Collecting and Analyzing the Mid-Air Gestures Data in Augmented Reality and User Preferences in Closed Elicitation Study

Jieqiong Li(✉), Adam S. Coler, Zahra Borhani, and Francisco R. Ortega

NUILab, Department of Computer Science, Colorado State University, Fort Collins, CO 80521, USA
{ljq915,AdamWil,zahra.borhani,f.ortega}@colostate.edu

Abstract. For users of AR-HMDs (i.e., AR glasses), one option for interaction with the HMD is through midair gestures. This modality allows users to move their hands to grab objects, move objects, select menus, and interact with a multiple set of features provided by AR-HMDs. However, midair gestures are still difficult to recognize. For example, the Microsoft HoloLens 2 only provides a few gestures, and there are many false positives in the gesture recognition of the system. Further, it is not clear what gestures the users prefer. As a result, we performed a study with data collection on midair gestures for AR-HMDs with the aim of providing a gesture dataset that can be utilized in machine learning models. In addition, we analyzed the user gesture preferences for object manipulation. Our collected dataset includes the trajectory of hand gestures and eye gaze, which are captured by Articulated Hand Tracking (AHAT), inertial measurement unit (IMU) and other sensors. Our analysis showed that users have different gesture preferences for different categories of referents.

Keywords: Augmented Reality · Elicitation Study · Gesture Dataset · AR-HMD · Data Collection

1 Introduction

Elicitation studies started in 2009 with [1], followed by multiple studies (e.g., [2–4]) in the search for improved gesture interfaces. This has included elicitation studies in augmented reality (AR) head-mounted displays (HMDs), such as Williams et al. [5–7] and others [2,8,9]. The objective is to improve everyday use of HMDs in everyday life, work, and entertainment, aligning with the vision of Mark Weiser [10] to make the computer invisible. However, it is not known if there are preferences for different gesture sets (e.g., a gesture from set A and a gesture from set B), which are one of the products of elicitation studies. In

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2024, LNCS 14707, pp. 201–215, 2024.
https://doi.org/10.1007/978-3-031-61044-8_15

this study, participants were asked to select their preferred gesture from different elicitation gesture sets while performing the gestures. In addition, we have collected all the videos and images, which can be found on the website[1].

The main contributions of this work are as follows: 1) preparing a well-curated mid-air gesture interaction dataset, aiming to improve the recognition of mid-air gestures in machine learning systems. 2) identifying users' preferences in gesture interaction for each referent by conducting a human-centric experiment.

2 Related Work

2.1 Mid-Air Gesture Data Collection

Li et al. (2019) pointed out that mid-air gestures can improve the overall user experience as compared to device-based remote control [11]. The collection of gesture data sets can greatly help the machine learning system learn and recognize mid-air gestures [11], thereby improving the user experience in the interactive environment, the system's flexibility, and the freedom of use. For example, Vogiatzidakis et al. (2020) improved the usability and user experience of smart furniture by adding mid-air gestures [4]. Williams et al. used an open elicitation study to develop a user-defined set of gesture interactions primarily for object manipulation to understand the user's gesture preferences in that direction [5]. Studying mid-air gestures by using elicitation studies has become one of the main research methods to understand user gesture types in different contexts.

The flexibility and feasibility of this research direction have gradually become an important topic in the field of HCI [12]. As a non-contact operation, mid-air interaction allows users to use mid-air gestures to remotely operate digital content displayed on the device [4,13]. This interaction method has the advantages of being intuitive, easy to execute, and memorable [14]. When wearing an AR-HMD (i.e., AR glasses), mid-air gestures are an option for interaction between the user and the HMD. Mid-air gestures allow users to move their hands to grab virtual objects, move objects, select menus, and interact with many of the features provided by AR-HMDs. The improvement and supplementation of the air gesture data set can increase the recognizability and diversity of AR-HMD's air gestures.

2.2 Closed Elicitation Study

An elicitation study is a classic method for researchers to study user action or interaction preferences [14]. This method was proposed by Vatavu et al. in 2009 [1]. Its main advantage is that the gestures are designed based on user preferences, thereby improving the user's interactive experience.

As a method suitable for gesture set design, elicitation studies can mainly help researchers understand user preferences for different devices [4], forms [3],

[1] http://tinyurl.com/nuilabgs.

or application domains [1]. Wobbrock et al. compared three methods of designing gesture sets and came to the following conclusions [9]: The gesture sets designed by researchers and users are more suitably designed by users only or researchers only. This result demonstrates the importance of participatory design for researchers to understand user preferences and collect relevant action forms early in the experiment. The results of the research on gestures by using an elicitation study show that, under the same gesture, the user will change the size of the gesture according to the size of the interacting object [2]. For different numbers of target objects, the users prefer to use both hands to interact with two objects, and for a single object, users prefer to interact with one hand [8].

As compared with an open elicitation study, a closed elicitation study requires users to complete suggestions for gesture preferences based on the appropriate gestures provided as a reference [7]. Its main purpose is still to find the natural choice after participants understand the command. The preference study mentioned in this article, also called a closed induction study [7], requires users to select their favorite interaction after each reference (i.e., 3 gestures, a total of 3 recordings for each referent). This preference study not only has the primary purpose of obtaining a more diverse and richer dataset of ego-centric mid-air gestures, but also explores the relationship between gesture preferences, referents, and gesture types in comparison to previous studies.

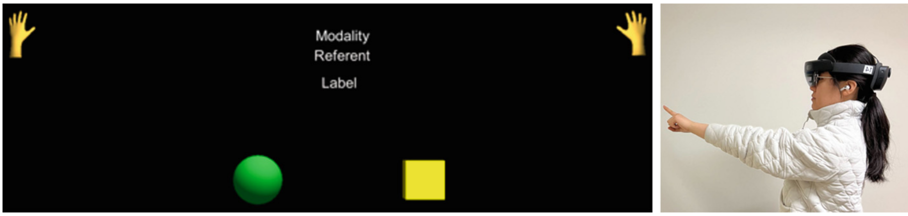

Fig. 1. Screenshot of the initial interface of the experimental software (Left), and a researcher testing HoloLens 2 (right)

3 Methods

An elicitation study was conducted to investigate gesture interactions for manipulating rendered 3D objects in optical see-through AR environments. The input modality used was a unimodal gesture modality. A Wizard of Oz (WoZ) [4] experimental design was employed, where the experimenter emulated a live system. In a Wizard of Oz experiment, participants interact with a mock interface controlled by the experimenter. Participants were presented with a pre-recorded video demonstrating a specific gesture, and they were instructed to replicate the gesture in three consecutive trials. A total of 23 referents were tested in a within-subjects experiment design. The study incorporated canonical referents (translation, rotation, and scale) alongside abstract referents (copy, paste, cut, create, destroy, and select). Table 1 shows a list of gesture referents used in this study.

Table 1. Type of gestures and relative gesture referents. *C: Clockwise, CC: Counter Clockwise

Manipulation	Gesture Referents		
Translation	Move Up	Move Down	Move Left
	Move Right	Move Forward	Move Away
Rotation	Roll C	Roll CC	Yaw Left
	Yaw Right	Pitch Up	Pitch Down
Scale\ Selection	Enlarge one object	Shrink one object	Select one object
	Enlarge two objects	Shrink two objects	Select two objects
Abstract	Copy	Paste	Cut
	Create	Destroy	

3.1 Apparatus

The AR environment was developed using Unity Game Engine 2019.2.18f1, Mixed Reality ToolKit (MRTK), and C# programming language. An Alienware Laptop with an Intel i7-8750h and NVIDIA Geforce RTX 2060 6 GB GPU was used for the development. The interface of the software is mainly composed of two 3D objects (cube and sphere) available for user manipulation, a text-based referent indicating the required manipulations, an animation illustrating the referents' actions, the name of the input modality (i.e., gesture only), and the trial number. Additionally, there are two hand-shaped icons serving as indicators of the user's hands within the device's field of view. If either of the user's hands is not detected on the screen, the corresponding hand icon disappears. Figure 1 (left) shows a screenshot of the participants' view during the experiment. Since there are task requirements for multi-object operation in this experiment, the yellow cube will be used as the main pseudo-interaction object for all of tasks, and the green sphere will be used as the auxiliary object for the multi-object task, as shown in Fig. 1 (left) and Fig. 2. In addition, the software collects the following data in the background: depth gesture images, finger joints, and eye gaze movement trajectories in the gesture data space.

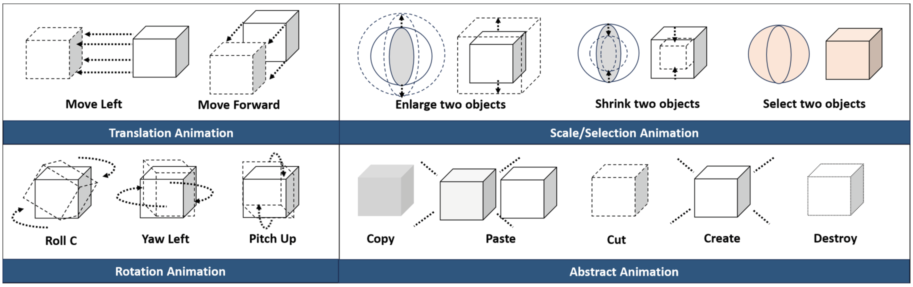

Fig. 2. Animation schematic diagram of some referents in the experiment application

This study was conducted using a Microsoft HoloLens 2 AR-HMD, which provides Articulated Hand Tracking (AHAT). The AHAT was used to collect the depth images of the participant's hand [15]. By using the AHAT depth camera, researchers captured the hand information from an ego-centric perspective. This information was then integrated with the Mixed Reality Toolkit (MRTK) in Unity to obtain depth gesture images, finger joint data, and the trajectory of eye gaze movement in a spatial position for each referent [15]. Furthermore, An inertial measurement unit (IMU) sensor was used to collect raw data from the accelerometer, gyroscope, and magnetometer during the experiment [15]. In addition to the depth camera, a GoPro Hero 7 mounted on an AR-HMD was employed to record color video.

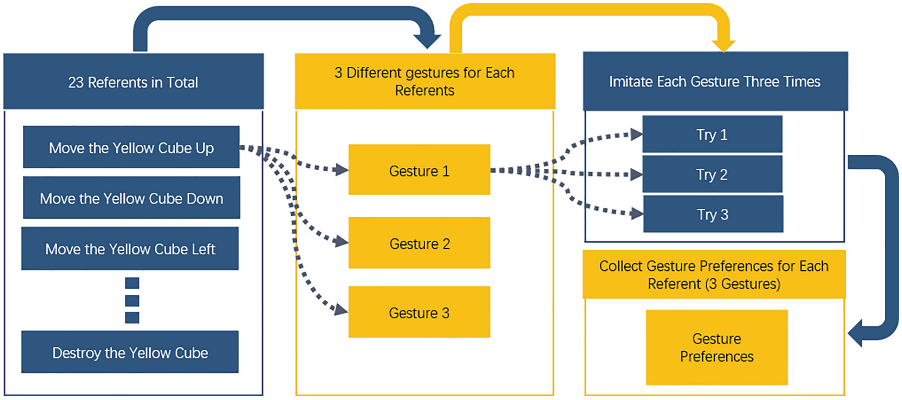

Fig. 3. Experiment Process

In order to improve the running performance of the software, depth gesture images were stored as RAW type images. After the experiment, the RAW type images were converted into PNG type gesture data images by using Unity scripts. In addition, the noise in data images was eliminated by using Python script. Its main methods involve eliminating noise in fixed areas, removing randomly appearing black spots, and deleting non-gesture data in error/experiment gaps. In addition to data processing, a Python script was utilized for image conversion video to generate a 4K video for each gesture. Each video contained three imitations of the gesture by the participant and used the name of referents to group and name the gesture depth videos. Except the processing of depth images, eye gaze spatial position data, position data of 19 joints of each hand, palm, and wrist, IMU data will be converted into CSV type files, and the color video will be cut and stored too.

3.2 Research Objective and Hypotheses

The aim of this research project was to explore the preferential differences between gesture types for various referents and different categories (e.g., translation, rotation, scale, etc.). This would also help us to summarize our findings and provide suggestions and ideas for future gesture designs. More specifically, we investigated the following research questions, which are inspired by similar previous work [5–7]:

- R_1**:** Does the type of referent in the translation category affect user preferences on how to perform a gesture (two fingers, one palm, or two palms)?
- R_2**:** Similarly, does the type of referent in the rotation category affect user preferences for associated gestures (two fingers, one palm, or two palms)?
- R_3**:** How do these preferences vary for referants in the translation and rotation categories?
- R_4**:** Do users strongly favor one of their options (pitch, grab x-axis, and graph y-axis) for the referents in the single-object scaling category?
- R_5**:** What about multi-object scaling; do users have a strong preference for any of their gesture options (palms two-hands, pitch two-hands, and grab two-hands)?
- R_6**:** How does the type of abstract referents (including copy, paste, and cut) affect user preference for the associated gestures (pitch three fingers, point two hands, and grab two hands)?

3.3 Procedure and Task

Before the experiment started, participants were asked to sign an informed consent form and were informed that the gesture data during the experiment would be recorded and used for research. An experimenter provided instructions on how to wear a Microsoft HoloLens 2 HMD and how to calibrate the eye gaze tracking. To ensure the accuracy of the depth image results and enhance the integrity of data collection, participants were asked to look straight ahead during the experiment (Fig. 4).

To minimize the impact of the order effect, the pre-recorded gesture videos for each referent were presented in a counter-balanced order using Latin squares. For each referent, the three different gestures were presented in random order. Participants were required to imitate the gestures shown in the pre-recorded videos and repeat them three times.Figure 3 shows this experiment process. After completing three gestures for one referent, the experimenter asked the participant about their gesture preference for that referent. Each experiment took around one hour. To reduce the experimental bias caused by fatigue, the researcher gave a break during the experiment based on agreement.

3.4 Participants

A total of 20 participants (10 Female, 10 Male) were recruited to participate in the experiment. All participants reported normal or corrected to normal vision. Two participants reported that they were left-handed. The participants included undergraduate and graduate students, faculty, or alumni of Colorado State University. Almost all participants were not familiar with AR and virtual

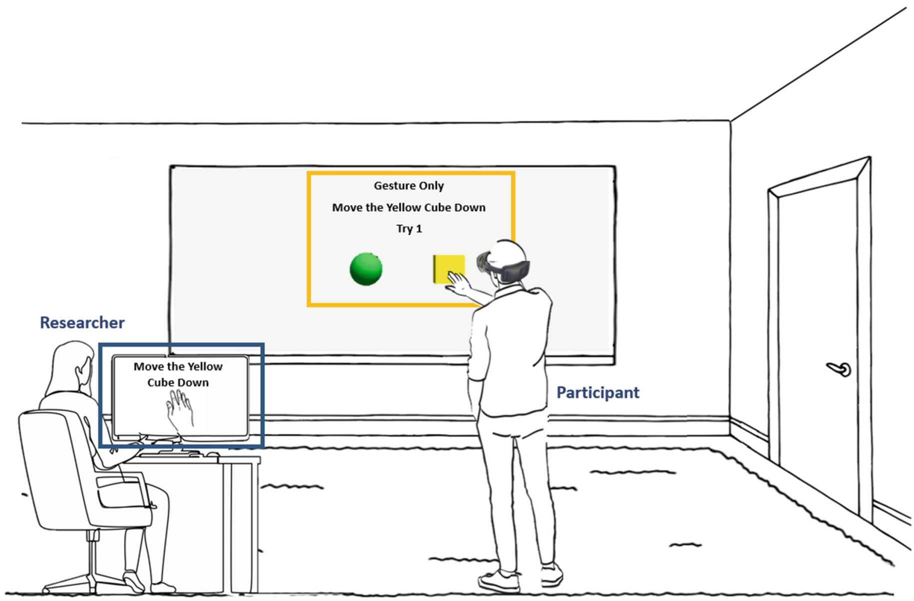

Fig. 4. Experimental simulation diagram, this figure includes a researcher (left), a participant (middle), and pre-recorded videos (left box), and the experimental animation demonstration which only can be seen when participants wearing HoloLens 2 (middle box).

reality (VR). Seventeen participants had never used an AR-HMD (e.g., Microsoft HoloLens 2 HMD) which can reduce the influence of participants' habits on the experimental results. Upon the completion of the experiment, participants could choose compensation in the form of 20$gift cards or in-class credit points. Or they could choose to participate in this experiment as volunteers.

4 Results

Our results consist of two main parts: 1) providing a mid-air gesture data set that can enhance the recognition of mid-air gestures. 2) assessing our hypothesis to explore users' preferences in gestures used for executing each referent.

4.1 Mid-Air Gesture Dataset

Our dataset includes three gestures per referent, each gesture was performed three times by users resulting in a total of nine gestures for each referent. Our study investigated a total of 23 referents across 20 participants, collecting information on 4,140 $(23 * 3 * 3 * 20)$ gesture trials. The AHAT depth camera on the Microsoft HoloLens 2 AR-HMD provided the ability to capture hand information from an egocentric perspective including depth gesture images, finger joints, and eye gaze trajectories in space. The recordings (depth images) were transformed, segmented, and denoised before being integrated with the data to produce independent depth videos for each gesture. Figure 5 shows some example images of these post-processed depth gesture videos. In addition, each of these 4k depth videos is accompanied by eye gaze spatial position data, and position data of 19 joints of each hand, palm, and wrist.

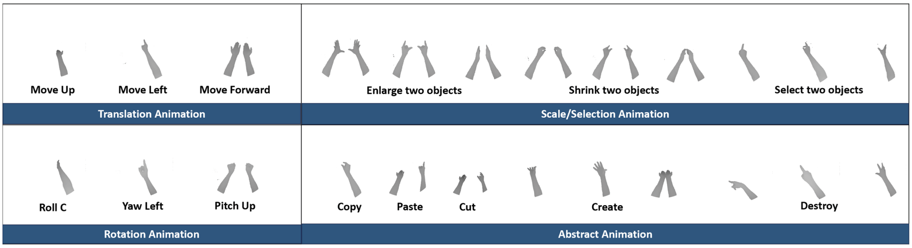

Fig. 5. Some results of gesture depth images

4.2 Gesture Preferences

The distribution of user preferences among three gesture types for canonical and abstract referents is presented in Table 2 and Table 3, respectively. Each table represents a distinct category, including its associated referents. For example, the translation category consists of six referents for translating an object in various directions: up, down, left, right, forward, and away. Users were presented with three different gesture options for each of these referents, as shown in the first column of each table. Figure 6 provides the image of these three gestures, in the same order as the rows of the tables, for each referent, along with the number of participants who chose that gesture among the three (out of 20).

4.3 Fisher−Freeman−Halton Exact Test

For each of the proposed research questions in Sect. 3.2, the statistical significance for gesture preferences is assessed using the Fisher−Freeman−Halton Exact test (FFH Exact test) [16], which is a statistical test used for analyzing contingency tables, particularly when the sample sizes are small (in our case, 20 participants) and the size of the contingency table is large (e.g., $3*6$ contingency table for the translation category). IBM SPSS Statistics (SPSS) was used for the analysis of the FFH Exact test [18]. Table 4 shows the p-value and value for the FFH Exact Test related to our six research questions.

R_1 explores how the type of referent in the translation category affects user preferences on how to perform a gesture. The translation category includes six referents (move up/down/left /right/forward/away), and three gesture types

Table 2. The percentage result of user preferences for first half of the referents. All gesture demonstrations and explanations can be found in Fig. 6. In each table, the last column labeled "All" displays the distribution of all referents within their corresponding categories. For each column, the largest value is presented in bold.

(a) The percentage result of user preferences for translation *F: forward, A: away,

Translation	**Gesture Referents**						**All**
Gestures	Move Up	Move Down	Move Left	Move Right	Move F*	Move A*	All
Two Fingers	15%	40%	45%	**50%**	15%	20%	32.5%
One Palm	**75%**	**60%**	**55%**	45%	**50%**	**55%**	**56.7%**
Two Palms	10%	0	0	5%	25%	25%	10.8%

(b) The percentage result of user preferences for rotation *C: Clockwise, CC: Counter Clockwise

Rotation	**Gesture Referents**						**All**
Gestures	Roll C*	Roll CC*	Yaw Left	Yaw Right	Pitch Up	Pitch Down	All
Two Fingers	**60%**	**70%**	**60%**	**60%**	**55%**	40%	**57.5%**
One Palm	40%	25%	30%	35%	40%	**55%**	37.5%
Two Palms	0	5%	10%	5%	5%	5%	5%

(c) The percentage result of user preferences for scale one object

Scale	**Gesture Referents**		**All**
Gestures	Enlarge One Object	Shrink One Object	All
Pitch	**80%**	**75%**	**82.5%**
Grab X-axis	20%	15%	17.5%
Grab Y-axis	0	0	0

(d) The percentage result of user preferences for scale two objects

Scale	**Gesture Referents**		**All**
Gestures	Enlarge Two Objects	Shrink Two Objects	All
Palm Two Hands	5%	15%	10%
Pitch Two Hands	**50%**	35%	42.5%
Grab Two Hands	45%	**50%**	**47.5%**

(two fingers, one palm, and two palms). The results from the FFH Exact test did not produce a statistically significant result ($p = 0.026$). This means that the differences in user preferences between the three gesture groups within the translation category were not large enough to be considered statistically significant. This indicates that there were no differences between participants' gesture choices for the referents in this category and the users' preferences were similar for these referents.

Similarly, R_2 examines whether there was any significant difference in user preferences when selecting among three gesture types for the referents in the rotation category (Roll C/CC, Yaw Left/Right, and Pitch Up/Down). The results from the FFH Exact test revealed no significant difference among the six rotation referents ($p = 0.745$), indicating that different rotation referents did not influence users' preferences in selecting gestures. As the values in Table 2 show, the participants' gesture choices were relatively uniform within this category, as well.

Table 3. The percentage result of user preferences for second half of the referents. All gesture demonstrations and explanations can be found in Fig. 6. In each table, the last column labeled "All" displays the distribution of all referents within their corresponding categories. For each column, the largest value is presented in bold.

(a) The percentage result of user preferences for select

Select	**Gesture Referents**	**Select**	**Gesture Referents**
Gestures	Select One Object	**Gestures**	Select Two Objects
Point	**55%**	Point (2 Times)	**35%**
Pitch	25%	Pitch (2 Times)	30%
Grab	20%	Grab (One Palm)	**35%**

(b) The percentage result of user preferences for abstract

Abstract	**Gesture Referents**			**All**
Gestures	Copy	Paste	Cut	All
Pitch Three Fingers	**90%**	35%	**75%**	**66.7%**
Point Two Hands	5%	**40%**	25%	23.3%
Grab Two Hands	5%	25%	0	10%

(c) The percentage result of user preferences for abstract *TS: Towards Self, SD: Swipe Diagonal, DX: Draw X

Abstract	**Gesture Referents**	**Abstract**	**Gesture Referents**
Gestures	Create	**Gestures**	Destroy
Bloom Up	**90%**	Index Finger SD*	**65%**
Bloom Up TS*	0	Index Finger Dx*	5%
Open Book	10%	Grab and Throw	30%

R_3 investigated user gesture preferences across both rotation and translation categories. Although the results for R_1 and R_2 showed the types of referents within each of the translation and rotation categories did not have any effect on user preferences, the outcome from the FFH Exact test revealed significant differences for the preferences across these categories ($p < 0.001$). As shown in Table 4, the p-value is less than 0.005, indicating that users preferred different gesture types while performing the referents of these categories.

R_4 and R_5 explore whether the act of shrinking or enlarging an object can result in different gesture preferences, for single objects and multiple objects, respectively. The results obtained from the FFH Exact test indicate no significant difference for either single-object or multi-object analyses (P-value = 0.407 for single objects and p-value = 0.44 for multi-objects). This suggests that the act of shrinking or enlarging an object does not affect gesture preferences, whether applied to single-object scaling or multi-object scaling. The participants' choice among the three gestures was more uniform for the referents in the single-object scaling category.

Finally, our last research question, R_6, investigated the effect of copy, paste, and cut referents on the gestures that users preferred (Point, Pitch, and Grab). The outcome of the FFH Exact test produced statistically significant results

Table 4. FFH Exact test result, The value of exact sig. (s-tailed) are written in bold and marked by one star (**) for values less than 0.005. This emphasizes the statistically significant outcome of these research questions

R_1	Value	Exact sig. (2-tailed)	R_2	Value	Exact sig. (2-tailed)
FFH **Exact test**	18.804	0.026	**FFH Exact test**	6.820	0.745
R_3	Value	Exact sig. (2-tailed)	R_4	Value	Exact sig. (2-tailed)
FFH **Exact test**	16.603	**0.00022 ****	**FFH Exact test**	–	0.407
R_5	Value	Exact sig. (2-tailed)	R_6	Value	Exact sig. (2-tailed)
FFH **Exact test**	1.53	0.44	**FFH Exact test**	16.249	**0.00087 ****

($p < 0.001$). Participants had certain different gesture choices for these three referent types.

5 Discussion

By analyzing the gesture preference data collected in the experiment, we learned that participants showed a preference for the "one Palm" gesture followed by the "Two Fingers" gesture for translation gestures. However, in the case of rotation, the "Two Fingers" gesture emerged as the most popular choice. When it came to scaling a single object, the majority of participants favored the "Pitch" gesture. However, for scaling two objects, participants almost equally preferred both the "Pitch Two Hands" and "Grab Two Hands" gestures.

In addition, our findings showed that users preferred simple pantomimic mid-air gestures (i.e., a gesture of performance or imitate for a special task [19], such as the "bloom up" gesture in the create) over two-handed and complex gestures (i.e., "Open book" gesture in the create). This result aligns with the trends observed in prior elicitation studies [5,6].

Participants preferred to use pantomimic gestures for referents with more complex meanings (i.e., create object/ destroy object). These pantomimic gestures enabled users to convey a richer meaning compared to the simple directional gestures observed, where participants traced out a square with their fingertips to create an object.

Furthermore, some users seemed to have a consistent preference among the gesture types. For example, they are more likely to choose the palm gesture for a referent if they have already selected a palm gesture for the earlier referents. This tendency may have been caused by participants wanting to increase the memorability of their set of preferred interactions.

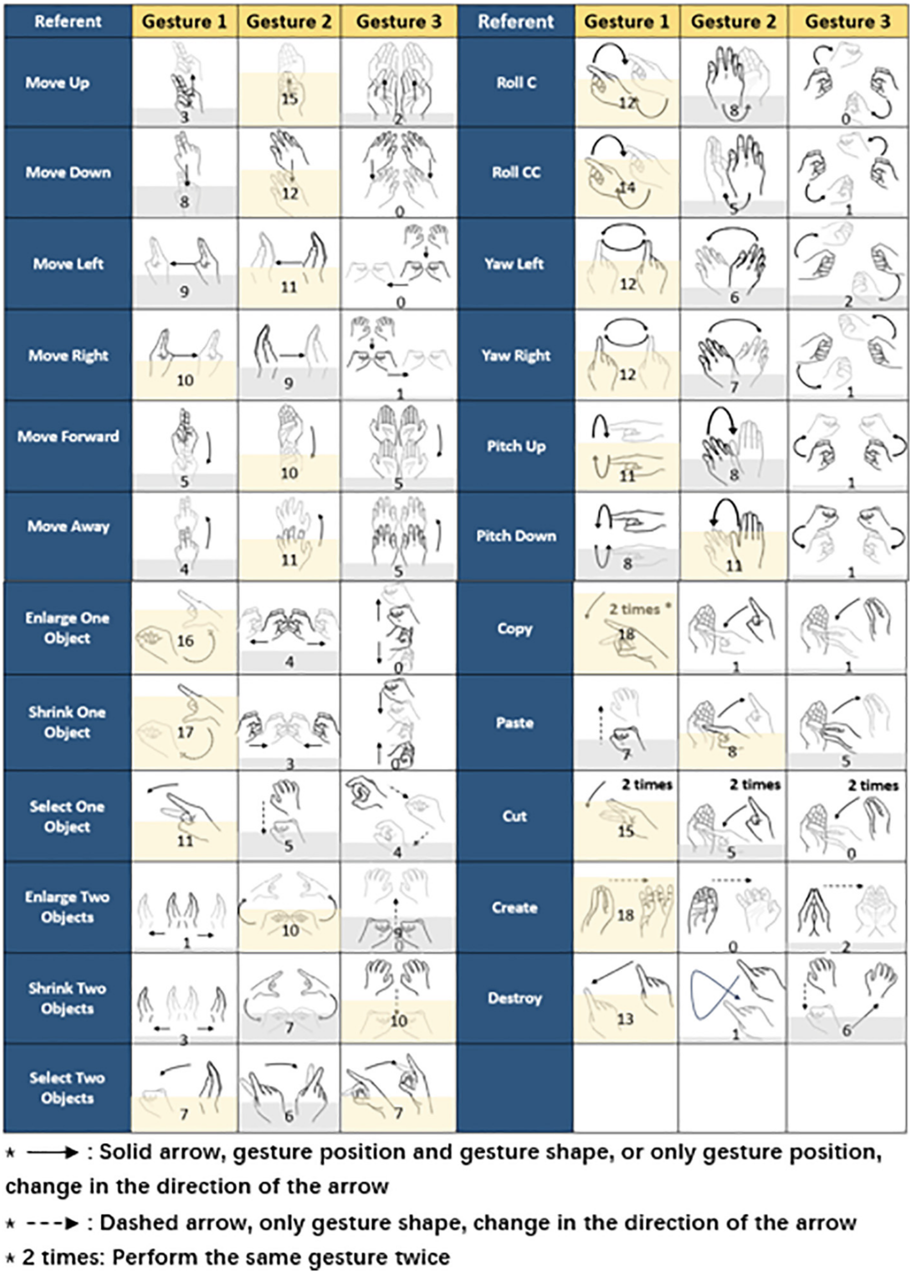

Fig. 6. Gestures schematic diagram for all of gestures with result of preferences. *C: Clockwise, CC: Counter

Our findings recommend following a certain level of uniformity for canonical gesture categories in the gesture design in future systems because users showed similar preferences among the three gesture types for the referents in each of these categories. However, this does not generalize to user gestures in different categories, so researchers need to consider the connection between user preferences and the corresponding category of the referent in future gesture designs.

Figure 6 shows that participants preferred to use the "Pitch Three Fingers" over "Point Two Hands" and "Grab Two Hands" gestures for both Copy and Cut referents. For the paste referent, they did not show any specific preferences among the three gestures. However, if we compare the one-hand gestures to two-hand gestures, our findings showed a preference for two-hand gestures over one-hand gestures (two-hand gestures: 65% and one-hand gestures: 35%). On the other hand, users favored one-hand gestures over two-hand gestures for Copy (one-hand gestures: 90%, two-hand gestures: 10%) and Cut (one-hand gestures: 75% and two-hand gestures: 25%) referents. A possible explanation for this preference may be attributed to the dynamic nature of the paste operation, involving transforming a single object into multiple objects. Participants generally agreed that this transition could be better conveyed using both hands.

6 Conclusion and Future Work

The user-centered mid-air gesture dataset presented here can be used to improve mid-air interactions in AR by improving training quality for real-time machine learning systems that enable gesture recognition. Deep gesture videos and photos contain more gesture information, and the data includes the spatial positions of finger joints, palms, etc., as well as the gaze position of the eyes. It can provide assistance and information supplements for future machine learning applications and analysis. In addition, due to issues such as gesture data sample size and referent classification, this article analyzes the independence between gesture types and referents. Furthermore, the insights from the analysis of participants' preferences can help better equip interaction designers when they seek to develop intuitive systems. One example of such a design guideline is that interactions for translation and rotation referents should be kept simple while abstract referents may benefit from pantomimic gestures. Last but not least, according to the data results, a unified type of gesture can be used for each operation type, providing more design ideas for subsequent gesture design and system development. For future work, we can increase the number of gesture options, increase the sample size, analyze the situation of closed elicitation study under object manipulation, and study more possibilities for users in gesture selection. In addition, we can utilize our existing gesture data information to train better gesture recognition systems.

Acknowledge. This work was supported by the National Science Foundation (NSF) awards NSF 1948254, NSF 2106590, NSF 2016714, NSF 2037417, and NSF 1948254. We would like to acknowledge Jo Carroll for the edits.

References

1. Vatavu, R.-D., Wobbrock, J.O.: Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 1325–1334. Association for Computing Machinery, New York, NY, USA (2015)
2. Morris, M.R., Wobbrock, J.O., Wilson, A.D.: Understanding users' preferences for surface gestures. In: Proceedings of Graphics Interface 2010 (GI '10). Canadian Information Processing Society, CAN, pp. 261–268 (2010)
3. Vatavu, A.-D.: The dissimilarity-consensus approach to agreement analysis in gesture elicitation studies. In: Proceedings of the 2019 CHI Conference on Human Factors in Commputing Systems, CHI '19, pp. 1–13. Association for Computing Machinery, New York, NY, USA (2019)
4. Vogiatzidakis, P., Koutsabasis, P.: Mid-air gesture control of multiple home devices in spatial augmented reality prototype. Multimodal Technol. Interact. **4**, 61 (2020). https://doi.org/10.3390/mti4030061
5. Williams, A.S., Garcia, J., Ortega, F.: Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation. IEEE Trans Visual. Comput. Graph. **26**(12), 3479–3489 (2020). https://doi.org/10.1109/TVCG.2020.3023566
6. Williams, A.S., Ortega, F.R.: Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation. Proc. ACM Hum.-Comput. Interact. **4**, 21 p., 202 (2020). https://doi.org/10.1145/3427330
7. Williams, A.S., Ortega, F.R.: A concise guide to elicitation methodology. Proc. ACM Hum.-Comput. Interact (cs. HC). B. **4**, 62 p., A. 2, Journals CoRR (2021). arXiv:2105.12865, https://arxiv.org/abs/2105.12865
8. Chan, E., Seyed, T., Sturzlinger, W., Yang, X.-D., Maurer, F.: User elicitation on single-hand microgestures. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), pp. 3403–3414. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2858036.2858589
9. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09), pp. 1083–1092. Association for Computing Machinery, New York, NY, USA (2009). https://doi.org/10.1145/1518701.1518866
10. Weiser, M.: The computer for the 21st century. SIGMOBILE Mob. Comput. Commun. Rev. **3**(3), 3–11 (1999). https://doi.org/10.1145/329124.329126
11. Li, X., Guan, D., Zhou, M., Zhang, J., Liu, X., Li, S.: Comparison on user experience of mid-air gesture interaction and traditional remotes control. In: Proceedings of the Seventh International Symposium of Chinese CHI (Chinese CHI '19), pp. 16–22. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3332169.3333570
12. Dewitz, B., Steinicke, F., Geiger, C.: Functional workspace for one-handed tap and swipe microgestures. In: Mensch und Computer 2019 - Workshopband, Bonn. Gesellschaft für Informatik e.V (2019)
13. Koutsabasis, P., Vogiatzidakis, P.: Empirical research in mid-air interaction: a systematic review. Int. J. Hum.–Comput. Interact. **35**, 1747–1768 (2019)
14. Morris, M.R., et al.: Reducing legacy bias in gesture elicitation studies. Interactions **21**(3), , 40–45 (2014). https://doi.org/10.1145/2591689

15. Ungureanu, D., Bogo, F., Microsoft etc.: HoloLens 2 Research Mode as a Tool for Computer Vision Research, Microsoft (2020). https://arxiv.org/abs/2008.11239
16. Freeman, G.H., Halton, J.H.: Note on an exact treatment of contingency, goodness of fit and other problems of significance. Biometrika **38**, 141–149 (1951)
17. Kim, H.Y.: Statistical notes for clinical researchers: chi-squared test and Fisher's exact test. Restor. Dent. Endod. **42**(2), 152–155 (2017). https://doi.org/10.5395/rde.2017.42.2.152. Epub 2017 Mar 30. PMID: 28503482; PMCID: PMC5426219
18. Mehta, C.R., Patel, N.R.: IBM SPSS Exact Tests. IBM Corporation, Armonk, NY (2011)
19. Aigner, R., et al.: Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI (2012)