

# Evaluating the Scalability of Non-Preferred Hand Mode Switching in Augmented Reality

Jesse Smith  
Department of CISE  
University of Florida  
Gainesville, Florida, USA  
jd.smith@ufl.edu

Isaac Wang  
Department of CISE  
University of Florida  
Gainesville, Florida, USA  
wangi@ufl.edu

Winston Wei  
Department of Computer Science  
University of Southern California  
Los Angeles, California, USA  
winstoww@usc.edu

Julia Woodward  
Department of CISE  
University of Florida  
Gainesville, Florida, USA  
julia.woodward@ufl.edu

Jaime Ruiz  
Department of CISE  
University of Florida  
Gainesville, Florida, USA  
jaime.ruiz@ufl.edu

## ABSTRACT

Mode switching allows applications to support a wide range of operations (e.g. selection, manipulation, and navigation) using a limited input space. While the performance of different mode switching techniques has been extensively examined for pen- and touch-based interfaces, investigating mode switching in augmented reality (AR) is still relatively new. Prior work found that using non-preferred hand is an efficient mode switching technique in AR. However, it is unclear how the technique performs when increasing the number of modes, which is more indicative of real-world applications. Therefore, we examined the scalability of non-preferred hand mode switching in AR with two, four, six, and eight modes. We found that as the number of modes increase, performance plateaus after the four-mode condition. We also found that counting gestures have varying effects on mode switching performance in AR. Our findings suggest that modeling mode switching performance in AR is more complex than simply counting the number of available modes. Our work lays a foundation for understanding the costs associated with scaling interaction techniques in AR.

## CCS CONCEPTS

- Human-centered computing – Gestural input

## KEYWORDS

Augmented reality, Mode switching, Non-preferred hand

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

AVI '20, September 28–October 2, 2020, Salerno, Italy  
© 2020 Association for Computing Machinery.  
ACM ISBN 978-1-4503-7535-1/20/09...\$15.00  
<https://doi.org/10.1145/3399715.3399850>

## ACM Reference format:

Jesse Smith, Isaac Wang, Winston Wei, Julia Woodward and Jaime Ruiz. 2020. Evaluating the Scalability of Non-Preferred Hand Mode Switching in Augmented Reality. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI'20)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3399715.3399850>

## 1 Introduction

Mode switching is an effective way to support a variety of operations (e.g., selection and manipulation) using a limited input space. Previous studies have investigated the performance of different mode switching techniques for pen- and touch-based interfaces [4, 7, 10, 15–17, 21, 23]. Li et al. [10] explored five different mode switching techniques for pen-based interfaces and found that the use of the non-preferred hand to switch modes was the most efficient and highest-rated. Non-preferred hand mode switching was also highly rated in touch-based interfaces [21]. Prior work examined the scalability of non-preferred hand as a mode switching technique for pen-based interfaces and found that it can be modeled by the Hick-Hyman law [15, 16].

While mode switching has been extensively examined for pen- and touch-based interfaces, mode switching in augmented reality (AR) is still relatively new. AR headsets allow users to see and interact with virtual objects projected onto a view of the real world. Users can interact with virtual objects through gaze, speech, and hand gestures. Smith et al. [19] investigated five different mode switching techniques for AR and found that the use of the non-preferred hand was the most efficient technique for mode switching. Although research has found that non-preferred hand is an efficient mode switching technique in AR [19, 21], it is still unclear how this technique performs as the number of modes increases beyond two.

We conducted a study examining the scalability of non-preferred hand mode switching in AR with two, four, six, and eight modes. During the study, participants used different non-preferred hand gestures to control the mode of a virtual object in

an AR headset. Each mode required participants to decide between a combination of shape (sphere or cube) and color (red, green, blue, or yellow). We found a significant increase in time between the two- and four-mode conditions; however, as the number of modes continued to increase, participants' performance plateaued resulting in no significant difference in task completion time for the four-, six- and eight-mode conditions. Our results suggest that increasing the types of decisions the user must make is more costly than increasing the number of modes. Based on these findings, we argue that more complex behaviors are involved for mode switching in AR than in pen- and touch-based interfaces.

Our examination into the scalability of non-preferred hand mode switching in AR provides the following insights:

1. Modeling mode switching performance in AR is more complex than simply counting the number of available modes, requiring a more in-depth analysis into the types of decisions being made and the presentation of modes.
2. Designers should consider how the complexity of articulating different hand gestures affects mode-switching performance.

We view our work as a starting point for understanding the costs associated with scaling interaction techniques in AR. This study provides new insights for developing mode-based AR applications.

## 2 Related Work

Prior work has focused on creating better interaction techniques for mode switching [10, 19–21, 23] and evaluating the scalability and performance of these techniques [13, 15, 16]. Researchers have created different mode switching techniques for pen-based interfaces [7, 10, 23], touchscreens [21], large displays [27], virtual reality [20], and augmented reality [19]. We focus our review on non-preferred hand mode switching, mode switching in AR, and modeling performance of mode switching techniques.

### 2.1 Non-Preferred Hand Mode Switching

Previous research has identified mode switching using the non-preferred hand as a technique that balances speed and accuracy. Li et al. [10] was the first to identify the non-preferred hand technique as a viable option for mode switching. They compared five different methods for pen-based interfaces, ranging from pushing a button on the pen to increasing the pressure of the pen to switch modes. They found non-preferred hand performed the best based on user preference, speed, and error rate. Lank et al. [7] extended this work to show how mode switching with the non-preferred hand can be done concurrently while the other hand is drawing with the pen. Ruiz and Lank [16] showed the advantages of bimanual pen interaction even when the number of modes increased. Surale et al. [21] further investigated mode switching in touch-based interfaces also finding that the non-preferred hand is the most efficient interaction for mode switching.

### 2.2 Mode Switching in AR

More recently, prior work on mode switching in AR focused on comparing different techniques for selecting modes, including

using the non-preferred hand. Smith et al. [19] evaluated five different mode switching techniques tailored to AR headsets, ranging from a physical button to voice-triggered interaction. Out of these five, they found that using the non-preferred hand to perform gestures to switch modes was the most efficient, balancing speed, error rate, and the ability to scale to a large number of modes. This aligns with previous findings for pen and touch interfaces. Likewise, Surale et al. [20] compared different dominant and non-dominant hand techniques in AR and also found that non-dominant techniques were faster and more accurate. However, both studies aimed to understand which techniques were more efficient and suitable for AR. Our goal is to understand how the non-preferred hand technique scales as the number of modes increases.

### 2.3 Modeling Mode Switching

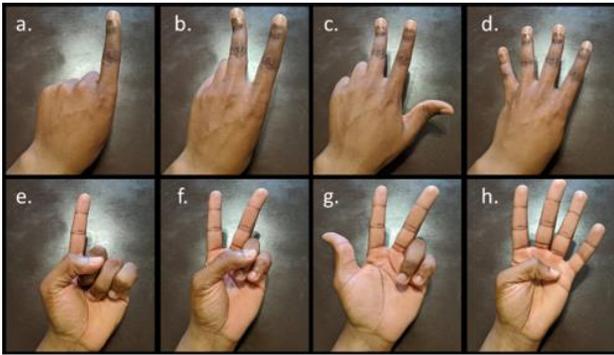
Prior work has focused on understanding how increasing the number of modes affects mode switching performance in pen-based interfaces. As mentioned above, Ruiz and Lank [16] showed how concurrent mode switching scaled to three- and four-mode interaction conditions. Their study suggests that mode switching time increases with respect to the number of modes, in accordance with the Hick-Hyman Law [3, 5]. They later supported this with a follow-up study [15] that modeled performance for two, four, six, and eight modes.

The Hick-Hyman Law states that the relationship between the time it takes to make a decision (i.e., response time) increases logarithmically with the number of choices. In the study by Ruiz et al. [15], participants completed tasks with increasing numbers of modes (up to eight), varying the thickness and color of their pen strokes as they drew on a tablet. Analysis of the participants' task times followed a logarithmic increase with an increase in modes, showing that non-preferred hand mode switching can be modeled using the Hick-Hyman Law. We seek to extend this work for mode switching in AR, particularly in finding if the increase in modes still follows the same model. It is possible that the increase in complexity of remembering and performing gestures in AR causes performance to scale differently compared to pressing buttons in a pen-based interface.

## 3 Mode Switching Gesture Selection

Before examining how increasing the amount of modes affects mode switching performance in AR, we first needed to determine what interaction technique to use for our study. Smith et al. [19] found that the non-preferred hand was the most efficient technique in AR, so we decided to use the non-preferred hand to examine mode switching scalability in AR. Next, we needed to determine which non-preferred hand gestures to use for our study.

Surale et al. [20] presented six classes of mid-air mode-switching gestures: 1) finger pinching, 2) finger extension, 3) closed-hand gestures, 4) open-hand gestures, 5) hand-raising, and 6) hand-to-body gestures. When choosing gestures for our study, we needed gestures that would stay within the sensor recognition area, therefore we excluded hand-raising and hand-to-body. We



**Figure 1. Gestures for non-preferred hand mode switching: a) 1-back, b) 2-back, c) 3-back, d) 4-back, e) 1-palm, f) 2-palm, g) 3-palm, and h) 4-palm.**

also excluded open- and closed-hand gestures because we needed gestures to correspond to a wider range of modes to examine scalability (i.e., not just two modes for open and close). Thus, we were left with finger pinching and finger extension gestures.

In our study, we used a Leap Motion hand controller [8] to track the participants' hand gestures. When testing gestures, we noticed that the built-in gesture recognizer was unreliable in classifying and differentiating pinching gestures, which is consistent with prior work [20]. Therefore, we excluded finger pinching gestures and only used finger extension gestures. In determining the specific finger extension gestures, they needed to be sequential (i.e., have a natural order in articulation) and chorded (i.e., allow stacking gestures) to remain consistent with previous work on the scalability of mode switching [18, 20]. We chose counting as a sequential gesture with palm orientation to allow for chording. Therefore, our final set of gestures included: 1-palm, 1-back, 2-palm, 2-back, 3-palm, 3-back, 4-palm, and 4-back (Figure 1). It should be noted that the 3-palm and 3-back gestures utilized the thumb, index, and middle fingers to resolve misclassifications between the gestures; a common solution found in previous literature [2, 11] to account for this misclassification.

## 4 Experiment

In this section, we present a study examining the scalability of non-preferred hand mode switching in AR that is based on the study Ruiz et al. [15] conducted for pen-based interfaces.

### 4.1 Participants

We had 31 adult participants in our study, ages 18 to 29 years ( $M = 20.9$ ,  $SD = 2.55$ ), none of whom were color blind. Eleven of our participants were female and 55% of our participants have had some prior experience interacting with AR devices. From the 31 participants, we excluded data from three; two due to equipment failure and one who withdrew from the study. Therefore, we report data for 28 total participants. All the participants were recruited from a local university, and either received course credit as compensation for their participation, or voluntarily

participated without compensation. Our protocol was approved by University of Florida's Institutional Review Board (IRB).

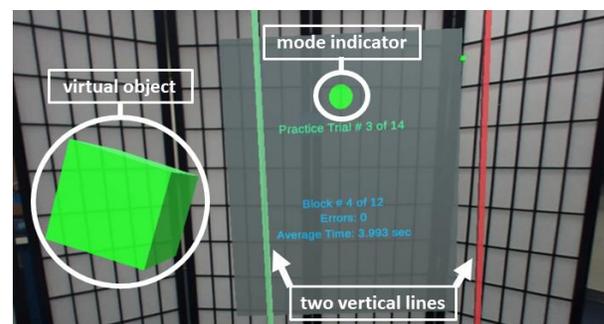
### 4.2 Apparatus

The experiment was conducted using the Meta 2 AR headset [12] tethered to a PC. The headset features an LCD display, with a resolution of 2560 x 1440px at a 72Hz refresh rate, projected onto a two-way mirrored lens allowing for a 90° field of view. A Leap Motion hand controller [8] attached to the Meta captured participants' hand movements, and the virtual environment was developed in Unity [24], a game development platform.

### 4.3 Task

Our experiment was separated into two tasks: 1) a gesture articulation task and 2) an object manipulation task. Participants were seated for both tasks to ensure no bias of fatigue from standing for an extended period of time. The goal of our gesture articulation task was to introduce and collect training samples of each gesture for each participant. We had participants perform each gesture four times and used the gesture data collected to train our recognizer for use in the object manipulation task. We utilized user-dependent sampling for our experiment since prior work [22, 25, 26, 29] has found that user-dependent recognition (i.e., testing data on training samples from the same user/source) results in more accurate recognition than user-independent sampling. The gesture articulation task took approximately five minutes.

We conducted an object manipulation task to examine the scalability of non-preferred hand modes for mode switching in AR. We based our task on Lank et al.'s [7] pen-based line cutting task, in which participants had to draw a line to bisect two lines. However, in order to adhere to more common interactions in AR and VR interfaces we modified the task from line drawing to object manipulation. We chose object manipulation because prior work has used it for examining mode switching techniques in AR [19]. During the object manipulation task, participants had to grab and move a virtual object (sphere or cube) from left to right, bisecting two vertical lines. The virtual object and lines were in the foreground of the virtual environment, and an information



**Figure 2. The task setup with the virtual object positioned to the left, followed by two vertical lines, and an information panel in the background displaying the mode indicator (annotated in figure).**

Block Design for Object Manipulation Task				
Mode Cond.	Block	Block Type	Available Modes (# of Trials)	Total Trials
two-	1	practice	m1 & m2 (5 each)	10
	2	experimental (u)	m1 & m2 (12 each)	24
	3	experimental (w)	m1 (6), m2 (18)	24
four-	4	practice	m1 & m2 (2 each) m3 & m4 (5 each)	14
	5	experimental (u)	m1 - m4 (6 each)	24
	6	experimental (w)	m1 & m2 (3 each) m3 & m4 (9 each)	24
six-	7	practice	m1 - m4 (2 each), m5 & m6 (5 each)	18
	8	experimental (u)	m1 - m6 (4 each)	24
	9	experimental (w)	m1 - m4 (2 each), m5 & m6 (8 each)	24
eight-	10	practice	m1 - m6 (2 each), m7 & m8 (5 each)	22
	11	experimental (u)	m1 - m8 (3 each)	24
	12	experimental (w)	m1 - m6 (2 each), m7 & m8 (6 each)	24

**Table 1. Block design for object manipulation task. Unweighted-mode blocks denoted by *u* and weighted-mode blocks denoted by *w*. Mode-1 is denoted by *m1*, mode-2 by *m2*, etc. Each mode condition is differentiated by color: two- (grey), four- (green), six- (blue), and eight- (orange).**

panel was in the background. The information panel consisted of a mode indicator, indicating the intended mode of the virtual object, trial and block counters, and statistics on the amount of errors and average trial time per block (see Figure 2).

The participants were instructed to 1) use their non-preferred hand to change the mode of the virtual object to the indicated mode, 2) use their preferred hand to grab and move that object from left to right, bisecting the two lines, then 3) release the object to the right of the rightmost line.

#### 4.4 Design and Procedure

To analyze the scalability of modes, we examined four different mode conditions during the object manipulation task (two-, four-, six-, and eight-mode). Each participant performed the conditions in chronological order, instead of using counterbalancing, to be consistent with Ruiz et al.'s [15] study in modeling pen-based non-preferred hand mode switching. We were concerned that initially introducing a large amount of information to the user (i.e., starting with eight modes) would increase cognitive load for the task and would require extensive training not feasible within the time constraints of a single study session. However, by controlling the mode condition order, we could effectively manage the amount of new information presented to the participant without them feeling overwhelmed or increasing mental fatigue.

In our experiment, each participant completed twelve blocks of the object manipulation task, separated into three blocks per

mode condition (4 mode conditions x 3 blocks each = 12 total blocks), shown in Table 1. The first block for each condition was a practice block where participants familiarized themselves with the introduced modes and mechanics of the experiment. After the practice block, the participant would then complete the two experimental blocks. The first experimental block balanced the occurrences of available modes (*unweighted-mode block*), whereas the second experimental block put more weight on the newly introduced modes (*weighted-mode block*), difference shown in Table 1. We designed our weighted-mode block to represent real-world applications in which some modes will be used more commonly than others, consistent with Ruiz et al [15]. We recorded trial times and errors for analysis (practice block data was not included in any analysis).

The procedure for each trial follows: 1) the mode indicator would appear after a two-second time delay, 2) the participant would then complete the mode switch for the indicated mode and the object manipulation task, 3) the current trial would end when the participant released the object. The participants were given the opportunity to take breaks between each individual block, and after three blocks of each condition the participants completed a subjective survey about their experience for that condition. We collected data from 5,376 trials (28 participants x 4 mode conditions x 2 experimental blocks x 24 trials).

Mode options ranged between two shapes (sphere or cube) and four colors (red, green, yellow, blue); these options were balanced through a Latin Square design. Recall our selected gestures from Section 3 are 1-palm, 1-back, 2-palm, 2-back, 3-palm, 3-back, 4-palm, and 4-back (Figure 1). Hand orientation (i.e., palm or back) controlled the shape while counting gestures (i.e., the number of fingers) controlled the color. In the two-mode condition, participants only changed the shape of the object, meaning they only altered their hand orientation. For the rest of the mode conditions (i.e., four-, six-, and eight-mode) the modes changed in both shape and color. For example, in the four-mode condition, a participant could switch between a red sphere, red cube, yellow sphere, and yellow cube. The hand orientation would change the shape (e.g., back orientation for a sphere), and the counting gesture would then control the color (e.g., one finger for red). Therefore, based on the example, if the participant completed a back gesture with one finger, then the mode would be a red sphere. We explained the color and shape gesture mappings to participants before each block. For each trial, the mode presentations were randomized such that the same mode would not appear consecutively more than twice.

## 5 Results

In this section we examine 1) errors associated with the object manipulation task, 2) mode switching trial times, and 3) evaluate our results based on the Hick-Hyman Law.

Mode Condition	Mode Error		Crossing Error	
	Mean	SD	Mean	SD
two-	2.46	3.16	6.18	6.48
four-	3.86	2.63	4.63	3.76
six-	4.09	3.47	4.46	4.93
eight-	3.79	2.78	3.27	3.9

**Table 2. Means and standard deviations of the percent of mode and crossing errors per mode condition (%).**

## 5.1 Error Analysis

Similar to prior work [10, 19–21, 23], we examined two types of errors: mode errors and crossing errors. Mode errors occurred when the object was in the incorrect mode state at the time of a grab event (e.g. blue cube was grabbed when the intended mode was yellow sphere). Crossing errors occurred when the participant failed to bisect both vertical lines with the grabbed object. This measure captures errors related to crossing accuracy and participants intentionally aborting the object manipulation task. Mode and crossing error percentages are listed in Table 2.

For our analysis on error rates, our data did not meet normality assumptions which was confirmed with a Shapiro-Wilks normality test. Therefore, we utilized the Aligned Rank Transform analysis of variance test (ART-ANOVA) [28]. This is a non-parametric alternative to a repeated measures analysis of variance (RM-ANOVA) when normality assumptions are not met. We used the transformed data for analysis, however, the mean errors rates presented in this paper are the actual measured values.

**5.1.1 Mode Error Rate.** In our analysis of mode error rates per number of modes, an ART-ANOVA found a significant main effect of number of modes on the mode error rates ( $F_{3,80} = 3.516, p < 0.05$ ). We ran a post-hoc comparisons using a pairwise t-test and found that participants made significantly less errors ( $p < 0.05$ ) in the two-mode condition when compared to both the four-mode and six-mode conditions.

**5.1.2 Crossing Error Rate.** We also analyzed crossing error rates per number of modes. An ART-ANOVA found no significant main effect of number of modes on the crossing error rates ( $F_{3,80} = 1.42, n.s.$ ). We do note that while there was no significant effect, the error rate decreased as the participants completed more blocks (i.e., as the number of modes increased). We attribute this decrease in error rate to participants getting used to the mechanics of dragging and releasing an object in 3D space over time.

## 5.2 Time Analysis

We model our trial time ( $T_{trial}$ ) based on prior work [15]:

$$T_{trial} = (T_p + T_a) + T_g + T_x \quad (1)$$

where  $(T_p + T_a)$  is the planning and mode activation (PA) time (i.e., time from start of a trial to the final mode selection),  $T_g$  is the grab time (i.e., time from final mode selection to the object grab event), and  $T_x$  is the crossing time (i.e., time from object grab event to object release event). We collected all the raw times from

our user studies, excluding trials with mode and crossing errors, then calculated the respective means by averaging the times for each participant over each of the conditions (two-, four-, six-, and eight-mode). Means and standard deviations of our time analysis can be found in Table 3. For our time analysis, we utilized a RM-ANOVA and checked normality with a Shapiro-Wilks normality test. When the times were not normal, we applied either a log transform or an inverse transform to the data. We utilized the transformed data for analysis, however, the mean times presented in this paper are the actual measured values. When the RM-ANOVA found significant effects, a Tukey post-hoc test was performed and reported significant of  $p < 0.5$ .

**5.2.1 Trial Time Analysis.** We first wanted to get an overview of how the number of modes affected trial time ( $T_{trial}$ ) before digging deeper into what components of the task impacted task performance. A RM-ANOVA found a significant main effect of number of modes on the trial time ( $F_{3,81} = 3.34, p < 0.05$ ). A Tukey post-hoc comparison test found that participants were significantly faster in the two-mode condition ( $M = 3.66$  seconds (s),  $SD = 2.01$ ) than the four-mode condition ( $M = 4.04$ s,  $SD = 2.17$ ). No other significance was found among the remaining comparisons. However, we do note that participants completed the six-mode ( $M = 3.86$ s,  $SD = 1.86$ ) and eight-mode conditions ( $M = 3.69$ s,  $SD = 1.69$ ) quicker than the four-mode.

**5.2.1.1 Planning and Activation Time Analysis.** We analyzed planning and activation (PA) time ( $T_p + T_a$ ) with a RM-ANOVA and found a significant main effect of the number of modes on the PA time ( $F_{3,81} = 61.62, p < 0.001$ ). When plotting the data, we saw a significant increase in the PA time when jumping from the two-mode condition to the four-mode condition. A Tukey post-hoc comparison test found that participants were significantly faster in the two-mode condition ( $M = 0.96$ s,  $SD = 1.23$ ) than the remaining conditions: four-mode ( $M = 1.78$ s,  $SD = 1.63$ ), six-mode ( $M = 1.71$ s,  $SD = 1.37$ ), and eight-mode ( $M = 1.71$ s,  $SD = 1.28$ ).

**5.2.1.2 Grab Initiation and Crossing Time Analysis.** We also examined the time it took participants to grab the object after making their final mode selection ( $T_g$ ). Using a RM-ANOVA we found a significant main effect of the number of modes on grab initiation time ( $F_{3,81} = 7.25, p < 0.001$ ). A Tukey post-hoc comparison test showed that the grab initiation time for participants was significantly slower in the two-mode condition ( $M = 1.09$ s,  $SD = 0.76$ ) when compared to the other conditions: four-mode ( $M = 0.93$ s,  $SD = 0.56$ ), six-mode ( $M = 0.93$ s,  $SD = 0.55$ ), and eight-mode ( $M = 0.86$ s,  $SD = 0.48$ ). Participants tended to improve grab initiation time as the study progressed.

We saw the same trend of improving performance for crossing time ( $T_x$ ). A RM-ANOVA found a significant main effect of the number of modes on crossing time ( $F_{3,81} = 20.08, p < 0.001$ ). A Tukey post-hoc comparison test showed that participants were significantly slower in the two-mode condition ( $M = 1.57$ s,  $SD = 0.92$ ) compared to the other conditions: four-mode ( $M = 1.34$ s,  $SD = 0.83$ ), six-mode ( $M = 1.24$ s,  $SD = 0.66$ ), and eight-mode ( $M = 1.12$ s,  $SD = 0.62$ ). Participants were also significantly slower in the four-mode condition when compared to eight-mode.

Mode Condition	$T_p + T_a$		$T_g$		$T_x$		$T_{trial}$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
two-	0.96	1.23	1.09	0.76	1.57	0.92	3.66	2.01
four-	1.78	1.63	0.93	0.56	1.34	0.83	4.40	2.17
six-	1.71	1.37	0.93	0.55	1.24	0.66	3.86	1.86
eight-	1.71	1.28	0.86	0.48	1.12	0.62	3.69	1.69

**Table 3. Means and standard deviations of planning and activation time ( $T_p + T_a$ ), grab time ( $T_g$ ), crossing time ( $T_x$ ), and total trial time ( $T_{trial}$ ). Times are given in seconds.**

5.2.2 *Analysis of Other Factors.* We also analyzed time in terms of the color of the object, shape of the object, unweighted-mode vs weighted-mode blocks (i.e., first vs second experimental block), and gesture type. We only examined PA time ( $T_p + T_a$ ) since it involves the number of modes. We did not find any significant effects for three of these factors when running a RM-ANOVA for main effects on PA time: color of object ( $F_{3,81} = 0.82, n.s.$ ), shape of object ( $F_{1,27} = 1.88, n.s.$ ), and unweighted-mode vs weighted-mode blocks ( $F_{1,27} = 1.48, n.s.$ ). The color and shape of an object did not have any effect on the study results. Our findings strengthen the notion that participants focused on the mode change instead of memorizing the frequency in which a mode would appear.

We did find a significant main effect of gesture type (1-palm, 1-back, 2-palm, 2-back, 3-palm, 3-back, 4-palm, and 4-back) on PA time when running a RM-ANOVA ( $F_{7,186} = 10.40, p < 0.001$ ). A Tukey post-hoc comparison test showed that there were significant differences among the counting gestures, namely that both the 1- and 4-gestures ( $M = 1.38s, SD = 0.68$  and  $M = 1.40s, SD = 1.05$  respectively) were significantly faster than the 2- and 3-gestures ( $M = 2.08s, SD = 1.01$  and  $M = 2.21s, SD = 1.37$  respectively,  $p < .05$  in all cases). There were no significant differences found for hand orientation among the counting gestures (i.e., palm vs back); see Figure 3. We must note that this analysis was only run on data from the unweighted eight-mode block when all gestures were used an equal amount of times.

### 5.3 Exploratory Model Fit Analysis

In Hick’s [3] original study, he found that modeling response time (RT) based on number of equally probable choices follows a logarithmic function. This function expressed linearly is:

$$RT = a + b * \log_2(n) \tag{2}$$

where  $a$  and  $b$  are constants, and  $n$  is the number of equal choices. Recall our model of trial time ( $T_{trial}$ ) from Equation 1. Similar to Ruiz et al. [15], we substituted response time (RT) with the mode planning and activation (PA) time ( $T_p + T_a$ ), resulting in the following equation:

$$T_p + T_a = a + b * \log_2(n) \tag{3}$$

where  $a$  and  $b$  are empirically determined constants, and  $n$  is the number of modes. In our analysis, a Pearson’s Product-Moment Correlation found a moderate correlation between observed PA time ( $T_p + T_a$ ) and the RT model ( $R^2 = 0.84$ ), however, the correlation was not significant ( $t_2 = 2.20, n.s.$ ).

We also analyzed this model for weighted mode occurrences. In our weighted-mode blocks, we increased the ratio of new modes to old modes. This allowed us to analyze Hick’s [3] original logarithmic model with choice variation against his model with equally probable choices (Equation 2). Allowing variation in the probability of any one choice produces the following equation:

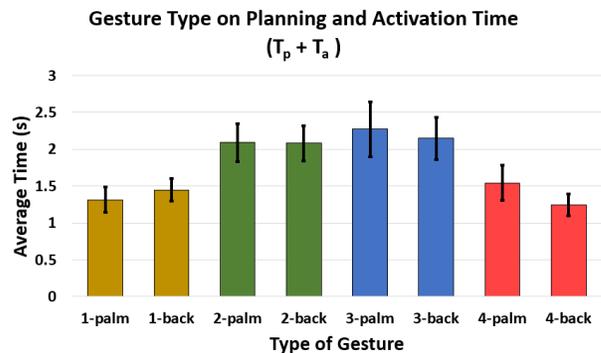
$$RT = a + b \sum_{i=1}^n p_i * \log_2\left(\frac{1}{p_i}\right) \tag{4}$$

where  $p_i$  is the probability of the  $i^{th}$  choice,  $a$  and  $b$  are constants and  $n$  is the number of unequal choices. Again, we substitute RT with PA time ( $T_p + T_a$ ) deriving the following equation for weighted modes:

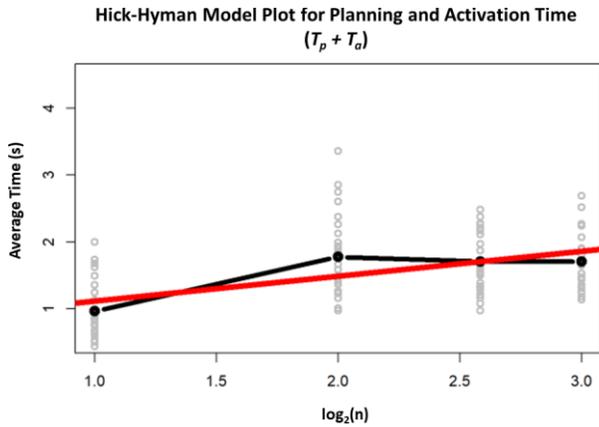
$$T_p + T_a = a + b \sum_{i=1}^n p_i * \log_2\left(\frac{1}{p_i}\right) \tag{5}$$

where  $p_i$  is the proportion of how often the  $i^{th}$  mode appears,  $a$  and  $b$  are empirically determined constants, and  $n$  is the number of modes. A Pearson’s Product-Moment Correlation found a moderate correlation between observed PA time ( $T_p + T_a$ ) and the RT model ( $R^2 = 0.82$ ) with no significance in this correlation as well ( $t_2 = 2.09, n.s.$ ). This result aligns with our earlier finding of no significance between weighted- and unweighted-mode blocks.

For both weighted- and unweighted-mode blocks, correlation values are lower than those previously reported [15]. This suggests that there may be other factors contributing to this fit as



**Figure 3. Average planning and activation time ( $T_p + T_a$ ) per gesture type. Error bars represent the 95% confidence interval.**



**Figure 4. Hick-Hyman model against observed planning and activation time ( $T_p + T_a$ ). Each gray point represents average times per participant, the black line represents the total average time across each condition, and the red line represents the model fit.**

shown by PA time ( $T_p + T_a$ ) plateauing after our two-mode condition. Overall, our observed data does not support the model for both the weighted- and unweighted-mode occurrences.

## 6 Discussion and Design Implications

In this section, we discuss our findings and how they relate to refining non-preferred hand mode switching for AR. We first discuss possible learning effects in the experiment. Following this, we discuss how our data compares to the Hick-Hyman model and how our original assumptions may have been incorrect. We finish with insights into gesture production and the implications they have for future gesture-based mode switching techniques.

### 6.1 Discussion of Learning Effects

Recall that we saw improvements in grab initiation time and crossing time as the modes increased. Ideally, this would remain constant across conditions. These improvements showed that learning was a possible factor in our object manipulation task. Our experimental design was based on work by Ruiz et al. [15]. Thus, their same concerns with overloading participants with too many modes at the start applies to our study.

We attribute the improvements in grab initiation time and crossing time to the chronological nature of the experiment design. Since participants saw each condition in order, two- then four- then six- then eight-mode, their mechanics of grabbing, dragging, and releasing an object improved. However, planning and activation (PA) times (more indicative of actual mode switching performance) showed no noticeable learning effect.

### 6.2 Modeling Mode Switching Performance

In Section 5.3 we analyzed how well our data fit the existing model for mode switching. Ruiz et al. [15] applied the Hick-Hyman Law [3, 5] for modeling non-preferred hand scalability of pen-based

mode switching. However, we discovered that their model does not accurately reflect the decision making within our task design.

We noticed differences in our data's behavior compared to the Hick-Hyman model. In the Ruiz et al.'s [15] model, PA time increased logarithmically with the number of modes (a large increase early on, and then smaller increases following that). In our data, we saw a significant increase between the two- and four-mode conditions, then times plateaued to around 1.7s (Figure 4) for the remaining mode conditions. The only temporal increase observed was moving from the two-mode to four-mode condition, which is still consistent with prior work [16]. So, we argue that the Hick-Hyman Law is still a good model for mode switching, but there are more complex behaviors present need to be considered.

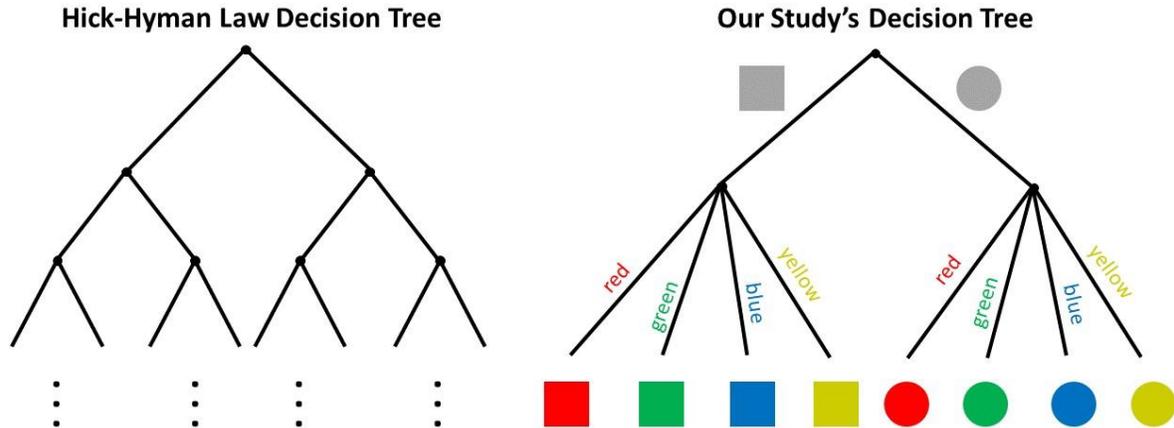
The original Hick-Hyman Law focused on the concept of decisions rather than number of possible options (e.g., modes); stating that a person's reaction time is based on the number of decisions to make. The original study by Hick [3] modeled a series of binary decisions, shown in Figure 5 (left). This resulted in an average of  $\log_2(n)$  decisions needed, with  $n$  representing the number of possible combinations. In the previous study [16] and in our study, we believed mode selection followed the same behavior, with  $n$  equal to the number of modes. However, we now see that there are fewer decisions than originally thought.

In our study, a person would need to choose from a max of two shapes and four colors. In the two-mode condition, there was only one decision to make, choosing between cube or sphere. By adding color in the four-mode condition, we increased the decisions to two: one for choosing the shape and one for choosing color. However, the number of decisions remained the same in the six- and eight-mode conditions. In those conditions, we only increased the number of color options available, rather than truly increasing the number of decisions, shown in Figure 5 (right).

As a result, we cannot generalize mode switching time to simple logarithmic behavior. According to Hick-Hyman, time depends more on the number of decisions than just the number of combinations. The two are related, but we cannot assume all mode switching follows the Hick-Hyman decision tree. This has interesting implications for mode switching in AR and may generalize to other 3D interfaces, such as VR. Apart from the initial addition of color with shape, adding more colors did not significantly increase PA time. Our findings show that for non-preferred hand mode switching in AR, adding another decision type is costly, but modes can be added with minimal cost.

### 6.3 Gesture Type Considerations

One interesting finding from our study was the effect of gesture type on PA time. Our results showed that the 1- and 4-counting gestures were significantly faster than the 2- and 3-counting gestures. The orientation of the hand did not make a difference. The performance of hand gestures is a relatively unexplored area. Prior work has focused on understanding the difficulty of articulating multi-touch gestures [14] and quantifying the performance of touch gesture production based on human kinematics [9]. However, this type of analysis has not been conducted for hand gestures. One study, by Erazo and Pino [1], looked at modeling the time it would take to execute a hand



**Figure 5. Left: Hick-Hyman Law decision tree model, Right: the decision tree model from our mode switching study. Both models show the traversal of making a choice based on the design of the task.**

gesture based on the mechanics of the motion (e.g., swiping or retracting the arm). Their study did not investigate the performance and difficulty of articulating the wrist and fingers to make different hand gestures.

Our findings make intuitive sense, as the 2- and 3-counting gestures are more complex. These gestures require moving the middle and ring fingers separately, which is difficult because they share a common flexor muscle [6]. Flipping the orientation of the hand, however, can be considered a simple motion, as the wrist only rotates about one axis. The fact that certain gestures resulted in worse performance has implications for gesture-controlled applications where temporal performance is critical. Different gesture attributes may affect the time it takes for a person to perform the gesture, such as finger movement constraints or other biomechanical considerations. Future work is needed to further understand how these attributes affect gesture production.

## 7 Limitations and Future Work

One limiting factor in our study involves the chronological introduction of modes (i.e., each participant experienced the two- then four- then six- then eight-mode conditions). As previously noted, this may have caused the improved performance for both grab time ( $T_g$ ) and crossing time ( $T_x$ ), introducing an unintended order effect. Future work can reproduce this study using a counterbalanced approach, however, the experiment needs to be carefully designed so that 1) participants are not overwhelmed by the introduction of a large amount of modes and 2) adequate training of the system and mode selection techniques does not independently bias performance from one condition to another.

We found that adding the extra complexity of color when going from the two- to four-mode conditions dramatically increased PA time, while adding more colors did not (i.e., during four-, six-, and eight-mode conditions). Adding decision types can be costly, while modes can be added with minimal cost. Future work is needed to explore the cognitive cost of adding decision types when compared to modes for mode switching in AR.

## 8 Conclusion

We examined the scalability of non-preferred hand mode switching in AR with two, four, six, and eight modes. We found that as the modes increased, performance plateaued instead of following a logarithmic curve. The previous model focused directly on relating response time to the number of modes. However, based on the Hick-Hyman Law, we believe response time to be related to the number of decisions a user must make. This behavior is more complex and changes depending on the way modes and options are presented; increasing the types of decisions (e.g. adding color to shape) is more costly than increasing the number of modes (e.g. adding more color options) for non-preferred hand mode switching in AR. We view our work as the starting point for understanding the costs associated with scaling interaction techniques in AR.

## ACKNOWLEDGMENTS

We thank the McKnight Doctoral Fellowship program for their support in making this work possible. In addition, this work is partially supported by National Science Foundation Grant Awards #IIS-1750840 and #CHS-1560243. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also want to thank our reviewers for their valuable feedback and insights.

## REFERENCES

- [1] Erazo, O. and Pino, J.A. 2015. Predicting Task Execution Time on Natural User Interfaces based on Touchless Hand Gestures. *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA, Mar. 2015), 97–109.
- [2] Freeman, D., Vennelakanti, R. and Madhvanath, S. 2012. Freehand pose-based Gestural Interaction: Studies and implications for interface design. *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)* (Dec. 2012), 1–6.
- [3] Hick, W.E. 1952. On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*. 4, 1 (Mar. 1952), 11–26. DOI:<https://doi.org/10.1080/17470215208416600>.

- [4] Hinckley, K., Baudisch, P., Ramos, G. and Guimbretiere, F. 2005. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2005), 451–460.
- [5] Hyman, R.A. 1953. Stimulus information as a determinant of reaction time. *Journal of experimental psychology*. 45, 3 (1953), 188–196. DOI:<https://doi.org/10.1037/h0056940>.
- [6] Kilbreath, S.L. and Gandevia, S.C. 1994. Limited independent flexion of the thumb and fingers in human subjects. *The Journal of Physiology*. 479, 3 (1994), 487–497. DOI:<https://doi.org/10.1113/jphysiol.1994.sp020312>.
- [7] Lank, E., Ruiz, J. and Cowan, W. 2006. Concurrent bimanual stylus interaction: a study of non-preferred hand mode manipulation. *Proceedings of Graphics Interface 2006* (Toronto, Ont., Canada, Canada, 2006), 17–24.
- [8] Leap Motion: <https://www.leapmotion.com/>. Accessed: 2018-09-20.
- [9] Leiva, L.A., Martín-Albo, D. and Vatavu, R.-D. 2018. GATO: predicting human performance with multistroke and multitouch gesture input. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain, Sep. 2018), 1–11.
- [10] Li, Y., Hinckley, K., Guan, Z. and Landay, J.A. 2005. Experimental analysis of mode switching techniques in pen-based user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05* (Portland, Oregon, USA, 2005), 461.
- [11] Liu, M., Nancel, M. and Vogel, D. 2015. Gunslinger: Subtle Arms-down Mid-air Interaction. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA, Nov. 2015), 63–71.
- [12] Meta Augmented Reality: <https://www.metavision.com/>. Accessed: 2018-09-20.
- [13] Ramos, G., Boulous, M. and Balakrishnan, R. 2004. Pressure widgets. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04* (Vienna, Austria, 2004), 487–494.
- [14] Rekek, Y., Vatavu, R.-D. and Grisoni, L. 2014. Understanding Users' Perceived Difficulty of Multi-Touch Gesture Articulation. *Proceedings of the 16th International Conference on Multimodal Interaction* (Istanbul, Turkey, Nov. 2014), 232–239.
- [15] Ruiz, J., Bunt, A. and Lank, E. 2008. A model of non-preferred hand mode switching. *GI '08: Proceedings of graphics interface 2008* (Toronto, Ont., Canada, Canada, 2008), 49–56.
- [16] Ruiz, J. and Lank, E. 2007. A study on the scalability of non-preferred hand mode manipulation. *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces* (New York, NY, USA, 2007), 170–177.
- [17] Saund, E. and Lank, E. 2003. Stylus input and editing without prior selection of mode. *Proceedings of the 16th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2003), 213–216.
- [18] Scheggi, S., Meli, L., Pacchierotti, C. and Prattichizzo, D. 2015. Touch the Virtual Reality: Using the Leap Motion Controller for Hand Tracking and Wearable Tactile Devices for Immersive Haptic Rendering. *ACM SIGGRAPH 2015 Posters* (New York, NY, USA, 2015), 31:1–31:1.
- [19] Smith, J., Wang, I., Woodward, J. and Ruiz, J. 2019. Experimental Analysis of Single Mode Switching Techniques in Augmented Reality. *Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019* (Kingston, Canada, Jun. 2019), 1–8.
- [20] Surale, H.B., Matulic, F. and Vogel, D. 2019. Experimental Analysis of Barehand Mid-air Mode-Switching Techniques in Virtual Reality. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), 196:1–196:14.
- [21] Surale, H.B., Matulic, F. and Vogel, D. 2017. Experimental Analysis of Mode Switching Techniques in Touch-based User Interfaces. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (Denver, Colorado, USA, 2017), 3267–3280.
- [22] Taranta II, E.M., Samiei, A., Maghoumi, M., Khaloo, P., Pittman, C.R. and LaViola Jr., J.J. 2017. Jackknife: A Reliable Recognizer with Few Samples and Many Modalities. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA, May 2017), 5850–5861.
- [23] Tu, H., Yang, X.-D., Wang, F., Tian, F. and Ren, X. 2012. Mode Switching Techniques Through Pen and Device Profiles. *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction* (New York, NY, USA, 2012), 169–176.
- [24] Unity: <https://unity3d.com>. Accessed: 2018-09-20.
- [25] Vatavu, R.-D., Anthony, L. and Wobbrock, J.O. 2018. \$Q: a super-quick, articulation-invariant stroke-gesture recognizer for low-resource devices. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain, Sep. 2018), 1–12.
- [26] Vatavu, R.-D., Anthony, L. and Wobbrock, J.O. 2012. Gestures as point clouds: a \$P recognizer for user interface prototypes. *Proceedings of the 14th ACM international conference on Multimodal interaction* (Santa Monica, California, USA, Oct. 2012), 273–280.
- [27] Vogel, D. and Balakrishnan, R. 2005. Distant freehand pointing and clicking on very large, high resolution displays. *Proceedings of the 18th annual ACM symposium on User interface software and technology* (Seattle, WA, USA, Oct. 2005), 33–42.
- [28] Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (Vancouver, BC, Canada, 2011), 143.
- [29] Wobbrock, J.O., Wilson, A.D. and Li, Y. 2007. Gestures Without Libraries, Toolkits or Training: A \$I Recognizer for User Interface Prototypes. *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2007), 159–168.