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

Majid Pourmemar[a] and Charalambos Poullis[a]

[a]Immersive and Creative Technologies Lab, Concordia University, Montreal, Canada

## 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

## 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[1] and medicine.[2] 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.[3] 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,[4] neck fatigue,[5] restrictions in hand gesture recognition, and high error rate[6] 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.

Further author information: (Send correspondence to A.A.A.)
A.A.A.: E-mail: mapourmemar@gmil.com,
B.B.A.: E-mail: charalambos@poullis.org

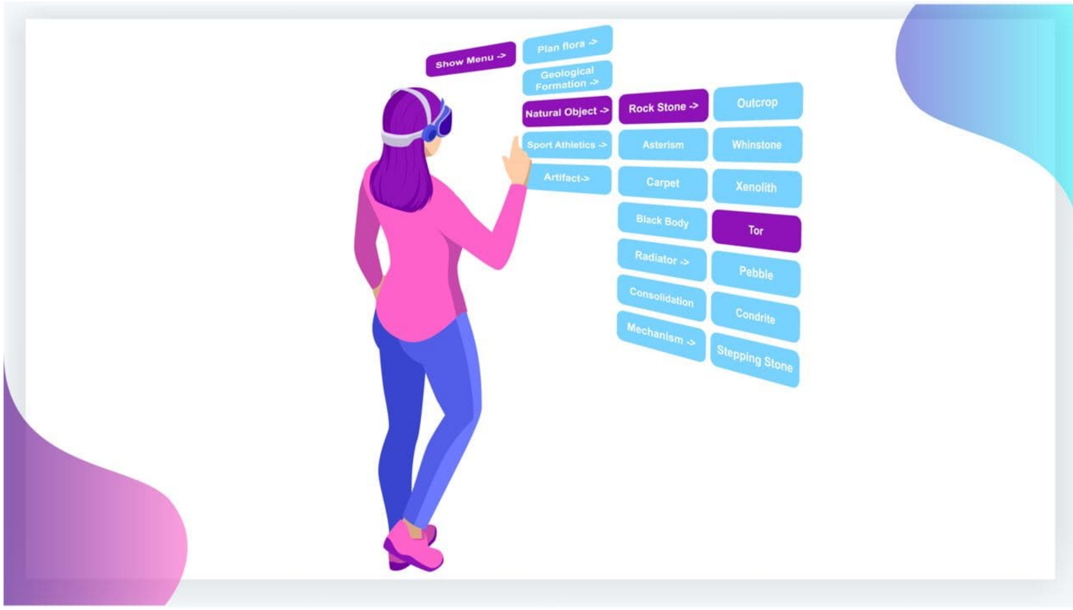


Figure 1: Hierarchical menus are widely used in various applications. They can be used in immersive AR to present numerous actions and options of the designed software.

### 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,[7] user fatigue, and cognitive performance of users[8] 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.

### 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.

## 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,[6] 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[9][10] and AI-based[8] 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,[6],[11].[12]

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

### 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.[13] 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[7] and immersive AR and VR environments.[8] 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[14] and Hick's law[15] for both 2D and 3D environments.

### 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[16] 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,[4] and[17] 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[8] offered an AI-based method for predicting the physical activity of users in immersive AR based on the metric proposed in[4] 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 base on the tasks and user-related factors.

## 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.

### 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.

#### 3.1.1 Learning Effect

The learning effect is the ability of the users to learn a user interface after multiple selections from the same hierarchy.[7] 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 2 shows the normal pointing times for two users. Secondly, to calculate the learning rate, we multiplied the slope of the trend line, as it is depicted in Figure 2, to 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 a number of tasks, the slope of the trend line will be higher. The multiplication to R2 aids the metric in avoiding cases with a high number of outliers.

#### 3.1.2 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[6] 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.

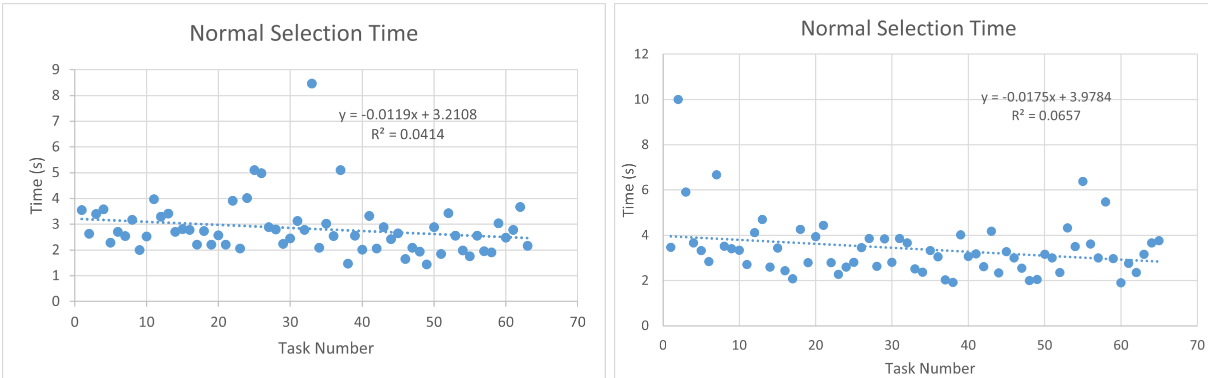


Figure 2: Normal pointing time for two users. The learning rate for the user on the left is the multiplication of the trend line's slope and $R^2$.

### 3.2 WAIS IV

The Wechsler Adult Intelligence Scale (WAIS) is an intelligence test used to evaluate the cognitive abilities of adults.[18] 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 3 shows a sample of Symbol Coding and Symbol Search sub-tests of WAIS test.

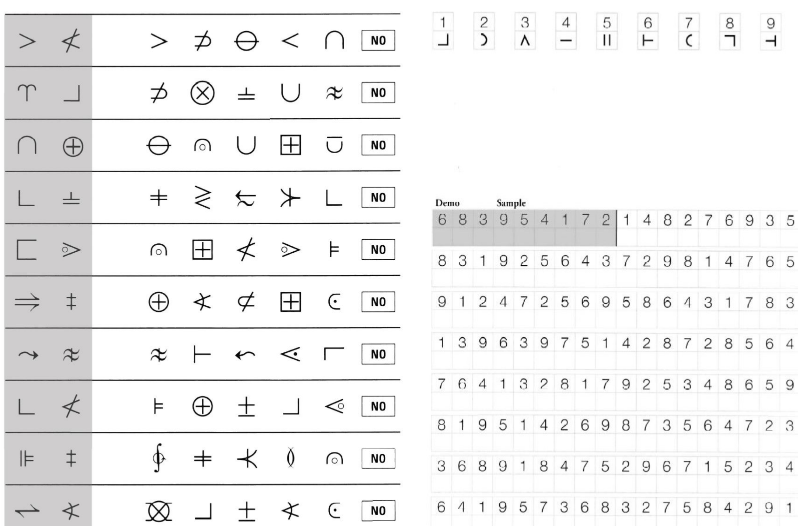


Figure 3: Symbol Search Test (left), the examinee had to find the exact symbol that appeared on the left among the right bar's symbols in each line. Symbol Coding Test (right), the examinee was asked to find the symbols associated with each number from the top bar and draw the exact shape under each digit.

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

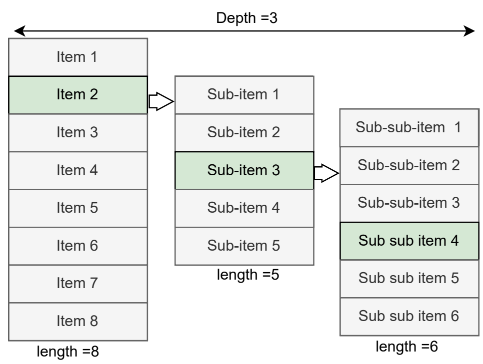


Figure 4: Example of a three-level hierarchical menu structure. Using the equation 1, the hierarchical menu task is comprised of the multiplication of 8, 5, and 6. A weighted average of all of their tasks will be determined to calculate the MF for each user.

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

- **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 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 2. 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.

$$ErrorIndex = \frac{MF + \frac{SC+SS}{AG}}{AE \times 10}, \tag{2}$$

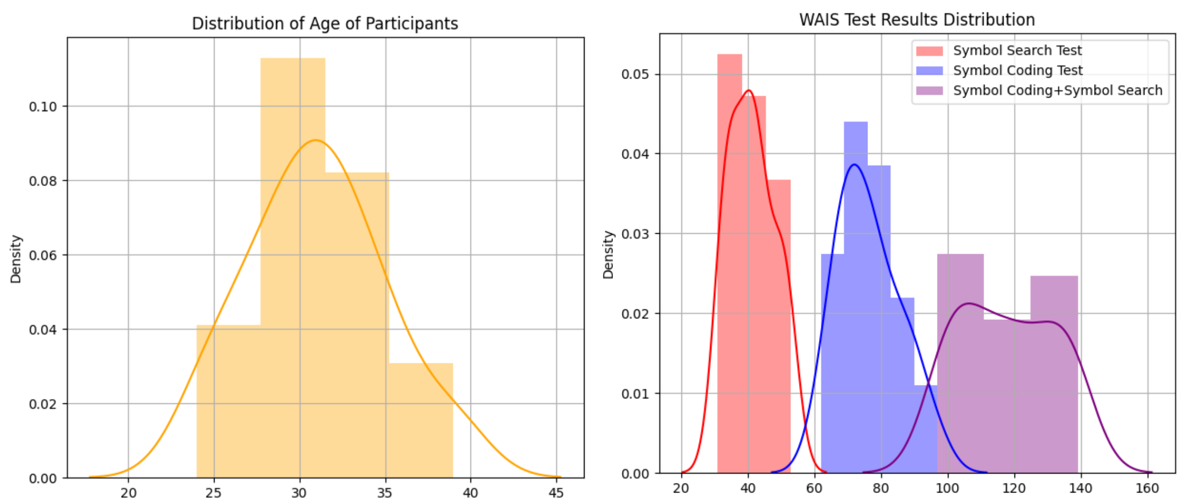


Figure 5: Participants' Demographics. (right) The red distribution represents the Symbol Search (SC) test results, while the blue distribution represents the Symbol Coding (SC) test results. The Processing Speed Index of an adult is determined by combining the symbol coding and symbol search tests. (left) The horizontal axis is the age, and the vertical one is the distribution density.

## 4. USER STUDY

The proposed index in Section 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 6. Figure 5 shows the participant's age and WAIS test result distributions.

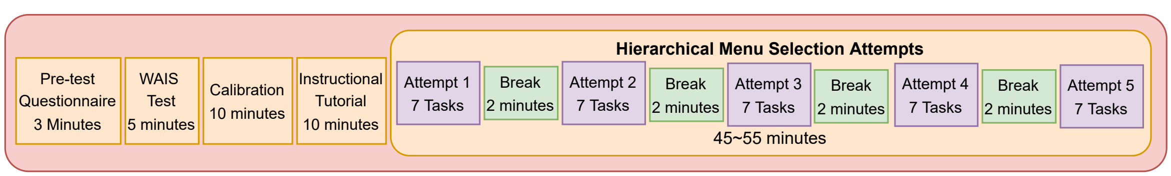


Figure 6: The participants were required to complete a pre-test questionnaire, followed by a 5-minute standard WAIS-IV test and a 10-minute headset calibration. As the next step, a 10-minute tutorial was presented to familiarize the participants with hand gestures used in immersive augmented reality, 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.

## 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 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. 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 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.

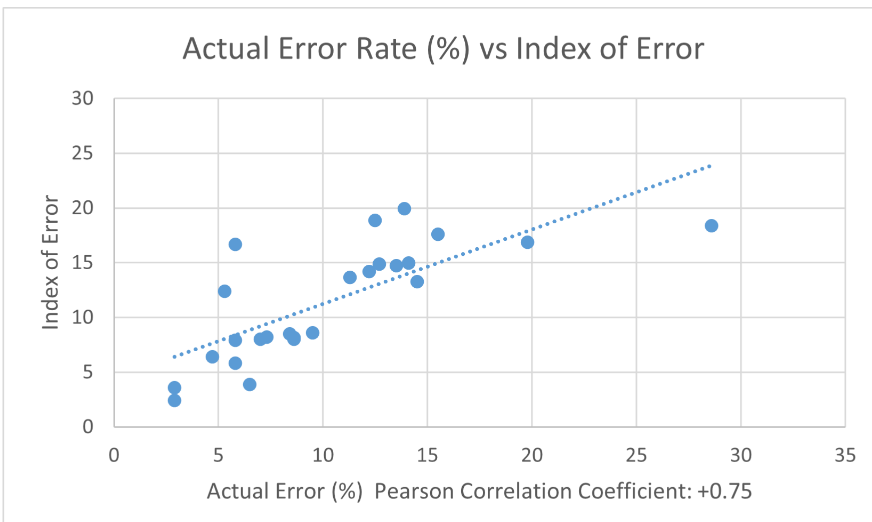


Figure 7: The index of error, vertical, and the real values of error for different people, horizontal, sorted increasingly.

In addition to the high correlation between the index and the real error rates, Table 1 shows the Pearson correlations between the learning rate calculated as it is explained in Section 3.1.1 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.

# 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

Table 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 |

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.

## ACKNOWLEDGMENTS

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

## REFERENCES

[1] Kurul, R., Ögün, M. N., Neriman Narin, A., Avci, Ş., and Yazgan, B., "An alternative method for anatomy training: Immersive virtual reality," *Anatomical Sciences Education* **13**(5), 648–656 (2020).

[2] Galati, R., Simone, M., Barile, G., De Luca, R., Cartanese, C., and Grassi, G., "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).

[3] Pathmanathan, N., Becher, M., Rodrigues, N., Reina, G., Ertl, T., Weiskopf, D., and Sedlmair, M., "Eye vs. head: Comparing gaze methods for interaction in augmented reality," in [*ACM Symposium on Eye Tracking Research and Applications*], 1–5 (2020).

[4] Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., and Irani, P., "Consumed endurance: a metric to quantify arm fatigue of mid-air interactions," in [*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*], 1063–1072, ACM (2014).

[5] Saidi, H., Dubois, E., and Serrano, M., "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*], 1–17 (2021).

[6] Pourmemar, M. and Poullis, C., "Visualizing and interacting with hierarchical menus in immersive augmented reality," in [*ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry*], ACM (2019).

[7] Li, Y., Bengio, S., and Bailly, G., "Predicting human performance in vertical menu selection using deep learning," in [*Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*], 29, ACM (2018).

[8] Pourmemar, M., Joshi, Y., and Poullis, C., "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)*], 1–8, IEEE (2022).

[9] Gebhardt, S., Pick, S., Leithold, F., Hentschel, B., and Kuhlen, T., "Extended pie menus for immersive virtual environments," *IEEE transactions on visualization and computer graphics* **19**(4), 644–651 (2013).

[10] Batmaz, A. U. and Stuerzlinger, W., "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).

[11] Bach, B., Sicat, R., Beyer, J., Cordeil, M., and Pfister, H., "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).

[12] Monteiro, P., Coelho, H., Gonçalves, G., Melo, M., and Bessa, M., "Comparison of radial and panel menus in virtual reality," *IEEE Access* **7**, 116370–116379 (2019).

[13] Samp, K. and Decker, S., "Supporting menu design with radial layouts," in [*Proceedings of the International Conference on Advanced Visual Interfaces*], 155–162, ACM (2010).

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

[15] Hick, W. E., "On the rate of gain of information," *Quarterly Journal of experimental psychology* **4**(1), 11–26 (1952).

[16] Hart, S. G., "Nasa-task load index (nasa-tlx); 20 years later," in [*Proceedings of the human factors and ergonomics society annual meeting*], **50**(9), 904–908, Sage publications Sage CA: Los Angeles, CA (2006).

[17] Jang, S., Stuerzlinger, W., Ambike, S., and Ramani, K., "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*], 3328–3339 (2017).

[18] Wechsler, D., "Wechsler adult intelligence scale–fourth edition (wais–iv)," *San Antonio, TX: NCS Pearson* **22**, 498 (2008).