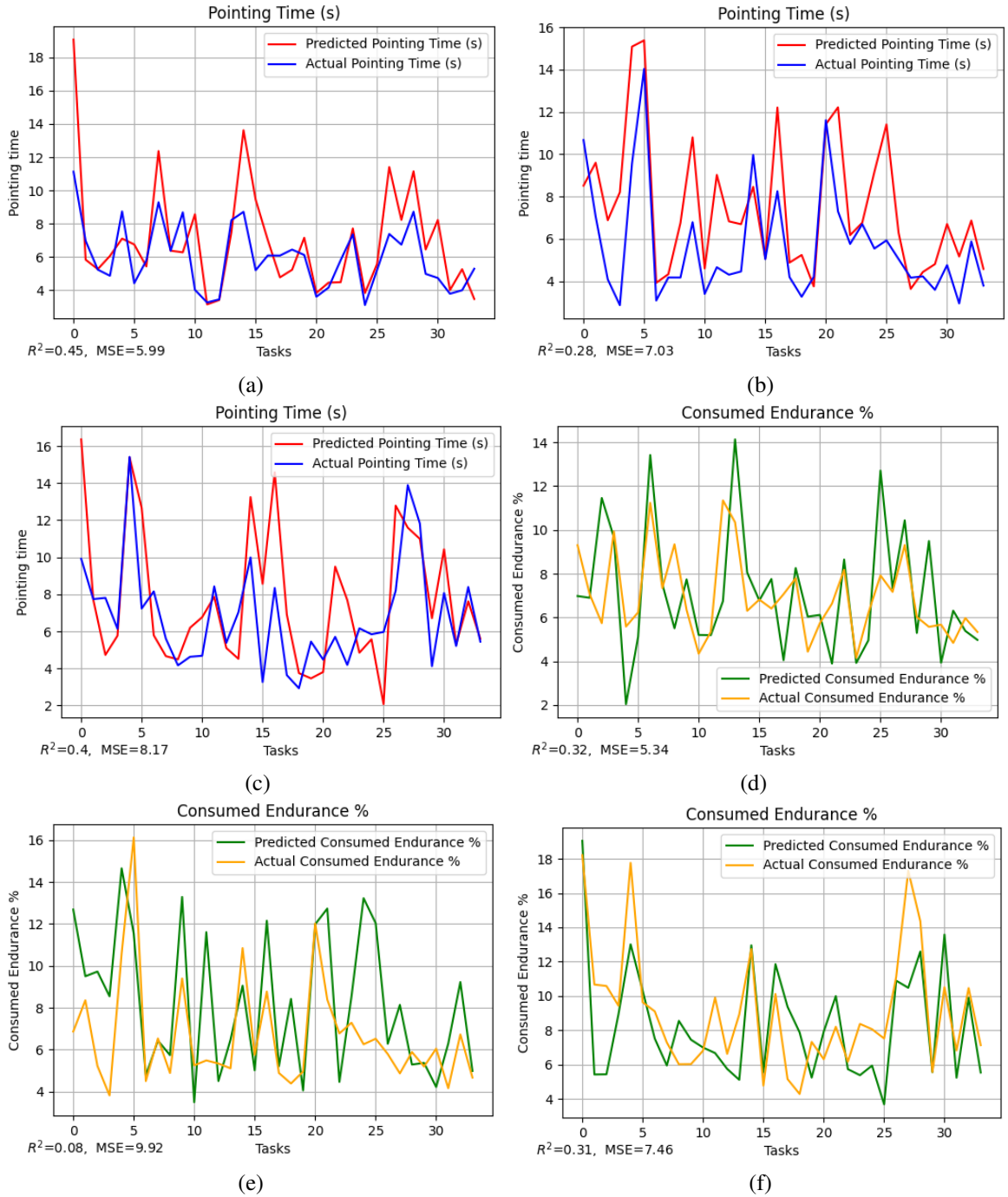


Figure 4.7: The predicted CE/PT versus the actual CE/PT through 35 tasks done by a participant in the evaluation user study. The vertical axis is the CE/PT and the horizontal axis represents the order of tasks from 0 to 34 (as it is depicted in Figure 5.6 each participant performed 35 tasks in 5 attempts each one included 7 tasks).

Table 4.1: Results of Average R^2 and MSE for predictions PT and CE on the evaluation dataset

User	R^2 CE	MSE CE	R^2 PT	MSE PT
1	0.32	5.34	0.45	5.99
2	0.08	9.92	0.28	7.03
3	0.31	7.46	0.4	8.17
4	0.26	6.06	0.31	6.37
Average	0.24	7.19	0.36	6.89

Table 4.2: Results of Ablation study to determine the importance of each feature

Feature to Remove	R^2 CE	MSE CE	R^2 PT	MSE PT
No Remove	0.24	7.19	0.36	6.89
Without WAIS	0.11	11.23	0.25	8.50
Without Semantics	0.03	8.74	0.14	7.11
Without Organization	0.0	15.67	0.1	12.37

this work, we employ hierarchical menu designs instead of a simple drop-down menu with only one depth level. Having a hierarchical task may increase the chance of error and inconsistency by itself. Thirdly, our experiments on the test data for immersive AR tasks, shows a positive correlation between the accuracy of the model and increasing the size of the training data.

4.5.1 Ablation

To evaluate the effect of each input component on the predictions, an ablation study is conducted by removing components from the input vector. WAIS-IV data, semantics,s and organization of menus are the components that were removed at each step. The results of training the network without these components are shown in Table 4.2. The ablation study shows that the combination of semantics, WAIS, and the organization of menus appeared to be accounts for much of the increase in the accuracy of prediction.

The analysis of the results also shows a positive correlation between the WAIS-IV test results and the performance of users in immersive AR, and thus integrating the WAIS results for a user improves the accuracy of the prediction. This also confirms the reported correlation between the

Table 4.3: Pearson’s correlation coefficient, ρ , indicating the correlation between the WAIS-IV data and our performance measures

Performance Factor	Total WAIS	Symbol Search	Symbol Coding	Age
Consumed Endurance	-0.324	-0.443	-0.167	-0.11
Pointing Time	-0.284	-0.295	-0.214	-0.434

WAIS-IV test result and the PT and amount of physical activity for each task. A Pearson’s correlation analysis has been applied on the results of the WAIS-IV (Symbol Search and Symbol coding) for different participants and their CE measurements, PT and error rate. The results clearly showed that there is a moderate degree of negative correlation between the WAIS-IV results and the CE values as well as PT. The Table 4.3 shows the correlation between PT, CE, WAIS data, and the age of the participants.

4.6 Conclusion and Future Work

Predicting human performance is beneficial in designing and developing user interfaces. Various works currently predict human performance in desktop or smartphones that works perfectly but not any recent work in immersive AR. Interacting with virtual content in immersive environments is mainly done by natural interactions such as hand-gestures and head-gaze. These natural interactions can increase the possibility of error, noise, and user fatigue. In this work, we presented a method for predicting human performance of hierarchical menu selection in immersive AR using hand-gestures and head-gaze as the input modalities. Our model builds upon previous work in the area for vertical menu selection in desktop environments [71]. The features involved in the performance of menu selection are divided into two categories: user-related and GUI-related measures. We have applied both categories to reach better results for predicting human performance through a sequence of tasks. To this end, we conducted two user studies to gather the data used for training and evaluating the machine learning model. The results show that applying the cognitive WAIS-IV test improves the predictions and reduces the effect of environmental noise and users’ fatigue. A natural extension of this work is its application to other graphical user interfaces such as radial menus and input modalities such as hand-gestures and head-gaze.

Acknowledgements

This research is based upon work supported by the Natural Sciences and Engineering Research Council of Canada Grants No. N01670 (Discovery Grant).

Chapter 5

Analysis of Error Rate in Hierarchical Menu Selection in Immersive Augmented Reality

This chapter is the paper published and presented at SPIE Conference on AR VR MR 2023, San Francisco, California, USA [93].

Abstract

The emergence of new immersive AR/VR headsets recently resulted in major improvements in hand-gesture-based user interfaces. Devices such as MS HoloLens II and Oculus Quest II support hand-gestures. Although using hand-gestures increases the sense of presence and ease of natural interactions, it has been shown that hand-gestures require extensive physical activity. Furthermore, it has been shown that the error rate in hierarchical menu selection is much higher when using hand-gestures than when using a desktop environment or the controllers. Therefore, assessing the difficulty of a hierarchical menu design when using hand-gestures and gaze for menu selection will enable UI designers to develop more effective user interfaces. In this work, we provide a validated index for estimating the hierarchical menu selection error using hand-gesture and head-gaze as input modalities. The index is informed by cognitive WAIS data gathered from participants, which measures subjective cognitive performance. The proposed index is the result of a user study that includes

hundreds of hierarchical menu selections using MS HoloLens, and is validated against the data of a group of different participants. The results demonstrate that the index can successfully capture the trend of the users' errors in selecting the hierarchical menu items in immersive environments.

Keywords: Immersive Augmented Reality, Hierarchical Menus, Hand-gesture, Menu Selection

5.1 Introduction

Immersive Augmented Reality is one of the most promising technologies that has emerged recently. Through immersive AR, the user's view will be augmented with computer-generated content. There are currently widespread applications for immersive AR in various fields, such as education [67] and medicine [52]. One of the most important features of immersive AR applications is the ability of the users to interact with computer-generated content. Interactions in immersive AR are applied through various input modalities and graphical user interfaces (GUI). According to the current state-of-the-art research in immersive AR, the combination of hand gestures and head gaze is a widespread interaction technique among AR headset designers and users because of its natural relationship with real-world 3D content manipulation [87]. There is also a controller option as an alternative interaction technique in recent immersive headsets such as HoloLens and Meta Quest for users that are not comfortable with hand gestures. Although there are problems such as Gorilla arm [59], neck fatigue [97], restrictions in hand gesture recognition, and high error rate [92] in menu selection, some users prefer to have no controller in their hand while working with immersive content which confirms the fact that fewer peripheral devices and more AI features for recent immersive AR applications are vital. Therefore, there is high interest in solving current interaction problems using hand gestures and head gaze for immersive AR.

Addressing the problem of a high error rate in hierarchical menu selection in immersive AR using hand gestures and head gaze, in this research, we aim to develop an error index for analyzing and evaluating the error rate in hierarchical menu selection for each person given the appearance of the menus and the user's cognitive performance. We applied both user-related and GUI-related measures, such as cognitive mental performance, age, and hierarchical menu complexity, to calculate the error index. The index was later evaluated through a user study with 25 participants.

The proposed index appeared to follow the trend of the error rates for each user and discriminate between the error rates of different users with different cognitive abilities and AR experiences.

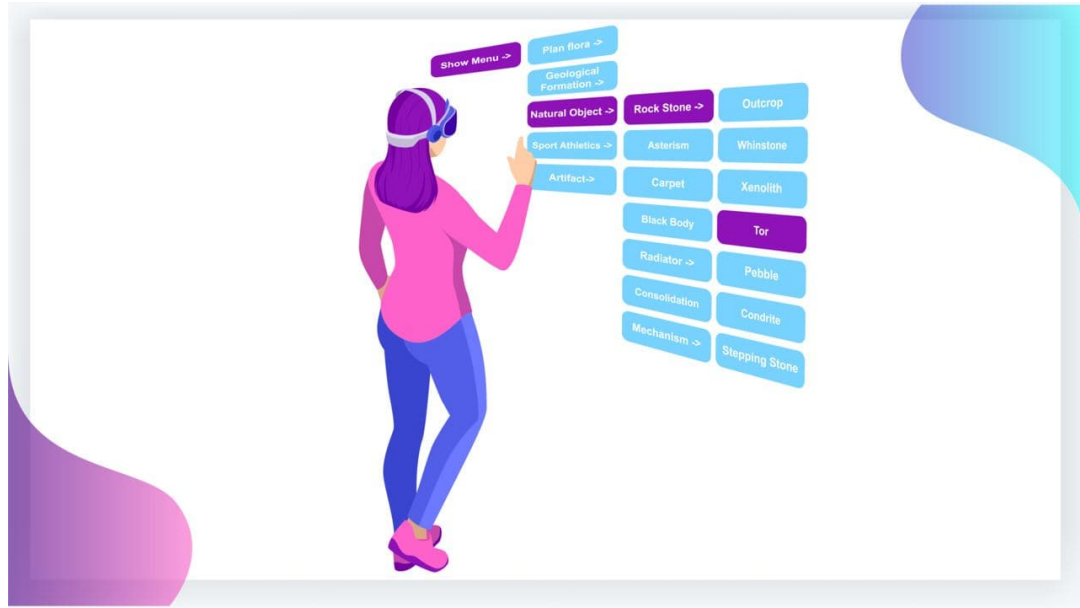


Figure 5.1: Hierarchical menus are widely used in various applications. They can be used in immersive AR to present numerous actions and options of the designed software.

5.1.1 User-related factors

The user-related factors depend on various users' abilities to perform menu selection tasks. The value of an individual user-related aspect varies amongst users. The learning effect [71], user fatigue, and cognitive performance of users [91] are examples of user-related factors. In this research, we employ the cognitive test result, AR Experience, and Age of users to calculate the Error Index in each user's menu selection.

5.1.2 GUI-related features

GUI-related factors are the features related to the appearance of the GUI, such as the number of menu items, the number of levels, and how the menus are gathered together (i.e., drop-down, radial, or tile menus). In this research, we consider only menus having drop-down forms and apply the number of elements in each level to calculate the provided index.

5.2 Related Work

Gaze and hand gesture-based interactions have been widely used in recent AR/VR devices to provide a more realistic sense of the immersive environment. Although these interactions seem to be the point of interest for many users [92], there are still many areas to improve to have more user-friendly and optimized interactions. Various works have recently addressed issues raised in gaze-gesture-based interactions in immersive AR. From providing mathematical-based [53] [25] and AI-based [91] predictive models for determining the performance of the users during the immersive experience to the numerous user studies conducted for evaluating different aspects of the immersive experience with gaze and gesture combinations [92], [22], [82].

Below, we review the factors relating to user performance when interacting with menu selection in immersive AR.

5.2.1 Pointing time and task completion time

Pointing time is the time that it takes for the user to select an item from a menu list or hierarchy [98]. Predicting pointing time and task completion time is the point of interest in many pieces of research related to human performance prediction in HCI. Numerous works have been published for predicting pointing time in menu selection in both desktop [71] and immersive AR and VR environments [91]. There are two types of approaches for predicting pointing time in menu selection tasks: 1) Machine learning-based models that are trained on menu selection data and mathematical models that are mostly based on Fitt's law [50] and Hick's law [58] for both 2D and 3D environments.

5.2.2 Physical and mental workload

Interactions in immersive AR include more physical and mental activity than in desktop environments. Therefore, it is important to have tools and methods for estimating the amount of physical and mental workload on the users while wearing the immersive headset. The subjective methods such as NASA TLX [56] are widely used to this end to estimate the mental and physical activity of users. The evaluation is based on the user's rating after each task. However, the limitation of these methods is that they contain a high level of subjectivity and error since they are typically done

through questioning after the activity. Researchers in [59], and [61] provided vision-based methods for calculating the user's arm fatigue during mid-air hand gesture tasks. Both of the works require MS Kinect to track users' hands while performing the tasks which can not be applied in many situations. To resolve this issue, authors in [91] offered an AI-based method for predicting the physical activity of users in immersive AR based on the metric proposed in [59] which is named Consumed Endurance.

Many works have been written about performance metrics such as pointing time and mental/physical workload. To the best of our knowledge, there are no studies evaluating the rate of error in hierarchical menu selection in immersive AR. Therefore, in this research, we focused on providing an index that can estimate the hierarchical menu selection error rate based on the tasks and user-related factors.

5.3 Methodology

In this research, we propose an index for estimating error rates based on GUI-related and user-related factors. The index receives a sequence of hierarchical menu selection tasks and each user's result of the cognitive WAIS test and estimates the number of errors in menu selections for each user. The index was evaluated through a user study with 25 participants. We found a strong correlation between the proposed index and the actual errors in the user study for different users. During the user study, we recorded the user's physical activity (Consumed Endurance), Pointing time, and Error rate for menu selection for each user.

5.3.1 Metrics

When dealing with hierarchical menu selection using hand gestures and head gaze, there are several important metrics to consider while evaluating the performance of the menu selection. In this section, the metrics are explained in depth.

Learning Effect

The learning effect is the ability of the users to learn a user interface after multiple selections from the same hierarchy [71]. Therefore, it will be easier for a user to point to a desired menu item after several selections. To quantify the learning rate in hierarchical menu selection, first, we normalized the hierarchical menu pointing time based on the depth of the menus. For example, if a user selects a menu item in depth 3 in 10 seconds, we divide the number 10 by 3 to get the normal time for the menu selection for that specified task. Figure 5.2 shows the normal pointing times for two users. Secondly, to calculate the learning rate, we multiplied the slope of the trend line, as depicted in Figure 5.2, by the goodness of fit of the regression, which is R^2 value of the trend line. Therefore, as a user learns a menu after completing several tasks, the slope of the trend line will be higher. The multiplication to R^2 aids the metric in avoiding cases with a high number of outliers.

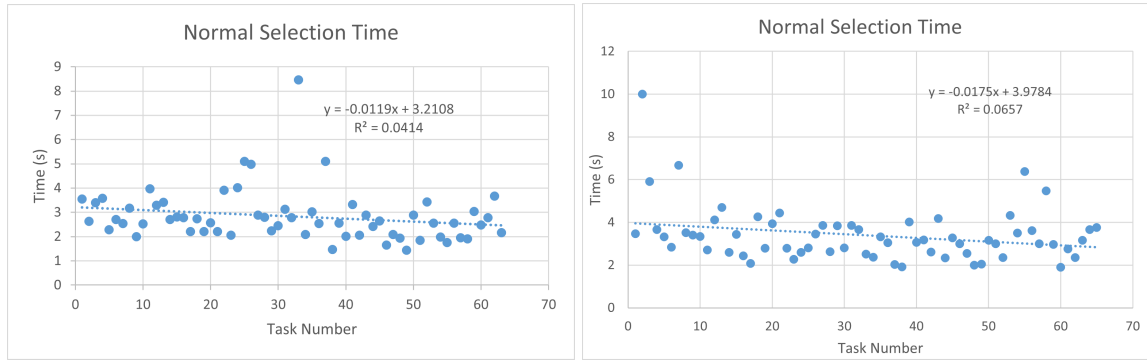


Figure 5.2: Normal pointing time for two users. The learning rate for the user on the left is the multiplication of the trend line's slope and R^2 .

Error Rate

The Error Rate is defined as the number of wrong menu selections for each user in a set of tasks divided by the total number of tasks. Since the immersive AR environment is still unfamiliar to many users, the error rate is high in most cases. As the researchers in [92] reported, 41% of drop-down hierarchical menu selection tasks in immersive AR led to errors in selecting the menu item or recognizing the hand gesture. Consequently, having an index for evaluating the error rate for different users with varying mental performance and prior AR experience will enable designers

and developers of user interfaces to predict how different users will interact with the design.

5.3.2 WAIS IV

The Wechsler Adult Intelligence Scale (WAIS) is an intelligence test used to evaluate the cognitive abilities of adults [105]. It is the most common intelligence test for adults and is often used in clinical, academic, and research settings. The test comprises several sub-tests that each assesses a distinct intelligence component, such as verbal comprehension, perceptual organization, and working memory. The results of the subtests are added together to get a total IQ score, as well as values for verbal and performance IQ. The WAIS is a standardized test; thus, results can be compared to those of other people in the same age group because it was normalized on a sizable representative sample of the population. Before performing the menu selection tasks, the participants were asked to complete a paper-based sub-test of the WAIS-IV. The data adds the user-related aspect of menu selection to the analysis and helps the model discriminate between different users. For our experiments, the Processing Speed Index (PSI) is measured as part of Performance IQ (PIQ). The PSI test consists of two sub-tests: Symbol Search and Symbol Coding. Figure 5.3 shows a sample of Symbol Coding and Symbol Search sub-tests of WAIS test.

5.3.3 The Proposed Index

The index is computed using subjective and objective measures that can be extracted before menu selection based on user- and GUI-related factors from users and GUIs. The user-related factors consist of the user's age, WAIS scores, and prior AR experience. The considered GUI-related is the depth and number of menus in each menu level.

- **Age (AG):** According to observations made during user studies, the age of participants is a significant factor in the error rate for hierarchical menu choices. The age of the participants was used to standardize the WAIS test results, which is used to calculate the error index.
- **Menu Complexity Factor (MF):** The factor is calculated according to Equation 3. It determines the average difficulty of a user's assigned tasks. If a user has a greater number of tasks with longer menus and deeper hierarchies, the MF for that user is greater. In the equation,

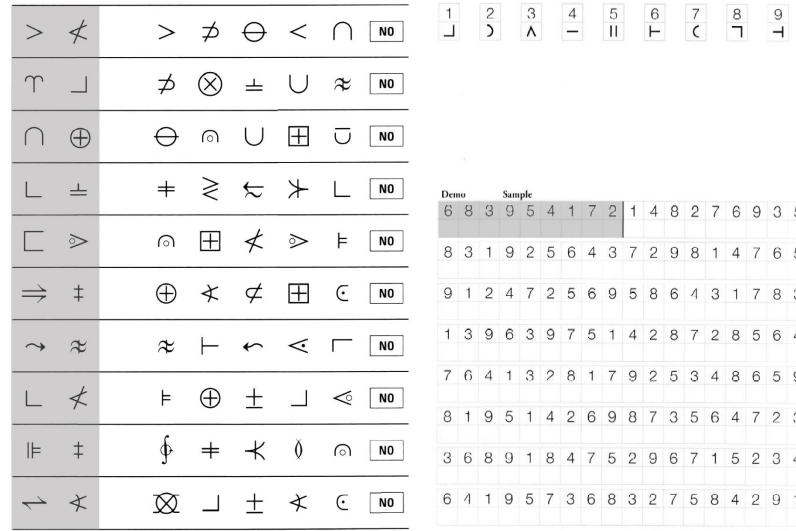


Figure 5.3: Symbol Search Test (left), the examinee had to find the exact symbol that appeared on the left among the right bar's symbols in each line. Symbol Coding Test (right), the examinee was asked to find the symbols associated with each number from the top bar and draw the exact shape under each digit.

N represents the total number of tasks assigned to a user, and l represents the length of each menu list.

$$MF = \frac{\sum_{i=1}^N (\prod_{j=1}^{depth} length_j)_i}{N} \quad (3)$$

- **WAIS Symbol Coding (SC):** The Symbol Coding subtest of the WAIS measures a person's processing speed, attention to detail, and ability to follow instructions. During the test, the individual is presented with a series of symbols and must use a key to quickly and accurately identify and transcribe the symbols. The test is timed, and the score is determined by the number of symbols correctly transcribed within the allotted time (120 seconds). Typically, the test is administered as part of a comprehensive evaluation of cognitive skills. The output will be a positive integer between 0 and 135, inclusive.
- **WAIS Symbol Search (SS):** As another component of the WAIS test, the Symbol Search test is intended to assess an individual's ability to quickly search and identify specific symbols among a group of symbols. During the test, the individual is presented with a series of rows

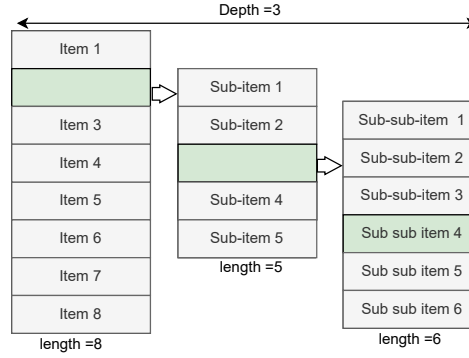


Figure 5.4: Example of a three-level hierarchical menu structure. Using the equation 3, the hierarchical menu task is comprised of the multiplication of 8, 5, and 6. A weighted average of all of their tasks will be determined to calculate the MF for each user.

of symbols and must identify and mark the target symbols as quickly as possible using a key. The test is timed, and the individual's score is determined by the number of symbols they correctly identify within the 120-second time limit. The test result will be a positive integer between 0 and 63.

- **AR Experience (AE):** The metric represents the subjective experience that users have when interacting with immersive AR headsets. The value is a number ranging from 0 to 5. The data was collected during the user study by surveying the participants along a Likert scale.

Finally, the index is calculated based on Equation 4. In the index formula's nominator, MF represents the difficulty of the menu selection. Consequently, as menu complexity increases, the Error index will also rise. SC and SS are the age-normalized Symbol Search and Symbol Coding values for the participants, respectively. The value AE represents the user experience with immersive AR devices.

$$ErrorIndex = \frac{MF + \frac{SC+SS}{AG}}{AE \times 10}, \quad (4)$$

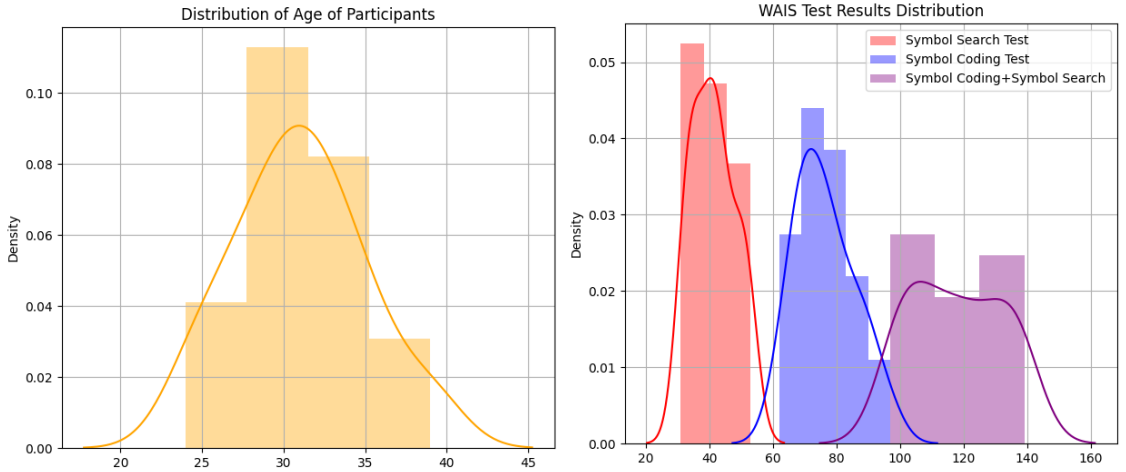


Figure 5.5: Participants' Demographics. (right) The red distribution represents the Symbol Search (SC) test results, while the blue distribution represents the Symbol Coding (SC) test results. The Processing Speed Index of an adult is determined by combining the symbol coding and symbol search tests. (left) The horizontal axis is the age, and the vertical one is the distribution density.

5.4 User Study

The proposed index in Section 5.3.3 is evaluated with the data gathered from a user study with 25 participants. Users are required to complete 70 menu selection tasks (two sets of 35 selection tasks) within an immersive AR application containing hundreds of menu hierarchies. The hierarchical menu selection tasks were generated at random with depths of 2 and 3. The user study setup is described in Figure 5.6. Figure 5.5 shows the participant's age and WAIS test result distributions.

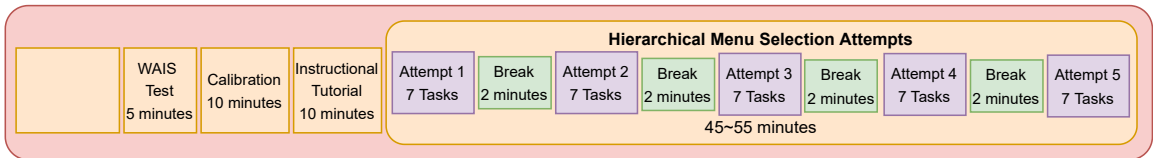


Figure 5.6: The participants were required to complete a pre-test questionnaire, followed by a 5-minute standard WAIS-IV test and a 10-minute headset calibration. As the next step, a 10-minute tutorial was presented to familiarize the participants with hand gestures used in immersive augmented reality, such as air-tapping. The participants were then required to complete the main portion of the test: menu selection tasks. The section contains five menu selection attempts, seven of which are hierarchical. Each attempt was followed by a 2-minute break to prevent user fatigue.

5.4.1 Immersive AR Application

We have designed and developed an immersive augmented reality (AR) application for Microsoft HoloLens I to collect the data necessary for analyzing the error rate of hierarchical menu selection. Additionally, a Microsoft Kinect is used to record the user's activity while performing immersive tasks. The integrated application is created using Microsoft's Mixed Reality Toolkit and Unity3D. (MRTK). The graphical user interface of the application featured multiple hierarchical menu selection tasks with a maximum depth of 3, a common value for hierarchical menus across applications. As depicted in Figure 5.1, users were tasked with completing the user study, which included menu selections. Before performing the menu selection tasks, all users were required to complete a tutorial on MS HoloLens interactions and a pre-test questionnaire about their experience with immersive AR and biographical information.

5.5 Discussion and Analysis of the Results

According to the results of the user study, there is a strong positive correlation (Pearson's Correlation Coefficient = 0.75) between the proposed index value and the real error rate value for the users. This confirms that the Error Index estimated higher values in cases where the user's error rate was higher. Therefore, the proposed Error Index takes the cognitive WAIS test score, age, experience with augmented reality, and the average complexity of the hierarchical menus proposed to the users and provides an estimate of error for the users. Figure 5.7 shows the sorted values of error rate for different users in the horizontal axis and their corresponding values of the proposed index in the vertical line. The trend line also shows the increasing values of the index based on the increasing values of the real error rate.

In addition to the high correlation between the index and the real error rates, Table 5.1 shows the Pearson correlations between the learning rate calculated as it is explained in Section 5.3.1 and the user-related and GUI-related factors involved in the proposed index. We have found a strong correlation between the learning rate and Symbol Search test results as well as the Menu complexity factor.

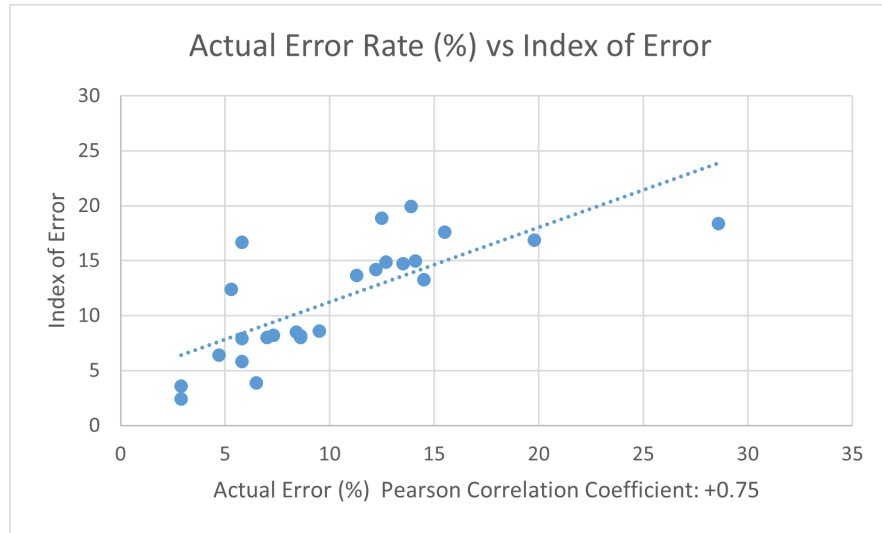


Figure 5.7: The index of error, vertical, and the real values of error for different people, horizontal, sorted increasingly.

Table 5.1: Pearson Coefficient correlation analysis describing the correlations between Learning and Error Rate with the user-related and GUI-related factors.

	Age	Symbol Search	Symbol Coding	AR Experience	Menu Complexity
Learning Rate	0.168	-0.38	-0.13	0.11	0.47
Error Rate	-0.223	-0.220	-0.29	-0.63	0.2

5.6 Conclusion and future works

Many software projects, from simple websites to complex 3D authoring applications with hundreds of menus and hierarchies, make extensive use of hierarchical menus. With the recent emergence of immersive AR headsets and the immense interest in projects like the Metaverse, it is also essential to have these menus in immersive environments. Recent research indicates that because hierarchical menu selection in immersive AR is based on input modalities such as hand gestures, head gaze, and other body movements, the possibility of error is greater. In this study, we sought to develop an index for estimating the error rate for a series of hierarchical menu selection tasks in immersive augmented reality. Since having the error can be related to both the User's capabilities and the structure of the hierarchical menus presented to the user, we have included both characteristics

in our index. A user study involving 25 participants revealed that the index could follow the trend of the real error rate for users with varying cognitive abilities.

For future work, we would like to add support for additional menu types, such as radial and tile menus, and validate our index based on user studies that include radial/tile menu-related tasks.

Acknowledgements

This research is supported in part by the Natural Sciences and Engineering Research Council of Canada Grants DG-N01670 (Discovery Grant).

Chapter 6

Analysis of Hand Movement and Head Orientation in Hierarchical Menu Selection in Immersive AR

This chapter is the paper published and orally presented at the 25th IEEE International Symposium on Multimedia (ISM) 2023, Laguna Hills, California, USA [94].

Abstract

Immersive Augmented Reality (AR) has become pervasive across multiple domains, ranging from medicine and education to interior design, and other fields. Leading technology giants like Apple, Meta, and Microsoft have introduced their proprietary Head Mounted Displays (HMD), joining the competitive market. This technology offers users an immersive experience by superimposing virtual 3D content onto real-world objects, allowing for various interactions and input modalities. Numerous Graphical User Interfaces (GUIs) have also been devised for immersive AR, integrating with input modalities such as hand gesture, head pointing, and voice commands. Since the most favored interaction method in immersive AR involves the interplay between hand gesture and head pointing, our research aims to analyze the workload exerted on users during hand gesture-head pointing interactions. For instance, a hierarchical menu selection task relies on both hand

gesture (for item selection or click) and head pointing (to navigate the cursor). The degree of reliance on each of these input modalities affects the user's workload. We applied both objective and subjective means to assess the physical workload. Two user studies were conducted, from which we derived a unified metric to evaluate the interdependence of hand and head in immersive AR. A significant correlation was discovered between the metric and both objective and subjective workload assessments. Our findings provide insights for AR designers, suggesting the combined use of head pointing and hand gesture when creating hierarchical menu selection interfaces in immersive AR.

Keywords: Immersive AR, Hand Gesture, Head Pointing, Menu Selection

6.1 Introduction

An Immersive AR system is defined as one that blends real and virtual objects within a real environment, provides real-time interactivity, ensures the registration and alignment of real and virtual objects, and displays the augmented content on a mobile device [42]. The immersive AR system must understand the environment's geometry, including surfaces, walls, and object locations. Typically, AR applications utilize a pre-processing scanning method to capture the environment's structure and store a 3D reconstruction model on the device. Once this 3D map is loaded, the application can position virtual objects on real surfaces using computer vision and machine learning algorithms.

As it is mentioned in the definition being interactive is one of the main features of AR applications. AR interactions can be categorized into three types: selection, manipulation, and navigation [29]. Selection involves identifying a virtual 3D model for further actions, consisting of pointing to change cursor location through methods like head pointing, eye tracking, and hand gestures, and then picking the desired object using methods such as clickers [57], hand gestures, or dwell-time [73]. Manipulation entails altering the features of the selected 3D model using input modalities like clickers or gestures [100] and navigation is for moving across virtual and real contents.

Recently, advances in computer vision, sensors, and optics have facilitated the emergence of a range of immersive HMDs designed to provide users with an unparalleled immersive experience. Notable examples of these advancements include the MS HoloLens II [8], Meta Quest Pro [7],

Magic Leap II [5], and the recent Apple Vision Pro [1]. These HMDs allow users to seamlessly engage with virtual content through a variety of interaction techniques such as hand gestures, head pointing, voice commands, and eye tracking. The richness of these interaction techniques enhances users' sense of presence, fostering a more realistic experience within these immersive environments [18]. Notably, the integration of hand gesture and head pointing has garnered significant attention, becoming a preferred interaction technique among leading AR/VR HMD manufacturers.

Prominent producers such as Microsoft, Meta, Apple, and Magicleap have all emphasized the use of this combined interaction approach in their designs [51]. For all interactions within immersive AR, these procedures demonstrated a combination of gaze and gesture to assist users in achieving tasks such as menu selection or object manipulation. For instance, HoloLens II employs gaze for cursor navigation and hand gestures for selecting and interacting with virtual content. Conversely, Meta prioritizes hand gestures for cursor navigation in the environment, with gaze playing a more secondary role in control. Given the importance of coordinating hand gestures with head pointing in immersive interactions, this paper presents an evaluation of hand movement data procured from an immersive experience user study. This is juxtaposed with the head orientation data obtained during immersive experience tasks performed in an immersive AR setting. Since hierarchical menus are widely used in all computer applications, the focus of immersive tasks in this paper is on hierarchical menu selections.

Hierarchical menus are prevalent in numerous computer applications, highlighting their importance for UI/UX designers and developers. This is particularly relevant when designing actions and items for immersive AR applications [92].

Recognizing that the specific hierarchical menu type plays a pivotal role in menu selection performance, this study employs two established graphical user interfaces: drop-down menus and radial menus. We offer a comprehensive analysis of hand movements and head pointing patterns observed when users navigate and select items within these menu hierarchies. In this context, the "hand path" is delineated as the trajectory a user's hand follows when navigating from the initial to the final menu item. Similarly, "head pointing displacement" is conceptualized as the cumulative orientation changes of a user's head while executing a menu selection task.

To enable comparison across varying interaction settings, we introduced a ratio incorporating

both "hand path" and "head pointing displacement". The values of the proposed ratio may vary based on the extent to which an interaction relies on hand gestures or head pointing. The results from our user studies demonstrate a significant relationship between our proposed metric for hand and head movements during hierarchical menu selection. These findings indicate that as an interaction increasingly employs head pointing for menu selection, there's a corresponding reduction in reported and measured workload. While the study confirms a preference for controlling the cursor via head movements for menu selection, subsequent findings suggest that relying solely on head pointing can result in increased errors during menu item selection. Figure 6.1 shows the user while performing menu selection tasks.



Figure 6.1: The figure depicts a user executing menu selection tasks in immersive AR. The experimental setup incorporated a tutorial to acquaint participants with the hand gestures utilized in the MS HoloLens. We conducted two separate experiments, with distinct participants for each. Our user study was announced in the university through email notifications sent to students.

6.2 Related Work

In this section, we survey recent studies pertinent to interactions, input modalities, and ongoing research regarding hand gestures and head pointing in immersive AR. Concerning the evaluation of interaction techniques within immersive AR, numerous studies, such as [92] and [68], have leveraged statistical methodologies like ANOVA and ANCOVA to elucidate differences in efficiency, performance, and other quantitative metrics across various interaction types and input modalities. The primary aim of these investigations is to conduct user experiments with a specified number of participants, who are directed to execute designated tasks. Throughout these experiments, a plethora of data is amassed, ranging from subjective measures such as NASA TLX [56] for workload assessment, to more objective metrics like task completion time, cursor's path, and Consumed Endurance [59]. Another strand of research is dedicated to devising predictive models, predominantly utilizing machine learning and deep neural networks to estimate various metrics like human performance metrics, task completion time, and workload indices. Insights from these research avenues are instrumental for HMD manufacturers and immersive AR developers in gauging the final performance of their designs. Exemplars of such research are delineated in works like [91], which predicts human performance in hierarchical menu selection, and [93], which offers an error index for hierarchical menu selection.

6.2.1 Hand gesture and head pointing in immersive AR

In HMDs, the user's head orientation is continually tracked using built-in sensors. The selection of a virtual or real object is facilitated by a ray cast from the user's head, based on the specified orientation. The point where the ray intersects with the object is termed the cursor. Kyto et al. [68] highlighted that head pointing offers superior targeting accuracy compared to eye tracking or gesture, although it is slower. Blattgerste et al. [27] juxtaposed the head-gaze and eye-gaze methods for dwell-time-based UI interactions in AR. For this, the HMD was equipped with an eye-tracking device. Their findings suggest that eye-gaze surpassed head gaze in terms of speed, task load, and error ratio. Replacing traditional desktop interaction methods, such as controllers, in immersive AR direct manipulation of virtual objects using hand gesture is proposed for more natural

interactions which were originally glove-based [44]. Later on, advances in computer vision led to the creation of the Leap Motion hand tracker [4], which could be placed in front of the user [64] or atop the HMD [88]. MS Kinect and HoloLens were the next devices to support vision-based hand gesture employing an array of cameras. While hand gesture in immersive AR effectively increase the DOF, sustained hand elevation can induce Gorilla Arm, a phenomenon denoting user fatigue. Addressing arm fatigue remains a pertinent research topic within mid-air hand gesture. Hincapie et al. [59] introduced CE to characterize arm fatigue during AR hand gesture, eliminating the need for subjective user feedback.

6.2.2 Coordination of hand gesture and head pointing

The simultaneous use of hand gesture and head pointing presents a multidimensional interaction paradigm in AR. When these modalities are synergized, users can potentially achieve more complex tasks. For instance, a user might employ head pointing for selection and use hand gesture for manipulations like rotation or scaling. The interplay between these two methods, however, requires careful design considerations. Improper coordination lead to increased workload. On the other hand, sole reliance on head pointing can sometimes lead to inaccuracies [92], especially when precise selections are needed. Furthermore, continuous head movements could potentially strain users over time [77]. Authors in [74] delve into the intricate relationship between eye and hand coordination, emphasizing the potential of Gaze-Hand alignment in AR. Highlighting gaze as a precursor to manual action, the authors introduce the innovative concepts of Gaze&Finger and Gaze&Hand techniques, demonstrating their efficacy in AR input. A comparative evaluation showed that gaze-assisted methods surpassed hands-only input. The introduced Gaze-Hand alignment framework enriches the design paradigm for AR, especially when considering inputs beyond manual reach. The study provides valuable insights into challenges and design trade-offs in the realm of AR.

In summary, the imprudent fusion of hand gesture and head pointing can amplify user workload in immersive AR settings. In this research, our objective is to establish a metric for appraising the synergy of hand gesture and head pointing in immersive AR during menu selection activities.

6.3 Application

After receiving the ethics approval, we carried out two user studies on menu selection in immersive AR. Participants were tasked with menu selection activities. Throughout these tasks, we monitored users' hand movements and head orientations utilizing both HoloLens and Kinect. Additionally, we collected users' subjective feedback on the physical workload for each task. The software was developed for the MS HoloLens I, leveraging Unity3D and the MRTK library. Figure 6.2 shows the user while performing the tasks.

User Study 1

A 3D authoring application was developed, including 20 menu items on a 3D object. Users navigated through four combinations of menu types and input modalities, including (drop-down menu, hand gesture), (radial menu, hand gesture), (drop-down menu, voice control), and (radial menu, head pointing). Tasks such as rotation, scaling, movement, and color change were sequentially presented with precise instructions to mitigate bias. Users completed a total of 16 tasks across these combinations. The drop-down menu, a multi-layered 2D plane within the AR environment, facilitated hand gesture and voice control. Conversely, the radial menu employed concentric circles for options, supporting selections via a center-originated line controlled by head pointing or hand gesture. A pilot study revealed that the application initially presented a steep learning curve for new AR users, especially concerning hand gesture. However, an introductory tutorial on gestures and HMD usage significantly mitigated this learning curve. Among the 25 participants in the study, 32% were female and 68% were male. In terms of age distribution, 52% belonged to the 25-34 age bracket, while the 18-24 and 35-44 age brackets each comprised 24% of the total. Users in this study were provided with different configurations exhibiting varying levels of hand gesture and head-pointing engagement. The collected data was later utilized to evaluate the correlation between the levels of hand gesture and head-pointing with the physical workload and error rate experienced by users.

User Study 2

An AR application was developed with the integration of MS Kinect I. Interaction was facilitated through hand gesture (for menu selection) and head pointing (for cursor movement) coupled with a hierarchical drop-down menu. In the study, a total of 26 participants (12 females, 14 males) were engaged, with ages ranging from 24 to 39 years. A significant 73% of the participants were aged between 27 and 35 years, 11% fell within the 35 to 39 years, and 16% were aged between 23 and 27. Users were presented with a total of five attempts, each containing seven menu selection tasks. To ensure consistent conditions for all participants across both user studies, we implemented a calibration process for the MS HoloLens and the user's gaze for each individual. In Figure 6.2, the red dots depicted on the ground and on the wall serve to calibrate the users' viewpoint at the initial stage. The MS HoloLens has a personalized calibration procedure for each user, necessitating the execution of certain hand gesture tasks before interaction with the device.

6.4 Methodology

In user study 1, we collected a dataset comprising menu selections from various participants, along with their subjective ratings of the workload required to perform these selections. In user study 2, we employed the objective method known as Consumed Endurance [59] to document users' workloads during menu selection tasks. Our primary aim is to evaluate the displacements of the hand and head and their relation with the subjective rating of users as well as the objective method for evaluating workload in immersive AR. To this end, we have introduced two metrics to elucidate the relationship between menu selection and the movements of the hand and orientations of the head. These metrics are discussed in detail in Section 6.4.1. Figure 6.2 shows the user while performing menu selection tasks in the user study 2.

6.4.1 Metrics

We introduced a ratio to determine the dependency of each task on hand and head displacements by accumulating the values of these displacements. For hand gestures, the position changes during task execution, while for the head, the user's head orientation serves as the variable parameter.

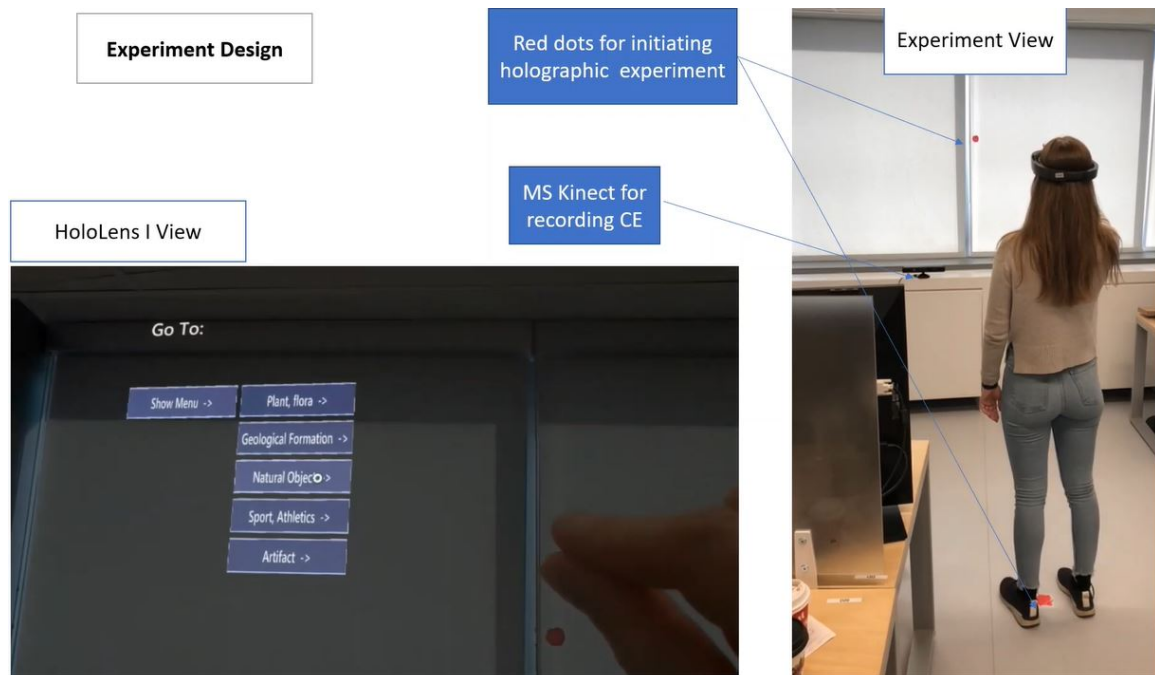


Figure 6.2: Participant executing menu selection tasks as the MS Kinect captures hand movements. Red dots on the floor and wall ensure a consistent starting configuration setup for all users.

Additionally, this section elaborates on both subjective and objective methods employed to assess the user's workload.

Hand-Head usage Ratio (HHR)

To assess the effort exerted through hand movements and head orientations during menu selections, we introduced two metrics representing their cumulative values. The AHM metric denotes Accumulated Hand Movements, while AHO signifies Accumulated Head Orientations. Essentially, these metrics quantify the effort involved with hand and head movements to achieve a menu selection. While the standard recording interval for these values is 10 ms, it can vary based on the device. Figure 6.3 illustrates users' menu selection tasks and the trajectories of the hand and head during these tasks. To better understand the relationship between hand movements and head orientations in a task, we introduced a new metric: the ratio of AHM to AHO. This metric sheds light on the reliance of a task on hand and head movements. As certain tasks in immersive AR might not involve hand movements (e.g., voice-only or head-pointing-only interactions) but always require head orientation, we positioned the AHM value in the numerator and the AHO value in the denominator

of the ratio, as presented in Equation 7. To avoid a zero in the numerator (no hand interactions), we incremented the numerator value by 1.

$$AHM = \sum_{t=0}^T (x_t - x_{t-1})^2 + (y_t - y_{t-1})^2 + (z_t - z_{t-1})^2 \quad (5)$$

$$AHO = \sum_{t=0}^T \sin^2(\phi_t - \phi_{t-1}) + \sin^2(\theta_t - \theta_{t-1}) + \sin^2(\psi_t - \psi_{t-1}) \quad (6)$$

$$HHR = \frac{1 + AHM}{AHO} \times 100 \quad (7)$$

Consumed Endurance (CE)

In the study referenced by [59], a metric was introduced to assess the workload of tasks involving mid-air hand gestures, particularly for AR/VR environments that use such gestures as input. This evaluation tool employs a Microsoft Kinect to monitor users' hand movements in real-time, attributing a workload score to 3D manipulative tasks. The underlying model is grounded in the biomechanical properties of the upper arm for both genders and is a popular metric for examining the Gorilla Arm phenomenon.

Subjective Physical Demand (SPD)

After each user study, participants were requested to complete a questionnaire assessing the physical workload associated with various input modalities and interaction techniques. They rated their experiences on a Likert scale ranging from 1 to 5. We subsequently utilized these scores to corroborate the objective methods used for workload calculation

Error Rate (ER)

As immersive AR is a novel concept for many users and the chance of wrong menu selection is high, it is crucial to ascertain the number of incorrect selections. To elucidate the relationship between incorrect menu selections and the proposed ratio, we recorded the instances of incorrect

menu selections during tasks done by users. This value was subsequently normalized by the total number of tasks completed.

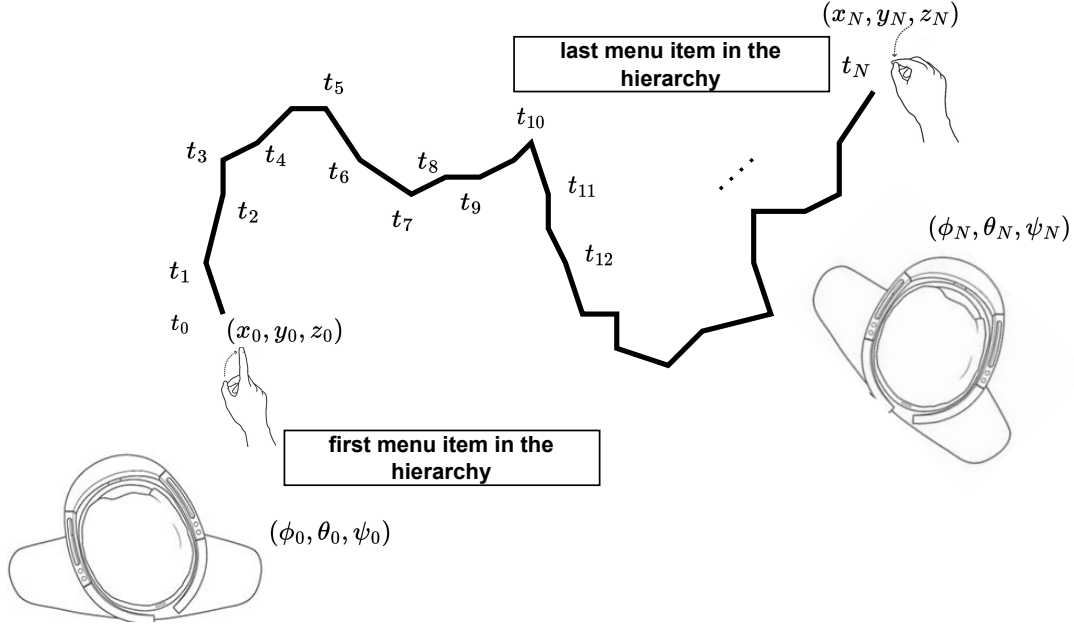


Figure 6.3: The figure depicts a user wearing the HMD and performing menu selection tasks. Using the MS HoloLens’s built-in procedure, we capture the user’s preferred hand positions and head orientations every 10 ms. The coordinate origin is set at the user’s initial head position. Data collection starts with the user’s first menu item selection and concludes upon selecting the final item in the hierarchy.

6.4.2 Analysis

In order to analyze the data and find the relation between the proposed ratio and the workload and error rate we used the Pearson correlation analysis. The Pearson Correlation Coefficient (PCC), denoted as r , is a statistical measure that computes the strength and direction of the linear relationship between two continuous variables. The coefficient values range from -1 to 1, wherein a value closer to 1 implies a strong positive linear relationship, a value closer to -1 indicates a strong negative linear relationship, and a value around 0 suggests no linear relationship.

6.5 Evaluation

In Table 6.1, the average values of AHM, AHO, and the ratio are provided, alongside the average subjective physical demand reported by users during user study 1, as well as the recorded error rate observed while performing menu selections. Each row delineates the values for a specific menu type and input modality pair. The averages were computed to obtain the overall values of hand movements and head orientations for each criterion.

Table 6.2 presents the average values of the proposed formulas and the ratio and the average Consumed Endurance recorded by MS Kinect during each menu selection task in user study 2. Given that in user study 2, each participant had five attempts, the average values of AHM and AHO were computed for each attempt across all users.

Table 6.3 presents the calculated Pearson correlations between the proposed ratio and several variables across two user studies. In user study 1, the correlations with error rate and subjective workload are shown, whereas in user study 2, the correlation with error rate and an objective workload analysis are displayed.

From the outcomes of user study 1, it was observed that (GUI, Input Modality) pairs requiring minimal hand interaction resulted in reduced subjective physical demand, as indicated by the average user ratings. This suggests that tasks with lesser hand involvement are less tiring for users to execute. The dependency ratio of hand and head for these interaction techniques converged to a value close to $(1/AHO) \times 100$. This is because hand involvement in interactions that are controlled only by voice recognition and head pointing is zero and only head orientations will be recorded in the experiments.

Another insight from the study highlighted that the proposed pair (radial, hand gesture), wherein the cursor was controlled by the user's hand, resulted in a higher value of the aforementioned ratio. This, in turn, corresponded to a heightened subjective physical demand as rated by participants. Additionally, the study revealed that greater values of this ratio could lead to increased errors in menu selection tasks.

To validate the findings from user study 1, which was subjective, we analyzed data from user study 2 and calculated the proposed ratio for each task, as outlined in Equation 7. Since the tasks

Table 6.1: The average values of Accumulated Hand Movements (AHM, m^2), Accumulated Head Orientations (AHO, rad^2), Hand-Head usage Ratio (HHR), Subjective Physical Demand (SPD) and Error Rate (ER). Each row represents a combination of (GUI, and input modality) in user study 1.

(Menu, Input modality)	AHM	AHO	HHR	SPD	ER %
(drop-down, hand gestures)	0.62	0.58	279	3.2	41
(radial, hand gestures)	0.51	0.37	408	4.2	46
(drop-down, voice)	0	0.53	180	1.2	13
(radial, head pointing)	0	0.39	250	1.5	10

in user study 2 involved a combination of hand gestures and head pointing, the initial ratio values were high at the task's onset. This can be predicted by the findings of the user study one as the outcome was higher values of ratio for the tasks that had both hand and head involvements. However, subsequent analysis of our user study 2 results indicated a consistent decline in the aforementioned ratio over time. This trend corroborates the findings of the study by Pourmemar et al. [91], which identified the user's learning effect as a significant factor in predicting task workload across consecutive menu selection experiments in immersive AR. The learning effect refers to users progressively familiarizing themselves with menu items and interactions over time. Similar observations about this effect have been made in prior studies, such as Li et al. [71] in menu selection for desktop environments.

The average values of AHM, AHO, and HHR, alongside objective workload evaluations using Consumed Endurance, are presented in Table 6.2. Regarding the Error Rate in menu selections made by participants in user study 2, the data indicates that higher values of the proposed ratio correspond to higher incorrect menu selections in immersive AR. This relationship is evident when comparing other factors such as Menu Type and Input Modality, confirming that increased engagement of both the hand and head results in higher errors. This observation aligns with feedback from participants in the post-test of user study 1, who reported that coordinating between both hand and head while doing the task can increase the workload on the user and lead to higher probable error.

Table 6.2: The average values of Accumulated Hand Movements (AHM, m^2), Accumulated Head Orientations (AHO, rad^2), Hand-Head usage Ratio (HHR), Consumed Endurance (CE) and Error Rate (ER) from tasks in user study 2. Participants undertook menu selections comprising five attempts, each containing seven menu tasks. Each line of the table presents the results of one attempt.

Attempt	AHM	AHO	HHR	CE	ER%
1	0.83	0.72	254	12.6	30
2	0.72	0.69	249	12.3	26
3	0.68	0.71	236	10.2	22
4	0.65	0.72	229	9.8	21
5	0.64	0.70	234	8.6	21

Table 6.3: The values of this table are provided based on Pearson correlation analysis extracted from the values in Table 6.1 and Table 6.2. The workload in user study 1 is Subjective Physical Demand (SPD) and in user study 2 is Consumed Endurance (CE).

	HHR, workload	HHR, ER
User Study 1	0.97	0.81
User Study 2	0.91	0.96

6.6 Discussion

The findings from our user studies highlight a prevailing discussion among immersive AR HMD developers, including industry leaders like Microsoft, Meta, and Apple. This debate focuses on determining key factors in the design and development of immersive HMDs, specifically regarding interaction techniques and input modalities. Hand movements and head pointing, the two primary input modalities across most immersive HMDs, are central to this discussion. For instance, Microsoft relies on head pointing for cursor control in routine interactions, whereas devices like the Meta Quest Pro use hand movements for navigating cursors within 3D user environments. While both hand-controlled and head-controlled cursor modalities have their respective advantages and drawbacks, our studies indicate a higher user preference and less workload for head-controlled cursors in immersive AR. However, our findings also show an elevated error rate associated with interactions dominated by hand control. The proposed ratio, detailing the dependency on hand and head interactions, offers valuable insights for AR HMD developers striving to optimize user workload

and physical demands in immersive AR tasks. These observations align with challenges identified in prior studies, such as those cited in [92] and [100].

6.7 Conclusion and Future Work

A significant challenge in immersive AR is the high workload because of gestural interactions. Since users and HMD manufacturers prefer hand gesture and head-pointing, this study aims to assess the combination of these modalities for menu selection—a ubiquitous task across computer applications. As the current literature shows a pronounced preference for hand-controlled and head-controlled cursors in immersive AR, we proposed a ratio to assess the relative reliance on hand and head displacements through user studies. Our results show that a higher reliance on hand movements escalates user workload during menu selection tasks, an observation verified through both subjective and objective workload assessments. Conversely, a high emphasis on head-pointing was associated with a reduced error rate. The primary insight from this study reveals a significant correlation between our proposed ratio and the error rate and workload. In the ensuing analysis, a noteworthy insight emerged, mirroring the learning effect outlined in [71] observed in user study 2. A substantial relationship was uncovered between the proposed ratio and the learning effect, suggesting that as users become more familiar with the menu, enhanced coordination between hand and head movements is achieved, leading to a decrease in the ratio during tasks. For future work, we aim to integrate additional AR critical components, such as eye-tracking, to assess the outcomes of merging hand gesture with eye-tracking.

Acknowledgements

This research is supported by the Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2021-03479 (Discovery Grant).

Chapter 7

Conclusion and Future Work

Immersive AR is gaining traction across various applications. Major tech giants like Apple, Meta, and Microsoft have released their immersive headsets. While this emerging technology is captivating, it presents several challenges that need thorough evaluation. Interactivity is a primary feature of HMDs, and most interactions occur via natural user interfaces and input modalities, which introduces unique challenges. Literature indicates that one significant challenge in immersive AR is the boosted mental and physical workload on users. In essence, users undertaking AR tasks often experience greater physical demand than in other environments, such as on desktops. Another concern is the increased likelihood of erroneous actions and selections, particularly in menu selection. This thesis primarily centers on hierarchical menu selection—a common task in most computer applications—with the aim of addressing these challenges.

Firstly, we carried out a user study to assess different hierarchical menu types and input modality combinations in immersive AR, aiming to comprehend their performance and associated challenges. This study analyzed the effectiveness of various pairings of hierarchical graphical user interfaces and input techniques for 3D authoring software in AR headsets. Combinations that utilized hand gestures were the least preferred because of their physical demands, underscoring the well-documented 'Gorilla arm' problem. Notably, the radial interface combined with head pointing emerged as the most efficient method. The combination of voice commands with suggested GUIs also appeared very well in terms of workload and accuracy but there was one significant challenge raised by some users and it was the privacy problem that might happen for users while using voice commands i.e.

in a library or a public area.

As the second goal of this thesis, understanding the anticipated performance of a GUI is crucial for designing intuitive UI/UX in immersive AR. In our second research, we conducted a user study involving numerous menu items and hierarchies from WordNet hierarchy, prompting users to complete menu selection tasks. Throughout these tasks, we captured data on both user-related and GUI-related features. User-related features pertained to individual performance variations in executing the tasks. Consequently, participants undertook a cognitive assessment known as WAIS-IV to gauge their cognitive capabilities. On the other hand, GUI-related features focused on the appearance of menu items, encompassing semantic relationships across hierarchy levels. We later employed a machine learning approach, incorporating a semantic encoder and an LSTM neural network, to predict the performance of each task. The insights from this study will equip immersive AR application designers and developers with knowledge of the anticipated efficiency of their designs.

Moreover, considering the importance of the error rate in immersive AR menu selections, we introduced an index in a subsequent study. This index leverages statistical analysis to project the expected error rate for various users, drawing from both user-related and GUI-related factors. Further statistical analysis showed that the index is able to track the trend of error rate in immersive AR menu selection.

Our final study focused on gestural and head-pointing interactions during menu selection in immersive AR. Utilizing data from previous user studies, we derived a ratio that illustrated the relationship between head pointing and hand gestures in menu selection. The analysis revealed that relying more on head pointing and less on hand gestures results in an improved workload and a reduced error rate for hierarchical menu selection in immersive AR.

In conclusion, both user interface and input modalities play pivotal roles in immersive AR. Through our research, we've offered an array of techniques and guidelines for designing user interactions within this immersive setup.

7.1 Limitations

Evaluating user interfaces in Human-Computer Interaction requires to test it with real users. To this end, user studies with a certain number of users will be conducted and performance factors such as time, error rate, and workload will be gathered. Especially in AR applications the users need to have Head Mounted Displays to test the application. The main objective of this research is to find a way to do an evaluation of a proposed user interface using Machine Learning. Therefore, the designers and developers can use our model to create more perfect user interfaces in terms of usability and human performance. Machine learning methods are proper ways in order to find predictive functions. To this end, we need to do a user study with a large number of users, working with menu items in immersive AR. Since immersive HMDs are not common enough among users it is difficult to find users and collect data from them through user studies.

Moreover having various headsets with different configurations is also another challenge in dealing with performance prediction in immersive AR. To avoid this problem, in this research, we only focused on one type of immersive headset. Therefore all the conditions will be similar for all users.

7.2 Future Perspective

Although we predicted the performance of users in immersive AR for drop-down hierarchical menus, still there are challenges in decoding all aspects of GUI and not only the semantics of the menu items. While existing research often harnesses models like BERT [102] to interpret hierarchical menus in immersive AR, there remains a gap in addressing more intricate GUIs, such as tile and radial menus. Moreover, most existing models are semantics-based, emphasizing the textual content of menu items. These models falter when confronted with visually rich menus, underscoring the need for an AI solution capable of decoding user-view scenes into vector formats, which can then guide performance predictions. To bridge this gap, we are currently working on employing the GPT-4 vision API [83], allowing us to interpret user-view screenshots within immersive AR. Alongside this, we assessed user performance metrics, like workload and error rates, from spatial user studies. This data subsequently trained another AI model to predict human performance.

In conclusion, the refined AI model serves as a potent tool to evaluate and anticipate user performance in immersive menu navigation tasks.

Appendix A

A.1 WAIS IV Test

Examinee Name: _____ Age: _____

Examiner Name: _____

Test Date: _____

Symbol Search

Demonstration Items

							NO
							NO
							NO

Sample Items

							NO
							NO
							NO

$>$	$\nless$	$>$	$\nless$	$\ominus$	$<$	$\cup$	NO
$\exists$	$\lrcorner$	$\nless$	$\otimes$	$\pm$	$\cup$	$\approx$	NO
$\cup$	$\oplus$	$\ominus$	$\odot$	$\cup$	$\boxplus$	$\supset$	NO
$\lrcorner$	$\pm$	$\neq$	$\gg$	$\curvearrowright$	$\nless$	$\lrcorner$	NO
$\sqsubset$	$\triangleright\circ$	$\odot$	$\boxplus$	$\nless$	$\triangleright\circ$	$\models$	NO
$\Rightarrow$	$\neq$	$\oplus$	$\nless$	$\nless$	$\boxplus$	$\subset$	NO
$\curvearrowright$	$\approx$	$\approx$	$\top$	$\curvearrowleft$	$\triangleright\circ$	$\lrcorner$	NO
$\lrcorner$	$\nless$	$\models$	$\oplus$	$\pm$	$\lrcorner$	$\triangleright\circ$	NO
$\models$	$\neq$	$\oint$	$\neq$	$\nless$	$\emptyset$	$\odot$	NO
$\uparrow$	$\nless$	$\otimes$	$\lrcorner$	$\pm$	$\nless$	$\subset$	NO

$\vDash$	$\neg$	$\vDash$	$\sim$	$\neg$	$\star$	$\subset$	NO
$\neq$	$\uparrow$	$\sim$	$\neg$	$\times$	$\star$	$\Rightarrow$	NO
$\boxplus$	$\otimes$	$\cup$	$\boxplus$	$\star$	$\neg$	$\approx$	NO
$\Rightarrow$	$\triangleright$	$\approx$	$\emptyset$	$\triangleright$	$\top$	$\times$	NO
$\mathbb{I}$	$\complement$	$\oint$	$\star$	$\complement$	$\uparrow$	$\neg$	NO
$\pm$	$\top$	$\neq$	$\sqcap$	$\top$	$\emptyset$	$\otimes$	NO
$\uparrow$	$\triangleright$	$\times$	$\cup$	$\neg$	$\otimes$	$\complement$	NO
$\odot$	$\sim$	$\Rightarrow$	$\complement$	$\star$	$\cup$	$\oplus$	NO
$\subset$	$\neq$	$\emptyset$	$\star$	$\uparrow$	$\neg$	$\supset$	NO
$\star$	$\triangleright$	$\sim$	$\star$	$\pm$	$\triangleright$	$\approx$	NO

$\approx$	$\sim$	$\nexists$	$\sim$	$\nexists$	$\approx$	$\boxplus$	NO
$\cup$	$\sqcap$	$\nexists$	$\otimes$	$\exists$	$\cup$	$\sqcap$	NO
$\neq$	∇	$\oplus$	$\sqcap$	$\rightarrow$	$\odot$	∇	NO
$\sqcap$	$\otimes$	$\emptyset$	$\sim$	$\sqcap$	$\nexists$	$\nexists$	NO
$\Downarrow$	∇	$\sqcap$	∇	$\sim$	∇	$\sqcap$	NO
$\times$	$\otimes$	$\cup$	$\otimes$	$\nexists$	$\exists$	$\emptyset$	NO
$\downarrow$	$\nexists$	$\perp$	$\nexists$	$\nexists$	$\sqcap$	$\sim$	NO
$\subset$	∇	$\subset$	$\times$	$\subset$	$\sqcap$	ϕ	NO
$\odot$	$\oplus$	∇	$\approx$	$\neq$	$\approx$	$\perp$	NO
$\downarrow$	$\approx$	$\cup$	$\perp$	$\approx$	$\approx$	$\odot$	NO

∇	$\rightarrow$	$\vdash$	$\neq$	$\cup$	$\models$	∇	NO
$\neq$	$\neq$	$\otimes$	$\neq$	$\boxplus$	$\neq$	$\cap$	NO
$\neq$	$\otimes$	$\pm$	$\rightarrow$	$\emptyset$	$\subset$	$\boxplus$	NO
$\perp$	$\neg$	$\neq$	$\approx$	$\neq$	$\neg$	$\Rightarrow$	NO
$\mathbb{I}$	$\models$	$\rightarrow$	$\subset$	$\vdash$	$\otimes$	$\models$	NO
$\Rightarrow$	$\neq$	$\oplus$	$\square$	$\approx$	$\subset$	$\Rightarrow$	NO
$\oplus$	$\rightarrow$	$\neq$	$\neq$	$\boxplus$	∇	$\approx$	NO
$\sim$	$\neq$	$\cap$	$\oplus$	$\rightarrow$	$\sim$	$\square$	NO
$\neg$	$\Rightarrow$	$\neg$	$\neq$	$\subset$	$\Rightarrow$	$\otimes$	NO
$\Rightarrow$	∇	$\neg$	$\oplus$	∇	$\pm$	$\cup$	NO

∇	$\top$	∇	$\top$	$\approx$	$\sqsupset$	$\complement$	NO
∇	$\curvearrowright$	$\curvearrowright$	∇	$\complement$	$\boxplus$	$\approx$	NO
∇	$\bowtie$	$\ominus$	∇	$\mathbb{I}$	$\neq$	$\downarrow$	NO
$\sqsubset$	$\star$	$\star$	$\angle$	$<$	$\star$	$\sqsupset$	NO
$\complement$	$>$	$\boxplus$	$\complement$	$\neq$	$\emptyset$	$\angle$	NO
$\boxplus$	$\approx$	$\cup$	$\ominus$	$\neq$	$\mathbb{I}$	$\boxplus$	NO
$\neq$	$\neq$	$\neq$	$\otimes$	$\complement$	$\star$	$\curvearrowright$	NO
$\emptyset$	$\complement$	$\approx$	$\boxplus$	$\sqsubset$	$\complement$	$\sim$	NO
$\Rightarrow$	$\curvearrowright$	$\curvearrowright$	$\complement$	$\otimes$	$\curvearrowright$	$\sqsupset$	NO
$\ominus$	$>$	$\neq$	$\complement$	$\complement$	$\approx$	$\sqsubset$	NO

$\subset$	$\oplus$	$\otimes$	$\approx$	$\nless$	$\subset$	$\Vdash$	NO
$\cup$	$\curvearrowright$	$\nless$	$\cap$	$\neq$	$\lessgtr$	$\emptyset$	NO
$\nless$	$\subsetneq$	$\perp$	$\nless$	$\otimes$	$\boxplus$	$\lrcorner$	NO
$\neq$	$\nless$	$\subset$	$\emptyset$	$\nless$	$\wr$	$\sim$	NO
$\neq$	$\nless$	$\nless$	$\wr$	$\lessgtr$	$\neq$	$\odot$	NO
$\emptyset$	$\approx$	$\nless$	$\approx$	$\nless$	$\nless$	$\nless$	NO
$\otimes$	$\perp$	$\nless$	$\perp$	$\sim$	$\nless$	$\perp$	NO
$\nless$	$\nless$	$\sqsubset$	$\curvearrowright$	$\nless$	$\emptyset$	$\boxplus$	NO
$\curvearrowright$	$\top$	$\rightarrow$	$\lessgtr$	$\nless$	$\subsetneq$	$\emptyset$	NO
$\sim$	$\wr$	$\neq$	$\sim$	$\sim$	$\nless$	$\supset$	NO

Coding

1	2	3	4	5	6	7	8	9
└	)	^	—		┐	(	└	┐

[illegible]

A.2 Pre-test Questionnaire

Pretest Questionnaire

This form contains personal information and technical background of each subject.

* Required

Untitled Section

1. Participant ID *

2. Age? *

Mark only one oval.

- ☐ 18-24
☐ 25-34
☐ 35-44
☐ 45-54
☐ 55-64
☐ >60

3. To which gender identity do you most identify? *

Mark only one oval.

- ☐ Female
☐ Male
☐ Prefer not to say
☐ Other:

4. Rate your experience using Augmented Reality (AR) devices? *

Mark only one oval.

	1	2	3	4	5	
I have never used an AR device	☐	☐	☐	☐	☐	I have used an AR device and I am comfortable in using this technology

5. Rate your experience using hand gestures for interaction? *

Mark only one oval.

	1	2	3	4	5	
I have no experience using hand gestures for interaction	☐	☐	☐	☐	☐	I am comfortable using hand gestures for interaction

Pretest Questionnaire

6. Rate your experience using 3D authoring software e.g. Maya, 3D Studio Max, Houdini, etc, for creating and manipulating 3D data? *

Mark only one oval.

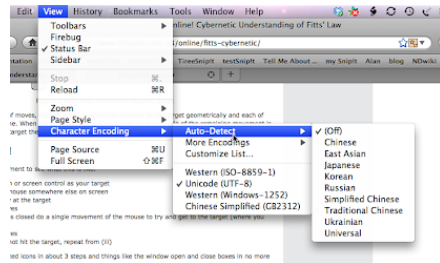
	1	2	3	4	5	
I have never used a 3D authoring software	☐	☐	☐	☐	☐	I am comfortably using a 3D authoring software

7. Rate your experience using voice input for interaction? *

Mark only one oval.

	1	2	3	4	5	
I have no experience using voice input for interaction	☐	☐	☐	☐	☐	I am comfortable using voice input for interaction

8. How difficult do you find using hierarchical (i.e. multi-level) drop-down menus in desktop applications? *



Mark only one oval.

	1	2	3	4	5	
Extremely difficult	☐	☐	☐	☐	☐	Extremely easy

9. Please provide an explanation for your rating in the previous question

This content is neither created nor endorsed by Google.

Google Forms

A.3 Post-test Questionnaire

Post test Questionnaire

This is the post-test questionnaire for the user study EHMAR

* Required

1. Participant ID

Rate your experience

2. Rate the physical demand when using each menu type *

Mark only one oval per row.

	1 - No physical demand	2	3	4	5 - Extremely physically demanding
Drop down	☐	☐	☐	☐	☐
Line Guided Radial (with hand gestures)	☐	☐	☐	☐	☐
Voice	☐	☐	☐	☐	☐
Line Guided Radial (with gaze control)	☐	☐	☐	☐	☐

3. Rate the level of frustration when using each menu type *

Mark only one oval per row.

	1 - No frustration	2	3	4	5 - Extremely frustrating
Drop down	☐	☐	☐	☐	☐
Line Guided Radial Menu (with hand gestures)	☐	☐	☐	☐	☐
Voice	☐	☐	☐	☐	☐
Line Guided Radial Menu (with gaze control)	☐	☐	☐	☐	☐

Post-Test Interview

Post test Questionnaire

4. Do you prefer the menus in AR to have a similar appearance to typical drop-down menus used in desktop applications?

Mark only one oval.

- ☐ Yes
☐ No
☐ Other: _____

5. What is your reason for choosing this type of interaction? (Choose all that apply)

Check all that apply.

- ☐ Easier to use
☐ Faster for object manipulations
☐ Less fatigue

Other: ☐ _____

6. In general, what challenges did you face during the test? (Choose all that apply)

Check all that apply.

- ☐ Difficulty performing hand gestures (e.g. fatigue, gestures not recognized, etc)
☐ Difficulty with voice commands (e.g. not recognized, etc)
☐ Difficulty using gaze as an input method (e.g. turning head, fatigue, etc)

Other: ☐ _____

7. What is your overall rating for drop-down menu using hand gestures? *

Mark only one oval.

	1	2	3	4	5	
1 - Definitely worse than other menus	☐	☐	☐	☐	☐	5 - Definitely better than other menus

8. What is your overall rating for drop-down menu using voice control? *

Mark only one oval.

	1	2	3	4	5	
1 - Definitely worse than other menus	☐	☐	☐	☐	☐	5 - Definitely better than other menus

Post test Questionnaire

9. Would you choose to use voice control in a crowded environment?

Mark only one oval.

☐ Yes

☐ No

10. Explain your answer in the previous question *

11. What is your overall rating for Line Guided Radial Menu using hand gestures? *

Mark only one oval.

1 2 3 4 5

1 - Definitely worse than other menus ☐ ☐ ☐ ☐ ☐ 5 - Definitely better than other menus

12. What is your overall rating for Line Guided Radial Menu using gaze control? *

Mark only one oval.

1 2 3 4 5

1 - Definitely worse than other menus ☐ ☐ ☐ ☐ ☐ 5 - Definitely better than other menus

13. Do you have any comments/suggestions you would like to share about line-guided radial menus in AR?

This content is neither created nor endorsed by Google.

Google Forms

Bibliography

- [1] Apple vision pro. www.apple.com/apple-vision-pro/. Accessed: 2023-07-10.
- [2] Holoanatomy. <https://www.case.edu/holoanatomy/>. Accessed: 2023-07-10.
- [3] Ikea place. <https://itunes.apple.com/ca/app/ikea-place/id1279244498?mt=8>. Accessed: 2023-07-10.
- [4] Leap motion. <https://www.ultraleap.com/>. Accessed: 2023-03-10.
- [5] Magic Leap 2. www.magicleap.com. Accessed: 2023-07-10.
- [6] Magicplan. <https://www.magicplan.app/>. Accessed: 2023-07-10.
- [7] Meta quest pro. www.meta.com/quest/quest-pro. Accessed: 2023-07-10.
- [8] Microsoft hololens. www.microsoft.com/en-us/hololens. Accessed: 2023-07-10.
- [9] Pokémon go. <https://pokemongolive.com/>. Accessed: 2023-07-10.
- [10] Pupil labs eye tracker. <https://www.pupil-labs.com/>. Accessed: 2023-07-10.
- [11] Room plan. <https://developer.apple.com/augmented-reality/roomplan/>. Accessed: 2023-07-10.
- [12] Shopify ar. <https://www.shopify.com/ar>. Accessed: 2023-07-10.
- [13] Spatial mapping. <https://www.stereolabs.com/docs/spatial-mapping/>. Accessed: 2023-07-10.
- [14] Tobii eye tracker. <https://www.tobii.com/>. Accessed: 2023-07-10.

- [15] Unity. <https://unity.com/>. Accessed: 2023-07-10.
- [16] Unreal. <https://www.unrealengine.com/>. Accessed: 2023-07-10.
- [17] D. Ahlström, A. Cockburn, C. Gutwin, and P. Irani. Why it's quick to be square: modelling new and existing hierarchical menu designs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1371–1380. ACM, 2010.
- [18] J. Aliprantis, M. Konstantakis, R. Nikopoulou, P. Mylonas, and G. Caridakis. Natural interaction in augmented reality context. In *VIPERC@ IRCDL*, pages 50–61, 2019.
- [19] T. Althoff, R. W. White, and E. Horvitz. Influence of pokémon go on physical activity: study and implications. *Journal of medical Internet research*, 18(12):e315, 2016.
- [20] T. Azai, S. Ogawa, M. Otsuki, F. Shibata, and A. Kimura. Selection and manipulation methods for a menu widget on the human forearm. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 357–360. ACM, 2017.
- [21] M. Bâce, T. Leppänen, D. G. De Gomez, and A. R. Gomez. ubigaze: ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, page 11. ACM, 2016.
- [22] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE transactions on visualization and computer graphics*, 24(1):457–467, 2018.
- [23] N. Bagher Zadeh Ansari, É. Léger, and M. Kersten-Oertel. Ventroar: an augmented reality platform for ventriculostomy using the microsoft hololens. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(4):1225–1233, 2023.
- [24] G. Bailly, A. Oulasvirta, D. P. Brumby, and A. Howes. Model of visual search and selection time in linear menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3865–3874. ACM, 2014.

- [25] A. U. Batmaz and W. Stuerzlinger. Effective throughput analysis of different task execution strategies for mid-air fitts' tasks in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3939–3947, 2022.
- [26] O. Bimber and R. Raskar. *Spatial augmented reality: merging real and virtual worlds*. CRC press, 2005.
- [27] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, page 1. ACM, 2018.
- [28] R. A. Bolt. "Put-that-there": Voice and gesture at the graphics interface, volume 14. ACM, 1980.
- [29] D. A. Bowman and L. F. Hodges. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing*, 10(1):37–53, 1999.
- [30] D. A. Bowman and C. A. Wingrave. Design and evaluation of menu systems for immersive virtual environments. In *Virtual Reality, 2001. Proceedings. IEEE*, pages 149–156. IEEE, 2001.
- [31] K. Browne and C. Anand. An empirical evaluation of user interfaces for a mobile video game. *Entertainment Computing*, 3(1):1–10, 2012.
- [32] F. Brudy. Interactive menus in augmented reality environments. *Beyond the Desktop*, page 1, 2013.
- [33] M. C. Cabral, C. H. Morimoto, and M. K. Zuffo. On the usability of gesture interfaces in virtual reality environments. In *Proceedings of the 2005 Latin American conference on Human-computer interaction*, pages 100–108. ACM, 2005.
- [34] J. Callahan, D. Hopkins, M. Weiser, and B. Shneiderman. An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 95–100. ACM, 1988.

- [35] C. Cao, Z. Li, P. Zhou, and M. Li. Amateur: Augmented reality based vehicle navigation system. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–24, 2018.
- [36] S. K. Card. *The psychology of human-computer interaction*. Crc Press, 2018.
- [37] S. K. Card, T. P. Moran, and A. Newell. Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive psychology*, 12(1):32–74, 1980.
- [38] S. K. Card, T. P. Moran, and A. Newell. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410, 1980.
- [39] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar, et al. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*, 2018.
- [40] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar, B. Strope, and R. Kurzweil. Universal sentence encoder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 169–174, Brussels, Belgium, Nov. 2018. Association for Computational Linguistics.
- [41] K.-E. Chang, J. Zhang, Y.-S. Huang, T.-C. Liu, and Y.-T. Sung. Applying augmented reality in physical education on motor skills learning. *Interactive Learning Environments*, pages 1–13, 2019.
- [42] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. Mobile augmented reality survey: From where we are to where we go. *IEEE Access*, 5:6917–6950, 2017.
- [43] M. M. Davis, J. L. Gabbard, D. A. Bowman, and D. Gracanin. Depth-based 3d gesture multi-level radial menu for virtual object manipulation. In *Virtual Reality (VR), 2016 IEEE*, pages 169–170. IEEE, 2016.

- [44] K. Dorfmueller-Ulhaas and D. Schmalstieg. Finger tracking for interaction in augmented environments. In *Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on*, pages 55–64. IEEE, 2001.
- [45] A. Easwara Moorthy and K.-P. L. Vu. Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, 31(4):307–335, 2015.
- [46] O. Erazo and J. A. Pino. Predicting task execution time on natural user interfaces based on touchless hand gestures. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 97–109. ACM, 2015.
- [47] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.
- [48] C. Fellbaum. Wordnet. *The encyclopedia of applied linguistics*, 2012.
- [49] L. Findlater, K. Moffatt, J. McGrenere, and J. Dawson. Ephemeral adaptation: The use of gradual onset to improve menu selection performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1655–1664. ACM, 2009.
- [50] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.
- [51] M. Frutos-Pascual, C. Creed, and I. Williams. Head mounted display interaction evaluation: manipulating virtual objects in augmented reality. In *Human-Computer Interaction—INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV 17*, pages 287–308. Springer, 2019.
- [52] R. Galati, M. Simone, G. Barile, R. De Luca, C. Cartanese, and G. Grassi. Experimental setup employed in the operating room based on virtual and mixed reality: analysis of pros and cons in open abdomen surgery. *Journal of healthcare engineering*, 2020, 2020.

- [53] S. Gebhardt, S. Pick, F. Leithold, B. Hentschel, and T. Kuhlen. Extended pie menus for immersive virtual environments. *IEEE transactions on visualization and computer graphics*, 19(4):644–651, 2013.
- [54] D. Gerber and D. Bechmann. Design and evaluation of the ring menu in virtual environments. *Immersive projection technologies*, 2004.
- [55] A. Grinshpoon, S. Sadri, G. J. Loeb, C. Elvezio, and S. K. Feiner. Hands-free interaction for augmented reality in vascular interventions. 2018.
- [56] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.
- [57] J. Hertel, S. Karaosmanoglu, S. Schmidt, J. Bräker, M. Semmann, and F. Steinicke. A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In *2021 IEEE International Symposium on mixed and Augmented Reality (ISMAR)*, pages 431–440. IEEE, 2021.
- [58] W. E. Hick. On the rate of gain of information. *Quarterly Journal of experimental psychology*, 4(1):11–26, 1952.
- [59] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1063–1072. ACM, 2014.
- [60] T. N. Hoang and B. H. Thomas. Augmented reality in-situ 3d model menu for outdoors. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 185–186. IEEE Computer Society, 2008.
- [61] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3328–3339, 2017.

- [62] J. Jankowski and M. Hachet. A survey of interaction techniques for interactive 3d environments. In *Eurographics 2013-STAR*, 2013.
- [63] J. P. Jokinen, Z. Wang, S. Sarcar, A. Oulasvirta, and X. Ren. Adaptive feature guidance: Modelling visual search with graphical layouts. *International Journal of Human-Computer Studies*, 136:102376, 2020.
- [64] W. Jung, W. Cho, H. Kim, and W. Woo. Boosthand: Distance-free object manipulation system with switchable non-linear mapping for augmented reality classrooms. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2017 IEEE International Symposium on*, pages 321–325. IEEE, 2017.
- [65] G. Kurtenbach and W. Buxton. The limits of expert performance using hierarchic marking menus. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 482–487. ACM, 1993.
- [66] G. Kurtenbach and W. Buxton. User learning and performance with marking menus. In *Proceedings of the SIGCHI Conference on Human factors in computing systems*, volume 94, pages 258–264. Citeseer, 1994.
- [67] R. Kurul, M. N. Ögün, A. Neriman Narin, Ş. Avci, and B. Yazgan. An alternative method for anatomy training: Immersive virtual reality. *Anatomical Sciences Education*, 13(5):648–656, 2020.
- [68] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 81. ACM, 2018.
- [69] E. Lank, Y.-C. N. Cheng, and J. Ruiz. Endpoint prediction using motion kinematics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 637–646, 2007.

- [70] É. Léger, J. Reyes, S. Drouin, T. Popa, J. A. Hall, D. L. Collins, and M. Kersten-Oertel. Marin: an open-source mobile augmented reality interactive neuronavigation system. *International journal of computer assisted radiology and surgery*, 15(6):1013–1021, 2020.
- [71] Y. Li, S. Bengio, and G. Bailly. Predicting human performance in vertical menu selection using deep learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 29. ACM, 2018.
- [72] S. Lin, H. F. Cheng, W. Li, Z. Huang, P. Hui, and C. Peylo. Ubii: Physical world interaction through augmented reality. *IEEE Transactions on Mobile Computing*, 16(3):872–885, 2017.
- [73] M. N. Lystbæk, K. Pfeuffer, J. E. S. Grønbæk, and H. Gellersen. Exploring gaze for assisting freehand selection-based text entry in ar. *Proceedings of the ACM on Human-Computer Interaction*, 6(ETRA):1–16, 2022.
- [74] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbæk, and H. Gellersen. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(ETRA):1–18, 2022.
- [75] S. Manjrekar, S. Sandilya, D. Bhosale, S. Kanchi, A. Pitkar, and M. Gondhalekar. Cave: an emerging immersive technology—a review. In *16th international conference on computer modeling and simulation*, pages 131–136. IEEE, 2014.
- [76] S. Marshall. Augmented reality’s application in education and training. *Springer Handbook of Augmented Reality*, pages 335–353, 2023.
- [77] L. M. Masopust, D. Bauer, S. Yao, and K.-L. Ma. A comparison of the fatigue progression of eye-tracked and motion-controlled interaction in immersive space. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 460–469. IEEE, 2021.
- [78] S. Matsui and S. Yamada. Optimizing hierarchical menus by genetic algorithm and simulated annealing. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pages 1587–1594. ACM, 2008.

- [79] G. McLean and A. Wilson. Shopping in the digital world: Examining customer engagement through augmented reality mobile applications. *Computers in Human Behavior*, 101:210–224, 2019.
- [80] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [81] A. Millette and M. J. McGuffin. Dualcad: integrating augmented reality with a desktop gui and smartphone interaction. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on*, pages 21–26. IEEE, 2016.
- [82] P. Monteiro, H. Coelho, G. Gonçalves, M. Melo, and M. Bessa. Comparison of radial and panel menus in virtual reality. *IEEE Access*, 7:116370–116379, 2019.
- [83] OpenAI. Gpt-4: Generative pre-trained transformer 4. <https://www.openai.com/>, 2023. Accessed: 2024-01-22.
- [84] K. Özacar, J. D. Hincapié-Ramos, K. Takashima, and Y. Kitamura. 3d selection techniques for mobile augmented reality head-mounted displays. *Interacting with Computers*, 29(4):579–591, 2016.
- [85] H. M. Park, S. H. Lee, and J. S. Choi. Wearable augmented reality system using gaze interaction. In *Proceedings of the 7th IEEE/ACM international Symposium on Mixed and Augmented Reality*, pages 175–176. IEEE Computer Society, 2008.
- [86] P. T. Pasqual and J. O. Wobbrock. Mouse pointing endpoint prediction using kinematic template matching. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 743–752, 2014.
- [87] N. Pathmanathan, M. Becher, N. Rodrigues, G. Reina, T. Ertl, D. Weiskopf, and M. Sedlmair. Eye vs. head: Comparing gaze methods for interaction in augmented reality. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–5, 2020.

- [88] B. Petry and J. Huber. Towards effective interaction with omnidirectional videos using immersive virtual reality headsets. In *Proceedings of the 6th Augmented Human International Conference*, pages 217–218. ACM, 2015.
- [89] S. Pick, A. S. Puika, and T. W. Kuhlen. Comparison of a speech-based and a pie-menu-based interaction metaphor for application control. In *Virtual Reality (VR), 2017 IEEE*, pages 381–382. IEEE, 2017.
- [90] I. Poupyrev, D. S. Tan, M. Billinghamurst, H. Kato, H. Regenbrecht, and N. Tetsutani. Tiles: A mixed reality authoring interface. In *Interact*, volume 1, pages 334–341, 2001.
- [91] M. Pourmemar, Y. Joshi, and C. Poullis. Predicting human performance in vertical hierarchical menu selection in immersive ar using hand-gesture and head-gaze. In *2022 15th International Conference on Human System Interaction (HSI)*, pages 1–8. IEEE, 2022.
- [92] M. Pourmemar and C. Poullis. Visualizing and interacting with hierarchical menus in immersive augmented reality. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 1–9, 2019.
- [93] M. Pourmemar and C. Poullis. Analysis of error rate in hierarchical menu selection in immersive augmented reality. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) IV*, volume 12449, pages 293–301. SPIE, 2023.
- [94] M. Pourmemar and C. Poullis. Analysis of hand movement and head orientation in hierarchical menu selection in immersive ar. In *2023 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2023.
- [95] P. Pratt, M. Ives, G. Lawton, J. Simmons, N. Radev, L. Spyropoulou, and D. Amiras. Through the hololens™ looking glass: augmented reality for extremity reconstruction surgery using 3d vascular models with perforating vessels. *European radiology experimental*, 2(1):2, 2018.
- [96] R. Reilink, G. de Bruin, M. Franken, M. A. Mariani, S. Misra, and S. Stramigioli. Endoscopic camera control by head movements for thoracic surgery. In *Biomedical Robotics*

- and *Biomechatronics (BioRob)*, 2010 3rd IEEE RAS and EMBS International Conference on, pages 510–515. IEEE, 2010.
- [97] H. Saidi, E. Dubois, and M. Serrano. Holobar: Rapid command execution for head-worn ar exploiting around the field-of-view interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.
 - [98] K. Samp and S. Decker. Supporting menu design with radial layouts. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 155–162. ACM, 2010.
 - [99] M. A. Schnabel, X. Wang, H. Seichter, and T. Kvan. From virtuality to reality and back. *Proceedings of the International Association of Societies of Design Research*, 1:15, 2007.
 - [100] B. Spittle, M. Frutos-Pascual, C. Creed, and I. Williams. A review of interaction techniques for immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
 - [101] A. Swearngin and Y. Li. Modeling mobile interface tappability using crowdsourcing and deep learning. In *Artificial Intelligence for Human Computer Interaction: A Modern Approach*, pages 73–96. Springer, 2021.
 - [102] I. Tenney, D. Das, and E. Pavlick. Bert rediscovers the classical nlp pipeline. *arXiv preprint arXiv:1905.05950*, 2019.
 - [103] J. Vince. *Introduction to virtual reality*. Springer Science & Business Media, 2004.
 - [104] D. Wechsler. The measurement and appraisal of adult intelligence. 1958.
 - [105] D. Wechsler. Wechsler adult intelligence scale–fourth edition (wais–iv). *San Antonio, TX: NCS Pearson*, 22:498, 2008.
 - [106] D. Wei, C. Yang, X. Zhang, and X. Yuan. Predicting mouse click position using long short-term memory model trained by joint loss function. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2021.

- [107] R. Xiao, J. Schwarz, N. Throm, A. D. Wilson, and H. Benko. Mrtouch: Adding touch input to head-mounted mixed reality. *IEEE transactions on visualization and computer graphics*, 24(4):1653–1660, 2018.
- [108] W. Xu, H.-N. Liang, Y. Zhao, T. Zhang, D. Yu, D. Monteiro, and Y. Yue. Ringtext: Dwell-free and hands-free text entry for mobile head-mounted displays using head motions. *IEEE transactions on visualization and computer graphics*, 2019.
- [109] A. Yuan and Y. Li. Modeling human visual search performance on realistic webpages using analytical and deep learning methods. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.
- [110] A. Yuan, K. Pfeuffer, and Y. Li. Human performance modeling with deep learning. In *Artificial Intelligence for Human Computer Interaction: A Modern Approach*, pages 3–31. Springer, 2021.
- [111] Y.-T. Yue, Y.-L. Yang, G. Ren, and W. Wang. Scenectrl: Mixed reality enhancement via efficient scene editing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 427–436. ACM, 2017.