Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality

Majid Pourmemar
Immersive and Creative Technologies Lab
Department of Computer Science and Software Engineering
Concordia University
Montreal, Quebec
mapourmemar@gmail.com

Charalambos Poullis
Immersive and Creative Technologies Lab
Department of Computer Science and Software Engineering
Concordia University
Montreal, Quebec
charalambos@poullis.org

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'.

CCS CONCEPTS

• **Human-centered computing → Pointing**; *User studies*; *Gestural input*.

KEYWORDS

Interaction paradigms, Graphical user interfaces, Interaction techniques, Gestural input, Head Pointing

ACM Reference Format:
Majid Pourmemar and Charalambos Poullis. 2019. Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality. In *The 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '19), November 14–16, 2019, Brisbane, QLD, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3359997.3365693

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 [Yue et al., 2017], to more complex-active actions where the user is able to create, edit, and manipulate the virtual objects [Brudy, 2013], [Xiao et al., 2018].

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 a 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 [Hoang and Thomas, 2008] 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 to be 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

VRCAI '19, November 14–16, 2019, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-7002-8/19/11…$15.00
https://doi.org/10.1145/3359997.3365693

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, the hand gestures were found to be the least preferred input modality for menu selection due to the increased physical activities required for prolonged use.

The paper is organized as follows: Section 2 provides a brief overview of the state-of-the-art in the area. In Section 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 4. Section 5 describes the evaluation of user study data including objective and subjective measures. A discussion is presented in Section 6 followed by the conclusion and future work in Section 7.

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.

2.1 Hierarchical Menus in AR

Since in the AR environment the work space is a boundless 3D area around the user, multiple varieties of GUIs coupled with different interaction methods have been proposed in order to manipulate 3D objects in the view of user and interact with them. From simple 2D drop-down menus [Hoang and Thomas, 2008], to more interactive 2D menus such as radial/ring menus [Gebhardt et al., 2013], [Davis et al., 2016], [Gerber and Bechmann, 2004] and more specifically 3D menus which can not be represented in 2D areas [Azai et al., 2017],[Bowman and Wingrave, 2001].

As previously mentioned, the 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 [Matsui and Yamada, 2008] 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 [Samp and Decker, 2010] or squares [Ahlström et al., 2010]. Radial menus since their introduction by Callahan et al. [Callahan et al., 1988] were shown to reduce the pointing time when compared to a traditional drop-down menu. At the time of the study in [Callahan et al., 1988] the authors only considered single-level drop-down menus. It was much later that [Samp and Decker, 2010] 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 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 the multi-level menus.

In [Gebhardt et al., 2013] 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 feed-back when considering the pointing time in immersive 3D environments.

Another variation of radial menus has been presented by in [Davis et al., 2016]. 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., 2017] 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.

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

(a) (b) (c)

Figure 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.

used for selecting or manipulating the virtual objects [Jankowski and Hachet, 2013].

[Lin et al., 2017] 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., 2018], 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 [Park et al., 2008], [Reilink et al., 2010].

In [Kytö et al., 2018] and [Blattgerste et al., 2018] 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 [Kytö et al., 2018] 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 in [Hincapié-Ramos et al., 2014] 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 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 [Callahan et al., 1988], [Samp and Decker, 2010] 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 1b. We have implemented two interaction methods for this menu: hand gestures (which includes 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

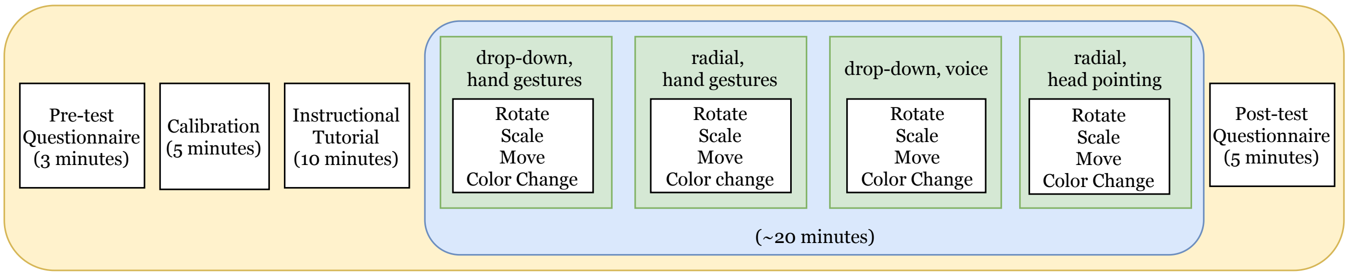


Figure 2: The experimental setting. All participants had to fill-out a pre-test questionnaire followed by a 5-minute system calibration to adjust the headset, and a 10-minute instructional tutorial to familiarize them with the hand gestures used in AR e.g. air-tapping, bloom, etc. Next, the participants had to perform four tasks with four (menu, input modality) pairs, after which they had to fill-out a post-test questionnaire.

centered at the object of interest. An example is shown in Figures 1a, 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 the 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 3a, 3b, 3c 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.

4 METHODOLOGY

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 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.

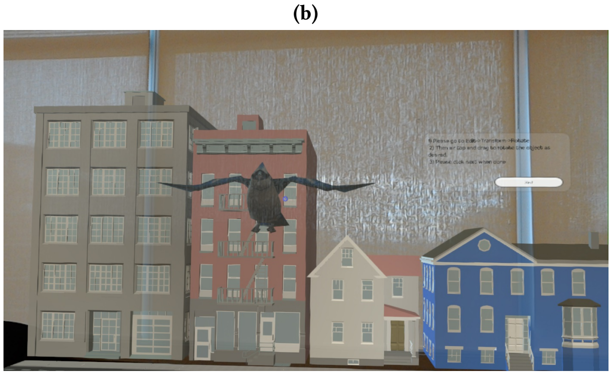
The participants were first asked to complete a pre-test questionnaire containing questions relating to demographics and background information on their experience in using AR headsets/ technologies. Following the pre-test questionnaire, the participants had to complete a 5 minute calibration process 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 2 summarizes the various steps of the experiments.

4.2 Pilot Test

A pilot test of the application were 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 two menu types.

(a)

(b)



(c)

Figure 3: (a) The application as seen from the user's perspective. The tasks include manipulating an 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.

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 [Faul et al., 2007]. 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.

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 [Samp and Decker, 2010] 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.

5 EVALUATION

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 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 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.

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)

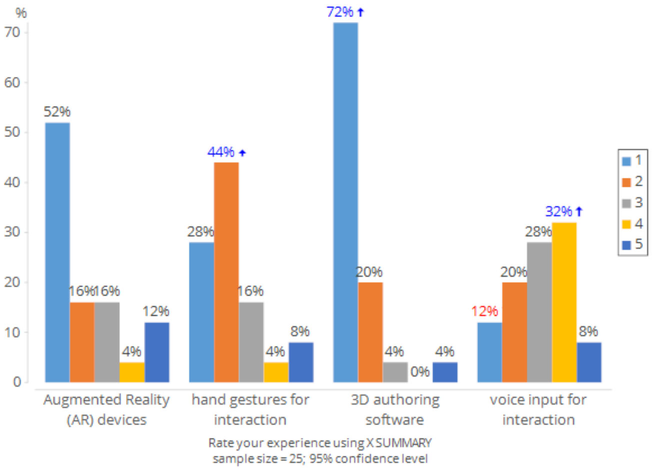


Figure 4: Pre-test responses for question 'Rate your experience using *X*' where *X* is each of [Augmented Reality (AR) devices, hand gestures for interaction, 3D authoring software, voice input for interaction]

Table 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%

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. 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 2. Figure 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).

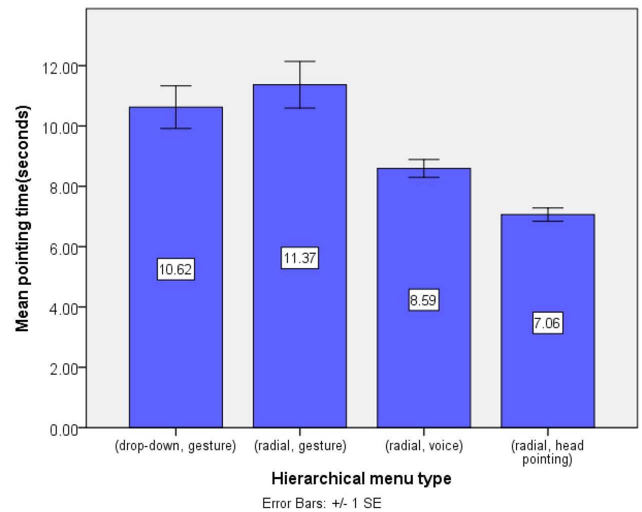


Figure 5: The mean pointing time per (menu, input modality) pair grouped by the tasks, and by the participants.

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 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 7 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.

Table 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)	-2.030	.739	**.043**	-4.021	-.040
	(radial, hand gesture)	-2.776	.736	**.002**	-4.758	-.794
	(radial, head pointing)	1.530	.316	**.000**	.679	2.381
(radial, head pointing)	(drop-down, hand gestures)	-3.560	.703	**.000**	-5.452	-1.669
	(radial, hand gestures)	-4.306	.782	**.000**	-6.412	-2.200
	(drop-down, voice)	-1.530	.316	**.000**	-2.381	-.679

Table 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

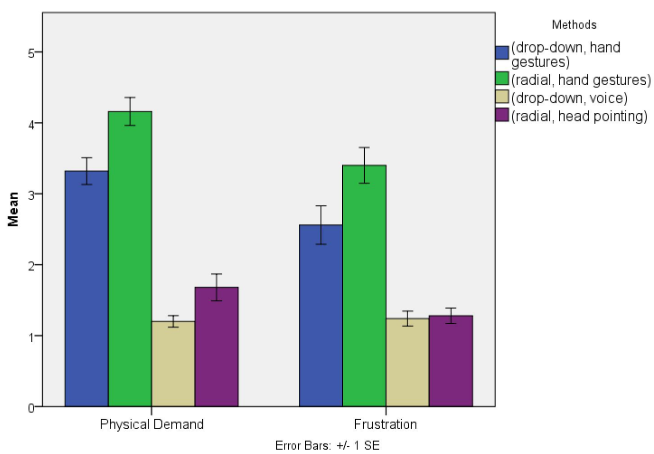


Figure 6: The mean responses from the users for physical demand and frustration when they were asked to rate the level of physical demand and frustration scaling from 1 to 5.

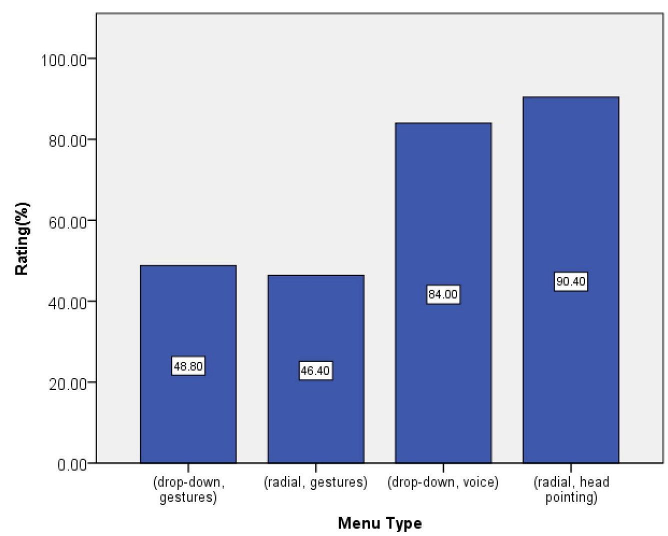


Figure 7: Rating of (menu, input modality) pairs. Each pair was rated on 5-point Likert scale.

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 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 [Özacar et al., 2016],[Xu et al., 2019] 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.

7 CONCLUSION AND FUTURE WORK

Graphical user interfaces in augmented/virtual reality are mostly inspired from 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 literature. The (radial, head pointing) pair was quantitatively shown to be the fastest. Further, when participants were asked to rate each (menu, input modality) pair it was also indicated as the most preferred.

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 in order 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 in order 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.

ACKNOWLEDGMENTS

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

REFERENCES

David Ahlström, Andy Cockburn, Carl Gutwin, and Pourang Irani. 2010. 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*. ACM, 1371–1380.

Takumi Azai, Shuhei Ogawa, Mai Otsuki, Fumihisa Shibata, and Asako Kimura. 2017. 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*. ACM, 357–360.

Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. 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*. ACM, 1.

Doug A Bowman and Chadwick A Wingrave. 2001. Design and evaluation of menu systems for immersive virtual environments. In *Virtual Reality, 2001. Proceedings. IEEE*. IEEE, 149–156.

Frederik Brudy. 2013. Interactive menus in augmented reality environments. *Beyond the Desktop* (2013), 1.

Jack Callahan, Don Hopkins, Mark Weiser, and Ben Shneiderman. 1988. An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 95–100.

Matthew M Davis, Joseph L Gabbard, Doug A Bowman, and Dennis Gracanin. 2016. Depth-based 3D gesture multi-level radial menu for virtual object manipulation. In *Virtual Reality (VR), 2016 IEEE*. IEEE, 169–170.

Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

Sascha Gebhardt, Sebastian Pick, Franziska Leithold, Bernd Hentschel, and Torsten Kuhlen. 2013. Extended pie menus for immersive virtual environments. *IEEE transactions on visualization and computer graphics* 19, 4 (2013), 644–651.

Dominique Gerber and Dominique Bechmann. 2004. Design and evaluation of the ring menu in virtual environments. *Immersive projection technologies* (2004).

Alon Grinshpoon, Shirin Sadri, Gabrielle J Loeb, Carmine Elvezio, and Steven K Feiner. 2018. Hands-Free Interaction for Augmented Reality in Vascular Interventions. (2018).

Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1063–1072.

Thuong N Hoang and Bruce H Thomas. 2008. Augmented reality in-situ 3D model menu for outdoors. In *Proceedings of the 7th IEEE/ACM International Symposium on*

Mixed and Augmented Reality. IEEE Computer Society, 185–186.

Jacek Jankowski and Martin Hachet. 2013. A survey of interaction techniques for interactive 3D environments. In *Eurographics 2013-STAR*.

Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head-and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 81.

Sikun Lin, Hao Fei Cheng, Weikai Li, Zhanpeng Huang, Pan Hui, and Christoph Peylo. 2017. Ubii: Physical World Interaction Through Augmented Reality. *IEEE Trans. Mob. Comput.* 16, 3 (2017), 872–885.

Shouichi Matsui and Seiji Yamada. 2008. Optimizing hierarchical menus by genetic algorithm and simulated annealing. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. ACM, 1587–1594.

Kasım Özacar, Juan David Hincapié-Ramos, Kazuki Takashima, and Yoshifumi Kitamura. 2016. 3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays. *Interacting with Computers* 29, 4 (2016), 579–591.

Hyung Min Park, Seok Han Lee, and Jong Soo Choi. 2008. Wearable augmented reality system using gaze interaction. In *Proceedings of the 7th IEEE/ACM international Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 175–176.

Sebastian Pick, Andrew S Puika, and Torsten W Kuhlen. 2017. Comparison of a speech-based and a pie-menu-based interaction metaphor for application control. In *2017 IEEE Virtual Reality (VR)*. IEEE, 381–382.

Rob Reilink, Gart de Bruin, Michel Franken, Massimo A Mariani, Sarthak Misra, and Stefano Stramigioli. 2010. Endoscopic camera control by head movements for thoracic surgery. In *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE, 510–515.

Krystian Samp and Stefan Decker. 2010. Supporting menu design with radial layouts. In *Proceedings of the International Conference on Advanced Visual Interfaces*. ACM, 155–162.

Robert Xiao, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko. 2018. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1653–1660.

Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Tianyu Zhang, Difeng Yu, Diego Monteiro, and Yong Yue. 2019. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. *IEEE transactions on visualization and computer graphics* (2019).

Ya-Ting Yue, Yong-Liang Yang, Gang Ren, and Wenping Wang. 2017. SceneCtrl: Mixed Reality Enhancement via Efficient Scene Editing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, 427–436.