Gesture Elicitation for 3D Travel via Multi-Touch and Mid-Air Systems for Procedurally Generated Pseudo-Universe

Francisco R. Ortega * Alain Galvan† Katherine Tarre‡ Armando Barreto§ Naphtali Rishe¶
Jonathan Bernal‖ Ruben Balcazar** Jason-Lee Thomas††

Florida International University
Miami, FL.
USA

Abstract

With the introduction of new input devices, a series of questions have been raised in regard to making user interaction more intuitive – in particular, preferred gestures for different tasks. Our study looks into how to find a gesture set for 3D travel using a multi-touch display and a mid-air device to improve user interaction. We conducted a user study with 30 subjects, concluding that users preferred simple gestures for multi-touch. In addition, we found that multi-touch user legacy carried over mid-Air interaction. Finally, we propose a gesture set for both type of interactions.

Index Terms: H.5.2 [User Interfaces]: User-centered design—Gesture Evaluation

1 Introduction

With the introduction of modern input devices such as multi-touch (MT) displays and mid-air (MA) vision-based systems (e.g., Microsoft Kinect and Intel RealSense), the search for the best set of gestures has been undertaken multiple times, starting with Wobbrock et al. who first attempted multi-touch gesture elicitation in 2009 [49]. An important application of gesture-based interaction is 3D navigation, which is divided into two sub-systems: travel and wayfinding [2, 30]. The current study delves into travel, which is the engine of navigation (e.g., a car takes us from point A to point B). In particular, we aim to learn whether it is possible to find a user-driven gesture set for 3D travel, using multi-touch and mid-air interactions with a 3M Multi-Touch display and the Intel RealSense camera for interactive displays.

1.1 Yet another gesture elicitation study?

It is important to understand the motivation for our user study. There have been plenty of studies (see Section 2) concerning gesture elicitation; however, we believe that there is still a need to conduct these studies as there is need to understand different devices and environments. Furthermore, this study uses legacy bias in its favor. This is the reason we conducted an experiment with two different devices and hope that in the future, this may enable us to measure legacy bias. Our motivation also points to the ability of users to interact with complex environment in a desktop setting. The desktop environment is where we spend most of our time, hence the importance of further improving our day-to-day interactions.

Some studies suggest that gestures are best created by experts [48, 11] while others have shown a clear preference for gesture set created by users [49, 24, 4]. While our study looks into finding gestures for 3D travel in a pseudo-universe, we do realize that comparison with expert-designed gestures merits a look and it is proposed as a follow-up study. The study presented here will also provide a framework for follow-up studies which: (i) compare future studies with expert-design gestures using the same environment and (ii) deploy gesture recognition algorithm to find the effectiveness of gesture-set created either by users or experts.

1.2 Contribution

The contribution of our study includes a proposed gesture set, and it arrives with the realization of gesture elicitation for 3D travel using a pseudo-universe, as well as the premise that it may be influenced by legacy bias. As opposed to previous studies, the present study is aimed at 3D travel in pseudo-universe for multi-touch and mid-air systems. It also asks the participants for gestures with different constraints or the lack of them.

An important application of our contribution is 3D travel in unconstrained domains (e.g., the cosmos, complex 3D dataset, or a 3D game). Finding the intuitive gestures for these type of environments provides a way for designers to create more intuitive interactions. We provide recommendations (see §6.2) based on our findings.

1.3 Six Degrees-of-Freedom and Zoom

Translations and rotations on x,y,z axes are commonly known as 6 degrees-of-freedom (DOF). The inclusion of scaling or zooming for a 3D navigation environment has been studied by other colleagues [10, 5] referring to it as 7-DOF. Warren and Holloway describe 7-DOF actions (6-DOF + scale) that provides a framework for a 7-DOF navigation environment for head-mounted displays (HMD) [46]. Our user study provides the additional DOF by adding a field-of-view (FOV) zoom for the user. Therefore, the user does not translate to a point but rather change the zoom of the lenses, providing a different experience. Hence, the use of 7DOF or 6DOF+ is appropriate for this study.

It is also important to address how an extra DOF (zoom) may be used in 3D navigation. By expanding or shrinking the world [46], or in our case, affecting the FOV of the user provides a different visual cue (e.g., finding a target) for 3D navigation (travel and wayfinding) [2, 30]. Additional cues are always recommended, as highlighted by Darken and Sibert [9].

2 Related Work

The primary focus of a gesture elicitation study is to find out whether a gesture set could be derived from the users themselves. A secondary objective is to understand user behavior and how users have potentially evolved since the first study by Wobbrock et al. in

*e-mail: fortega@cs.fiu.edu
†e-mail:agalv023@fiu.edu
‡e-mail:ktarr007@fiu.edu
§email:barretoa@fiu.edu
¶email:ndr@acm.org
‖email:jbern102@fiu.edu
**email:rbalc001@fiu.edu
††email:jasthoma@fiu.edu

2017 IEEE Symposium on 3D User Interfaces (3DUI)
March 18-19, 2017, Los Angeles, CA, USA
978-1-5090-6716-9/17/$31.00

Table 1: Gesture Elicitation Studies

Publication	Spatial			Selection and Manipulation			Travel			Additional Information					
(by reference)	ED	ID	$\mathbb{DOF}$	Selection	$\mathbb{R}.\mathbb{S}.\mathbb{T}.$	Other‡	$\mathbb{R}$	$\mathbb{T}$	Other¶	MT	MA	Type^{E}	Bias	G.A.M.	RC
Our Study	2/3	3	6+				✓	✓	$\mathbb{Z}$	✓	✓	D	✓	✓	14
[1]	2,3	3	6	✓	$\mathbb{R}.\mathbb{T}.$	✓	✓		✓		✓	D			10
[4]	3	2	2†			✓				✓		D	✓	✓	27
[6]	3	2	6	✓	✓					✓		$\mathbb{T}$		✓	21
[7]	2,3	3	6	✓	✓	✓					✓	Ω		✓	22
[22]	2	3	2	✓		✓					✓	⊕		✓	15
[24]	2	2	2†	✓	✓	✓				✓		T		✓	22
[31]	3	3	6	✓	✓	✓					✓	A		✓	40
[33]	3	3	6			✓		✓	✓		✓	M		✓	18
[36]	2	3	2			✓				✓		⊕			16
[44]	2	3	1			✓					✓	Ω		✓	12
[45]	2	3	2+	✓	$\mathbb{T}$	✓			✓		✓	Ω		✓	12
[49]	2	2	2†	✓	✓	✓				✓		T		✓	27

Notes: ED = Environment Dimensions; ID = Input Dimensions; D.O.F. = Degrees of Freedom

†= Participant may have tried to use additional ones; $\mathbb{R}.\mathbb{S}.\mathbb{T}.$ = Rotate, Scale, and Translate; ‡= Commands and other manip. techniques.

$\mathbb{R}$ = Rotate; $\mathbb{T}$ = Translate; ¶= Other travel techniques; $\mathbb{Z}$ = Zoom; Bias = Legacy Bias was considered; RC = Referent Count

E = Environment Type: Augmented Reality (A), Tabletop (T), Desktop (D), Mobile (M), TV (Ω), Multi-Modal (⊕).

G.A.M. = Use gesture agreement technique [49].

Note about Environment Types: Desktop (similar to tabletop) may include environments closer to Virtual or Augmented Reality.

2009 [49] (first described by [28]). Given the year that the experiment was conducted, we understand that the researchers were able to find users who had no experience with multi-touch devices (e.g., iPhone); therefore, they were able to derive a gesture set unaffected by the common use of multi-touch devices. Finding a similar set of participants today is significantly more challenging for multi-touch interaction. Many follow-up studies have created gesture sets using gesture elicitation, such as [6, 4]. The popularity of gesture elicitation can be seen in the variety of studies ranging from deformable displays [19, 42] to multi-touch surfaces [21, 4], and mobile devices [33], among others [6]. Table 1 provides a comparison of some previous gesture elicitation studies and our contribution; we have focused on the most important points of comparison. In relation to spatial considerations we are concerned with the Environment Dimensions (ED), Input Dimensions (ID), and Degrees of Freedom ($\mathbb{DOF}$). We have also considered the general methods of interaction by subdividing sections to Selection, Rotate, Scale, and Translate ($\mathbb{RST}$ symbol), and Other for more general commands and manipulation techniques. In relation to travel we have focused on Rotations ($\mathbb{R}$), Translations ($\mathbb{T}$), and Other travel techniques. For additional considerations we have also taken into account the Type of Environment used with a focus on Augmented Reality (A), Tabletop (T), Desktop (D), Mobile (M), TV (Ω), and Multi-Modal (⊕). We also note if a Gesture Agreement Technique (G.A.M.) is used, as well as if Bias is considered in the experiment and the total Referent Count (RC).

A follow-up paper to [49] concluded that user-created gesture sets performed better than expert-created sets [24]. However, this has created a controversy in the fields of Human-Computer Interaction (HCI) and 3D User Interface (3DUI). Some researchers have interpreted the results to the inexistence of "natural" gestures [29, 30, 13], while others argue that once the technology becomes pervasive, an expert-created gesture set may deliver better results.

It is important to clarify the meaning of the word "natural" in the context of gestures: to us, it means "intuitive" but we avoid using it – the contention for the word "natural" next to interaction or gestures in the HCI community may have its origin in semantics. To some, it means intuitive, while to others it means that the action is inherent in us. The discord may also be traced to the use of the word "natural" for marketing purposes by Steve Ballmer (former Microsoft CEO) [29], as well. However, we believe the intent of our research community is to make systems more intuitive. As Weiser wrote: "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it" [47] – therefore, our study does not aim to discover if gestures are natural or not but rather if we can derive a gesture set from users.

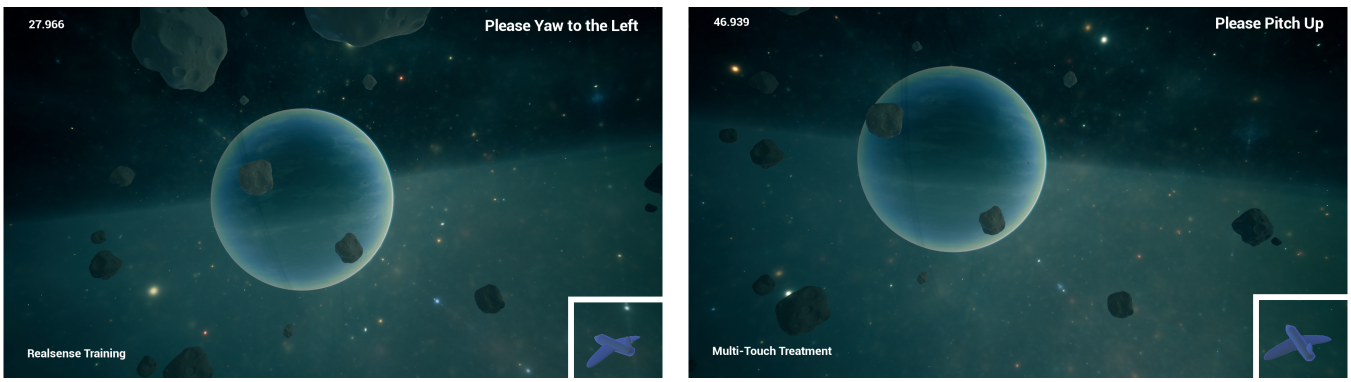
2.1 Legacy Bias

One of the primary concerns with gesture elicitation has been **legacy bias**, stemming from the users coming in with experience from previous interfaces and technologies (e.g,. WIMP). Morris et al. proposed possible steps to reduce legacy bias [23], including production (requiring users to produce multiple interactions), priming (asking users to produce new form factors), and partnering (creating groups to participate in the elicitation) [23]. The approach proposed by [23] was used for whole-body gestures in [34], and another study eliciting mid-Air gestures for music playlists [14].

However, legacy bias can also be used to one's advantage [17]: our experimental design aims to take advantage of legacy bias by creating a gesture set from what already feels "natural" to the user [17]. This will be accomplished through the comparison of gesture sets derived from the multi-touch and mid-air systems. Participants are also presented with the multi-touch and mid-air treatments in a random order to prevent immediate bias from the initial treatment presented to the participant.

(a) Space (b) Participant

Figure 1: Procedurally Generated Pseudo-Universe



(a) mid-Air Training Environment (b) Multi-Touch Treatment Environment

Figure 2: Additional Images of Generated Pseudo-Universe

2.2 3D Navigation

Our gesture evaluation study was designed to find 3D navigation gesture sets. While this is a wizard-of-oz experiment, it is important to understand previous 3D Navigation studies. For example, Santos et al. studied the difference between 3D navigation with a non-stereo, desktop display versus a HMD [40]. Fu et al. studied large-scale 3D astrophysical simulations. Their navigation approach used different gestures and touch widgets to allow different actions in a multi-dimensional environment to study astrophysics. Yu et al. studied touch-based 3D navigation [50]. Their environment was a representation of scientific data, where users navigated using single-touch gestures or a mouse. Their objective was to test a 3D navigation approach with 7-DOF that included translations and rotations on x,y, and z axes with an additional degree of freedom that allows users to scale the environment [50]. Feng et al. studied three types of 3D navigation techniques with 7-DOF [10]. The techniques include Spindle+Wheel – a variant of the Grab-and-Scale method – and One-Hand+Scale. The results showed that the three proposed techniques perform equivalently whenever no scale adjustments are needed. When scaling is needed, the Spindle+Wheel and Grab-and-Scale performed better and were faster than the One-Hand+Scale [10]. Stannus et al. implemented a 7-DOF 3D navigation technique, called AeroSpace, for geo-spatial maps [41]. Additional studies in 7-DOF for interaction (e.g., selection and/or manipulation) have also been studied: for example, Schultheis et al. compared the use of different types of input devices for 3D manipulation in 7-DOF [35]. For additional studies, see [43, 39, 20, 5].

3 Study Environment

Our travel experiment was designed for 7-DOF in a pseudo-universe with planets, asteroids, stars, gases, and other objects. The importance of this environment lies in helping encourage all the rotations and translations needed for a 6-DOF travel experiment. In addition, we added zoom (in and out) referents to the study, amounting to 7-DOF environment. Zooming is performed by adjusting the FOV (see also §1.3). There was a clear difference between zooming in/out compared to translating forward/backwards.

We chose to consider zooming due to this being an unknown environment to the user, and therefore, while not absolutely necessary, the ability to inspect what lies ahead prior to determining a path can prove rather helpful. As a post-experiment survey, we asked five of the original participants if they could differentiate between the translation forwards/backwards and the zoom in/out. All five participants were able to do so.

One of our primary concerns when dealing with 3D rotations is that they can become rather difficult for users, in particular when there is no frame of reference. Previous studies have cited that i) in order to make the required comparison they had to visualize one object rotated into the same orientation as the other and that they could only carry out this "mental rotation" at a limited rate; and (ii) since they perceived the two-dimensional pictures as objects in three-dimensional space, they could imagine the rotation around whichever axis was required with equal ease [38, 37]

The environment was designed in-house to provide a continuous procedurally generated pseudo-universe running under Unreal Engine 4, as shown in Figure 1a. The user experiences the environment from a first person perspective, which allows subjects to cover an "infinite" set of real-state while displaying different planets, stars, and objects. A pseudo-universe provides a generic way to test 3D travel that requires all the rotations and translations found in a 6-DOF system (plus zoom). The participant is shown in Figure 1b. This is ideal for search tasks [2, 30] as our environment will provide an endless set of space for 7-DOF interaction and objects to find. Additional Figures 2a and 2b provide the complete user interface that the participant would see during experiment. This included the referent in text and a visual representation using plane for reference. The plane was animated to represent the current referent.

The environment was generated using a shader applied to a cubemap texture, along with procedural algorithms that generated the asteroid field according to a specified seed that remained constant throughout every experiment. The environment built off the research of [3, 26], in the generation of star fields and planetary atmospheres. Our Unreal Engine 4 version of this Real-Time Celestial Rendering project will be available to download for anyone.

Please note that translation, rotation, or scaling of the world was performed by the experimenter to provide a real-time 3D navigation experience. In other words, this is a wizard-of-oz experiment where the velocity of the gesture did not affect the translation, rotation, or scaling of the environment. However, the translation and rotation moved at a constant speed while the FOV zoom was instantaneous.

3.1 Apparatus

The experiment was run with a procedurally generated pseudo-universe, running on a Windows 10 computer with an NVIDIA GeForce GTX 980M and Intel HD Graphics 530, SSD hard drive, Intel Core i7-6820HK @ 2.70 GHz, and 16 GB. In addition, this included a multi-touch 3M 22" display (M2256PW) and an Intel Real-Sense camera. While this was a gesture elicitation study, we decided to save all the raw data for future analysis. In addition, we used two Raspberry PI with video cameras to record the participants.

4 Gesture Elicitation User Study

We applied a standard elicitation study protocol, where each participant was asked to propose a gesture for each 3D travel task. We sought to determine if there existed a preference (natural or learned) for one-handed or two-handed gestures. For both the multi-touch display and mid-air camera participants, three conditions were presented: unconstrained (which entails creating a gesture as they see fit), one-handed gesture, and two-handed gesture. The unconstrained condition was always presented first to ensure no bias from the constrained conditions. The order in which the devices were presented was randomized to prevent bias from the device initially used. Our hypotheses are as follows:

- H_1: The unconstrained condition will yield a greater number of one-handed gestures for all referents (see [25]. The use of bi-manual interaction has been extensively studied (see [30, §7.3.1 and §8.3.7]). For example, a study of preferred-hand versus non-preferred hand was conducted by [15]. Another reason for the intuition we had for this hypothesis was that it is known that one-handed gestures are better suited for integral tasks (e.g., rotations) [25, 16]

- H_2: The time required by participants to develop gestures will be different for multi-touch and mid-air treatments. Furthermore, we believe that the time taken for the multi-touch gestures will be shorter in comparison to mid-air gestures.

- H_3: Agreement rates will be higher for one-handed gestures in comparison to two-handed gestures, as the former are more commonly used in today's multi-touch systems and will likely appear as a legacy from the use of a trackpad.

- H_4: Legacy from the use of multi-touch will appear in the development of gestures for mid-Air interaction.

4.1 Participants and Design Procedure

Thirty volunteers were recruited, 12 female and 18 males, between the ages of 18 and 46 (mean = 23.4, sd = 4.82). All participants had experience with multi-touch smartphone devices. All participants also reported that they had at least used either the Microsoft Kinect or the Nintendo Wii controller.

Participants were asked to perform instinctive gestures for each referent, explicitly indicating that all gestures would be user-defined. The fourteen referents are listed in Table 2. The information provided to the participants was brief in order to prevent exterior sources of bias. Participants were told the primary objective of the experiment was to evaluate a system which was designed to require minimal – if any – training, that is to say that the gestures built-in to the system should match the first that come to mind for the common user. Note that there were no on-screen controls for the users. However, there were basic instructions on the display, which were also read aloud by the experimenter. Participants were told that gestures were based on-screen contact for multi-touch, and hand-movement in the camera FOV for mid-air. The experimenter used a wireless keyboard to change the state of the application after the user performed a gesture to give the illusion of feedback which was based on a preset constant displacement. All referents were explained to participants using an instructional video which covered all referents in a third person and then a first person perspective which is what was used during the actual experiment. Focus was on the terms: yaw, pitch, and roll. Participant were also informed the third person perspective was only used for instructional purposes and the actual experiment would be conducted using a first person perspective.

Table 2: List of Referents

Ref #	Command	Description
R1	Move Up	Translate $+Z$ axis
R2	Move Down	Translate $-Z$ axis
R3	Move Forward	Translate $+Y$ axis
R4	Move Backward	Translate $-Y$ axis
R5	Move Left	Translate $-X$ axis
R6	Move Right	Translate $+X$ axis
R7	Pitch Up	Rotate on the $+X$ axis
R8	Pitch Down	Rotate on the $-X$ axis
R9	Roll CW†	Rotate on the $+Z$ axis
R10	Roll C-CW‡	Rotate on the $-Z$ axis
R11	Yaw Left	Rotate on the $-Y$ axis
R12	Yaw Right	Rotate on the $+Y$ axis
R13	Zoom In	Field of View $+10°$
R14	Zoom Out	Field of View $-10°$

Table Notes: †: Clockwise; ‡: Counter Clockwise

Participants proposed gestures in two randomized blocks, multi-touch and mid-Air. Each block consisted of three sets. Referents were presented in a random order. First the participant proposed

Table 3: Preferred Gestures with its Referents (bold denotes preferred)

			Multi-Touch		Mid-Air	
	Ref. #	Command	One-Handed	Two-Handed	One-Handed	Two-Handed
Translations	R_1	Move Up	**Swipe Up (I)**	Swipe Up (I,I)	Swipe Up$^{\ddagger\diamond}$	**Swipe Up**$^{\ddagger\diamond}$
	R_2	Move Down	**Swipe Down (I)**	Swipe Down (I,I)	**Swipe Down**$^{\ddagger\diamond}$	Swipe Down$^{\ddagger\diamond}$
	R_3	Move Forward	**Swipe Up (I)**	Swipe Up (I,I)	**Swipe Forward**$^{\ddagger\diamond}$	Swipe Forward$^{\ddagger\diamond}$
	R_4	Move Backward	**Swipe Down (I)**	Swipe Down (I,I)	**Swipe Back**$^{\ddagger\diamond}$	Swipe Back$^{\ddagger\diamond}$
	R_5	Move Left	**Swipe Left (I)**	Swipe Left (I,I)	**Swipe Left**$^{\ddagger\diamond}$	Swipe Left (I,I)
	R_6	Move Right	**Swipe Right (I)**	Swipe Right (I,I)	**Swipe Right**$^{\ddagger\diamond}$	Swipe Right$^{\ddagger\diamond}$
Rotations	R_7	Pitch Up	**Swipe Up (I)**	Swipe Up (I,I)	**Flickwrist Up**$^{\ddagger\diamond}$	Flickwrist Up$^{\ddagger\diamond}$
	R_8	Pitch Down	**Swipe Down (I)**	Swipe Down (I,I)	**Flickwrist Down**$^{\ddagger\diamond}$	Flickwrist Down$^{\ddagger\diamond}$
	R_9	Roll CC	**Rotate CC (I)**	Rotate CC (I,I)	Rotate CC$^{\ddagger\diamond}$	**Rotate CC**$^{\ddagger\diamond}$
	R_{10}	Roll CCW	**Rotate CCW (I)**	Rotate CCW (I,I)	Rotate CCW$^{\ddagger\diamond}$	**Rotate CCW**$^{\ddagger\diamond}$
	R_{11}	Yaw Left	**Swipe Left (I)**	Swipe Left (I,I)	**Flickwrist Left**$^{\ddagger\diamond}$	Flickwrist Left$^{\ddagger\diamond}$
	R_{12}	Yaw Right	**Swipe Right (I)**	Swipe Right (I,I)	**Flickwrist Right**$^{\ddagger\diamond}$	Flickwrist Right$^{\ddagger\diamond}$
Scaling	R_{13}	Zoom In	**Pinch Out (TI)**	Pinch Out (I,I)	Pinch Out‡§	**Pinch Out**‡§
	R_{14}	Zoom Out	**Pinch In (TI)**	Pinch In (I,I)	Pinch In‡§	**Pinch In**‡§

Notes: I = Index Finger; T = Thumb; CW = Clockwise; CCW= Counter Clockwise. ; $^{\ddagger\diamond}$ = Open hand
(F_1,F_2) = Finger left hand, Finger right hand; **Bold Font** = Preferred gesture for device.
§ = Straight hand towards display; $\diamond$ = Palm down.

an unconstrained gesture for each of the 14 referents, followed by a one-handed gesture for each referent, and finally a two-handed gesture for each referent. Referents were presented in the upper-right-hand corner of the screen and read out-loud. The currently tested gesture set was also present in the upper-left-hand corner at all times. With 30 participants creating 3 gesture sets per block, for 14 referents, we collected a total of 2,520 gestures. Each study session lasted approximately 45 minutes.

4.2 Data Analysis

Gestures were evaluated based on the use of left or right hand, subsets of fingers used (for mid-Air this was defined as the extended fingers used for pointing, for multi-touch as fingers making contact with the display), and direction of movement. Classification was completed using a set of rules developed by us with the aid of previous publications, such as [49, 24, 4, 12, 1, 27, 31]

All gestures were first classified by two independent reviewers. Gesture classifications by the primary raters were analyzed for agreement, and disagreements were sent to a third independent coder for review. Primary raters had an agreement rate of 85.99% (2167 out 2520 gestures). The remaining 353 gestures were sent to a third rater who in all cases agreed with the classification of one of two primary raters. Agreement rate between rater three and the primary raters was 41.93% (148 out of 353 gestures), and 58.07% (205 out of 353) for raters one and two.

5 Results

Out of 840 gestures in the unconstrained condition, 187 (22.3%) required the use of both hands, 631 (75.1%) used the right hand, and the remaining 22 (2.6%) used the left hand. The number of proposed left-handed gestures is consistent with the number of left-handed participants, which is 2. The unconstrained condition demonstrates a clear preference for one-handed gestures, supporting H_1; however, we must note that this preference is most apparent when using the multi-touch system. The use of one and two-handed gestures is more evenly distributed for the mid-air system. Based on the information gathered from participants after the experiment when using the multi-touch system, twenty-five out of thirty (83.3%) participants preferred one-handed gestures. However, when using mid-air system, only seventeen out of thirty (56.7%) participants preferred one-handed gestures. Note that participants who preferred one-handed gestures overall also preferred two-handed gestures for zoom and roll referents. Overall, participants preferred navigating our 3D environment using the mid-air system (22 out of 30 or 73.3%). Reasons for this preference included the system being more "intuitive", more "immersive", "allowed for more natural movements from previous use of motion devices" and "was easier to visualize the movements by using my hands to complete the referent motions". Those who preferred the multi-touch cited familiarity with similar media and one-handed gestures being more intuitive.

5.1 Multi-Touch

The multi-touch block, as previously mentioned, consisted of 3 sets, each containing the 14 referents in a randomized order. The unconstrained condition was presented first, to yield the most instinctual gesture, be it using one or two hands. By performing unconstrained elicitation first, we also prevented bias from a preferred gesture under both the one-handed and two-handed constraints, and fall back to – what we consider to be – less reasoned gestures. For this portion of the experiment the multi-touch display was placed 30 cm from the edge of the desk, which was found to be the most comfortable distance for ease of use.

The gesture set for multi-touch systems also proved most difficult, as gestures were oftentimes repetitive – particularly between translation and rotation referents. The distribution of gestures for "move up" (R_1) and "pitch up" (R_7) were identical, with "swipe up" using the index finger being by far the most common gesture used; this was also the case for referents R_2 and R_8, R_5 and R_{11} and R_6 and R_{12} (see Table 3). The same pattern was noted with two-handed gestures, where the only change entailed using the index fingers on both hands in unison. Participants noted that in the space available, one-handed gestures were more comfortable and easier to perform. For referents Roll and Zoom, a more even number of one-handed and two-handed gestures were used. It is apparent that participants required a greater number of constraints in order to develop more

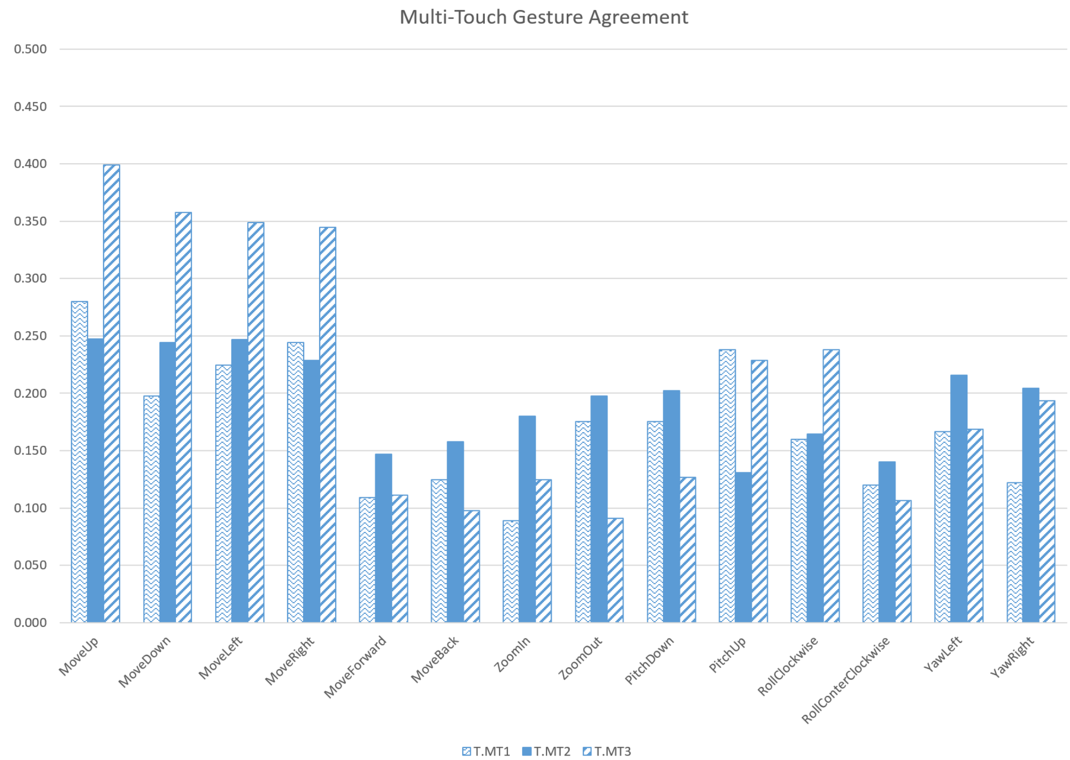


Figure 3: Multi-Touch Agreement

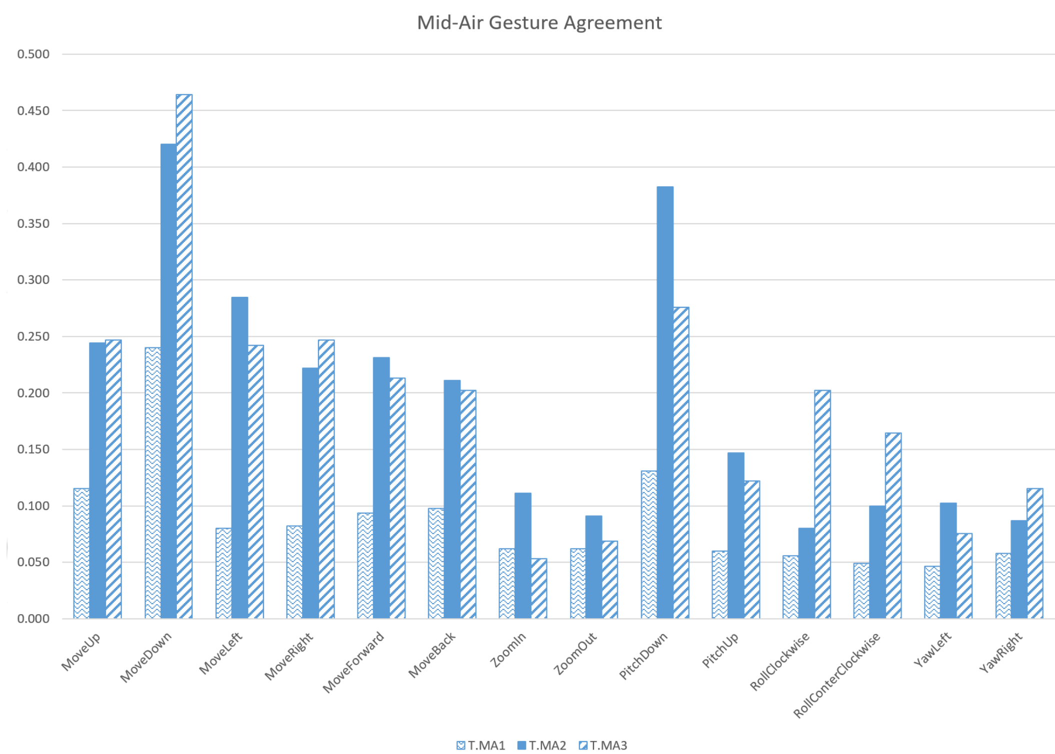


Figure 4: Mid-Air Agreement

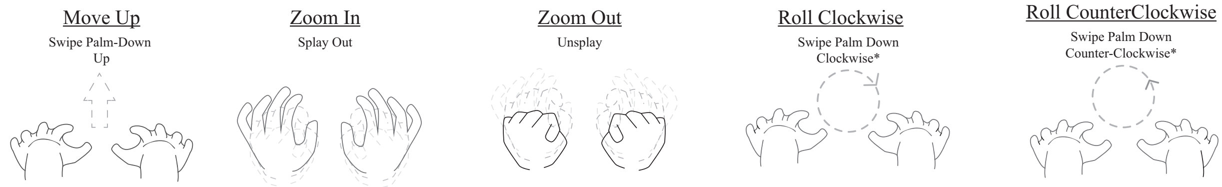


Figure 5: Mid-Air Two-Hands Gesture Set (also see Figure 6)

unique gestures. However, it is also important to mention the links between translations and rotations, as when seen from a first person perspective, the effect is fairly similar and this could be the primary cause of the confusion experienced by the users.

5.2 Mid-Air

The gesture set for the mid-air system showed greater variation than that for the multi-touch display. When using the mid-air system, participants displayed an even use of one-handed and two-handed gestures, which led to the creation of a bi-manual gesture set, unlike that of the multi-touch. The primary difference between translations and rotations was that translations were characterized by full arm movements in the appropriate direction, while rotations displayed a preference for what is referred to as a "Flick-wrist" (see Table 3) – in relation to the movement of the wrist. Visually, a flick-wrist is similar to the motions of a plane, in comparison to the rotational referents. In using the mid-air system, we also found a preference for two-handed gestures for the referent Roll, in the form of a rotation – meaning moving both hands to create a circular object; as well as for Zoom, in the form of a pinch by spreading hands apart or bringing them together. There was no clear preference between one and two-handed gestures for the pairs Move Up/Down, Move Forward/Back, and Pitch Up/Down therefore the overall most repeated gesture is reported. However, it is important to note the difference in use for the most repeated gesture and second most repeated gesture is one or two participants which does not represent a significant difference. For the remaining gestures, we still see a prevalent use of the right hand, which is consistent with the results of the multi-touch system.

5.3 Gesture Agreement

To further understand the interrelation of gestures between participants, we adopted Wobbrock et al.'s definition of agreement and calculated the agreement rates for individual referents[49], as can be seen in Figures 3 and 4 (where MA = mid-air; MT = Multi-Touch).

$$A(r) = \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|} \right)^2 \tag{1}$$

For multi-touch gestures, the agreement rates are approximately the same for similar gestures, e.g., "move up" and "move down". Overall agreement level is also highest for basic translations such as "move up" (.399 for two-handed condition), "move down" (0.358 for two-handed condition), "move left" and "move right" in comparison to rotational referents.

mid-Air gestures display a similar pattern to the multi-touch gestures, with the exception of the correlation between "move up" and "move down" and "pitch up" and "pitch down". The agreement level of the down movements is nearly double that of the up referents for the same translation and rotation. The highest levels of agreement are also present for "move down" (0.464 for the two-handed condition) and "pitch down" (0.382 for the one-handed condition).

Overall average agreement rates for the one-handed treatments are nearly identical for multi-touch and mid-air systems (0.193 and 0.194, respectively). The average agreement rate for two-handed gestures is higher than that of the one-handed gestures for the multi-touch block, and nearly identical for the mid-air. We can therefore conclude that H_3 is not supported, due to the similarity of the agreement rates.

A recurring pattern throughout our data is the prevalent use of index fingers when using a multi-touch display, while using a full hand for mid-air interaction. For multi-touch both one- and two-handed gestures were primarily composed of using the index finger for contact with the display. On the other hand, mid-air gestures used a full hand palm-down, or palm-straight to a lesser degree. These preferences support the development of simplistic gestures demonstrated by the users.

5.4 Gesture Set

While (H_3) agreement was not significant, we were still able to derive a gesture set with the most commonly used gestures. However, there are overlaps between translation and rotations (swipe gestures) for multi-touch. In Table 3, we provide a gesture for each referent per device. For each device, we listed the most common one-hand gesture and the most common two-hand gesture. Bolded is the preferred gesture between one and two hands for each device. Figure 5 shows the two-handed gestures that were preferred. A graphical representation for one-handed gestures for the unconstrained mode is shown in Figure 6 (for additional information see §6.1). The referents in which two-handed gestures were preferred represent variants of the one handed gestures and are therefore not represented in Figure 6 for simplicity (and are marked with an asterisk on the figure).

From these results there exists an evident overlap between the gestures developed for multi-touch and mid-air. This occurrence may be attributed to: a) the existence of an instinctual sense to navigate as in the real-world (use of a motion in the direction requested), or b) transferability of gestures to mid-air as a result of legacy from multi-touch devices (see H_4).

5.5 Gesture Completion Time

We expected participants to require less time to create a gesture set for multi-touch systems, as they are exposed to them on a daily basis. The time data for each block was tested for normality, using the Shapiro-Wilk Normality Test. For MT data the p-value was 0.60 and for mid-air the p-value was 0.12. The data does not show a severe deviation from normality. Once this was established, we performed a paired t-test to determine if the times were significantly different. The true difference of means was equal to 0 at a 95% level of significance ($t = -0.8494$, $df = 29$, p-value $= 0.40$) and H_2 was therefore not supported. The mean time for multi-touch was 282 seconds in comparison to 296 seconds for mid-Air.

6 Discussion

Our results show that participants had a preference for one-handed gestures, particularly in the use of multi-touch display (H_1 supported). However, the influence of legacy bias is clearly evident in

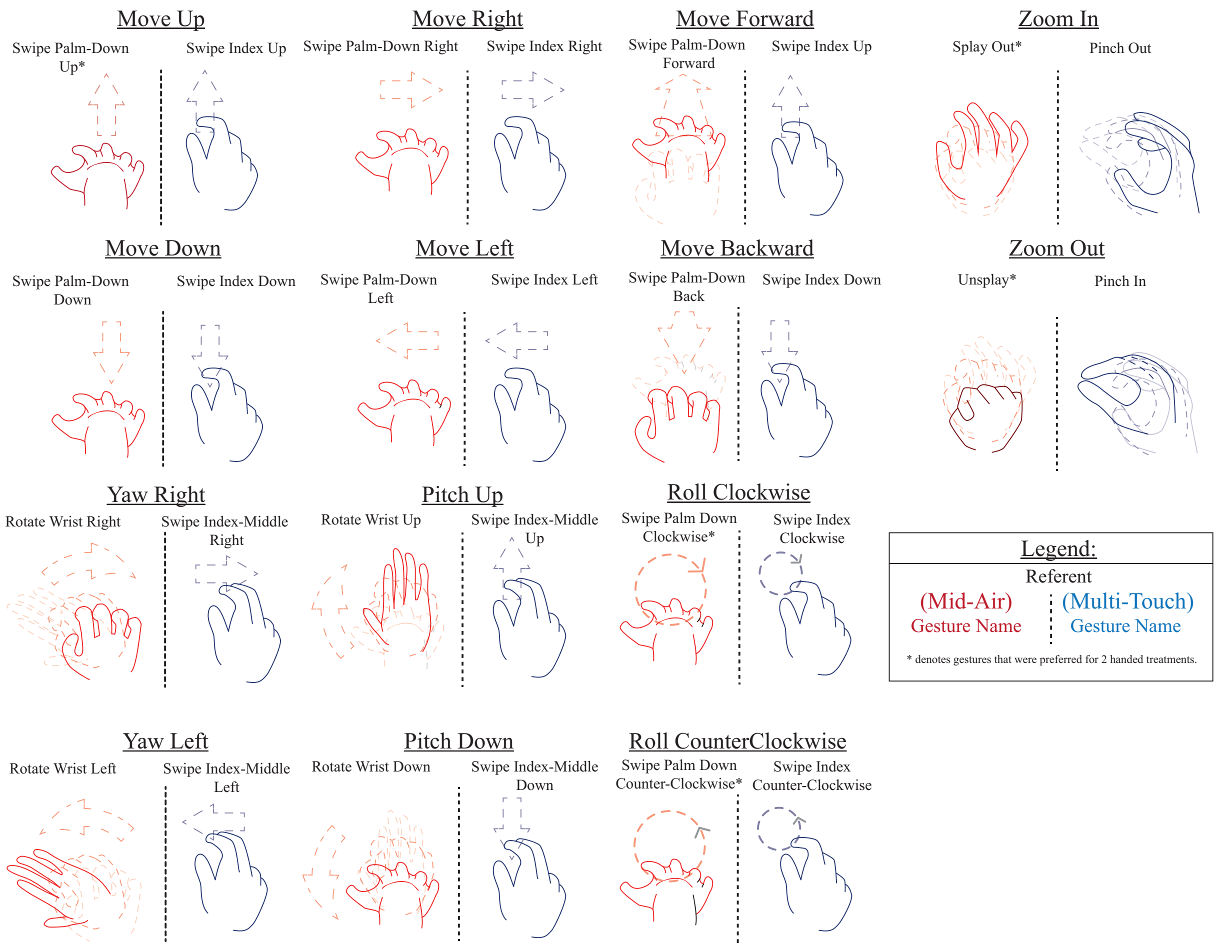


Figure 6: Gesture Set Derived from Unconstrained Treatments

the prevalent use of simple swipe gestures to complete a high proportion of the referents for both multi-touch and mid-air Systems. While this pattern opens up the possibility of eliciting a "better" gesture set, if steps are taken to limit legacy bias, it is important to remember that even if this is a previously existing gesture, it is not necessarily "natural" for the user. Köpsel and Bubalo highlight that "even if legacy bias is acquired through training, it can still be used to inform rules for creating new interfaces" [17]. A clear example of this is the use of the QWERTY keyboard, a legacy of the typewriter era that remains in use. We have interpreted this to mean that by taking advantage of what already feels familiar and "natural" to the user we can achieve the creation of a functional user-generated gesture set [17].

6.1 Unconstrained Gesture Set

The unconstrained gesture set (MT.1 and MA.1) was used to develop a final gesture set, which we present in Figure 6. While it is true that five of the mid-air gestures displayed preferences for two-handed: Swipe Up, Rotate CC, Rotate CCW, Pinch Out, and Pinch In (R_1, R_9, R_{10}, R_{13}, and R_{14}) – see Table 3. One question is – why Swipe Up was preferred with two-hands while Swipe Down was not? The reason is that it was the result of a technical tie (off by 1). Therefore, it is inconclusive. Note that Figure 6 shows only one-handed gestures because the five preferred gestures for mid-air that use two hands were identical in movement.

Another question is – why the graphical representation does not show two hands? Note that Figure 6 shows only one-handed gestures because the five preferred gestures for mid-air that use two hands were identical in movement. Nevertheless, Figure 5 provides the two-handed graphical representation for completeness. Finally, it is important to note that the difference in multi-touch was more definitive than mid-air where in certain cases, the preference was a technical tie between one and two-handed gestures; therefore, the overall most repeated gesture is reported.

6.2 Gesture Design Recommendations

After further analyzing the gestures collected, it becomes evident that agreement rates among treatments differ between conditions. Looking at the multi-touch treatment, it is clear that the difference in agreement rates between conditions can be attributed to the type of referent requested. For Move referents (Up, Down, Left and Right), the two-handed constraint condition provides significantly higher levels of agreements, as do the Roll referents (Clockwise and Counter-Clockwise). For other referents, the difference is either not significant, or the one-handed constraint has slightly higher agreement rates. In the mid-air treatment, agreement rates for unconstrained gestures are significantly lower than constrained gestures, both one and two-handed. Based on this evidence (as shown in Figures 3 and 4), multi-touch does not require constraints, because of the limited non-symbolic gestures. Conversely, mid-air gestures require some constraints, in order for participants to design similar gestures. In addition, the exit questionnaire, as well as the data collected, shows that users prefer simple gestures for multi-touch and mid-air (having many more options for mid-air interaction).

6.3 Limitations

This study has certain limitations: first, participants' legacy bias may have severely influenced their primary gestures. In particular, most users performed gestures adapted from multi-touch during the mid-air treatment, regardless of whether multi-touch was presented before or after the mid-air treatments. In follow-up studies, we may be able to evaluate the effects of legacy bias by having participants create a set of gestures for each referent, and have them select their preferred gestures from that subset. In addition, employing some recommendations by Morris et al. will help measure the legacy bias from users [23]; second, the use of multiple devices may have negatively influenced the participants' decision to transfer gestures from one device to the next. Concentrating on one device may yield better results for that specific medium. Finally, the use of the pseudo-space environment may have resulted in undue disorientation for the participants; for example, navigation in 3D environments is still a difficult task because the environment and user perception play an important role. To support this claim, various experiments have been conducted: Kozhevnikov and Dhond explored ways in which users rotate 3D images and in their findings, the authors concluded that navigation in 3D environments is influenced by the users frame of reference, e.g., egocentric vs allocentric [8], as well as by the objects around them [18]; Preus et al. conducted an experiment with a focus on perception processing times. They concluded that small changes in the rotation angle can greatly influence the allocation of the self in complicated environments [32].

7 Future Work

A modified experimental design would allow participants to develop a gesture set composed of unique gestures for each referent, followed by a sequence of requests to use the newly designed gestures in order to evaluate memorability and ease of use for said gestures. Participants could also be asked to create multiple gestures per referent in a team setting, which could act as a counter-balance against legacy bias. By allowing the participants to develop a full gesture set prior to the start of the experimental tests, we also would provide them with more time to consider unique gestures, removing the pressure of time as a consideration.

8 Conclusion

This study delved into 3D travel gesture elicitation for multi-touch and mid-air interactions, to find a gesture set for a series of 7-DOF referents. We found that unconstrained condition yielded a greater amount of one-handed gestures, more so for the multi-touch display; however, we were not able to support the time difference between multi-touch and mid-air systems for the completion of all tasks. In addition, agreement rates for one-handed and two-handed gestures were not significantly different. We also found that users transfer their multi-touch experience to the mid-air system. We believe that users preferred simple gestures for multi-touch, such as one- or two-finger gestures with one hand rather than complex ones, as stated in previous studies [25, 16].

While this study has limitations (see §6.3), the proposed gestures set and findings transfer to 7 DOF studies. First, as described in 3D navigation section 2.2, current research derived gestures from existing techniques. While this approach is valid, it doesn't take into account that some gestures may serve users better than others, but only provides comparison with other techniques. Our gesture-set provides a good approximation for large environments using the user-driven approach. Second, the fact that we use legacy bias in our favor provides alternate design guidelines for future researchers dealing with 7-DOF. Third, a clear take away from our studies is that users prefer simpler gestures whenever possible, as it was shown when using the multi-touch display. While it is up to the designer to choose between expert-driven gestures or user-driven gestures, we believe that the latter, with further study, may produce favorable results. In order to accomplish this, future work is needed, as described in Section 7.

Acknowledgements

We acknowledge the support by the National Science Foundation under Grant Nos I/UCRC IIP-1338922, III-Large IIS-1213026, MRI CNS-1429345, MRI CNS-1532061, MRI CNS-1532061, MRI CNS-1429345, RAPID CNS-1507611, and U.S. DOT Grant ARI73. We also acknowledge Lukas Borges for his continuous support during this study. Finally, Vesna Barbarogic for the time she spent proof-reading this document.

References

[1] R. Aigner, D. Wigdor, H. Benko, and M. Haller. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Technical report, 2012.

[2] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2004.

[3] E. Bruneton and F. Neyret. Precomputed atmospheric scattering. *Computer Graphics Forum*, 27(4):1079–1086, 2008.

[4] S. Buchanan, B. Floyd, W. Holderness, and J. J. LaViola. Towards user-defined multi-touch gestures for 3D objects. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*. ACM, 2013.

[5] I. Cho and Z. Wartell. Evaluation of a bimanual simultaneous 7dof interaction technique in virtual environments. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 133–136, March 2015.

[6] A. Cohé and M. Hachet. Understanding user gestures for manipulating 3D objects from touchscreen inputs. In *GI '12: Proceedings of Graphics Interface 2012*. CIPS, 2012.

[7] S. Connell, P.-Y. Kuo, L. Liu, and A. M. Piper. A wizard-of-oz elicitation study examining child-defined gestures with a whole-body interface. In *Proceedings of the 12th International Conference on Interaction Design and Children*, IDC '13, pages 277–280, New York, NY, USA, 2013. ACM.

[8] R. Darken, T. Allard, and L. Achille. Spatial Orientation and Wayfinding in Large-Scale Virtual Spaces: An Introduction. *Presence*, 7(2):101–107, 1998.

[9] R. P. Darken and J. L. Sibert. Wayfinding strategies and behaviors in large virtual worlds. In *CHI '96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 142–149, New York, New York, USA, Apr. 1996. ACM.

[10] J. Feng, I. Cho, and Z. Wartell. Comparison of device-based, one and two-handed 7dof manipulation techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, SUI '15, pages 2–9, New York, NY, USA, 2015. ACM.

[11] D. Freeman, H. Benko, M. R. Morris, and D. Wigdor. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. pages 165–172, 2009.

[12] A. M. Genest, C. Gutwin, A. Tang, M. Kalyn, and Z. Ivkovic. KinectArms: a toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. In *CSCW '13: Proceedings of the 2013 conference on Computer supported cooperative work*, pages 157–166, New York, New York, USA, 2013. ACM.

[13] K. Hinckley and D. Widgor. Input Technologies and Techniques. In J. A. Jacko, editor, *The Human-Computer Interaction*, pages 95–132. CRC Press, 2012.

[14] L. Hoff, E. Hornecker, and S. Bertel. Modifying Gesture Elicitation: Do Kinaesthetic Priming and Increased Production Reduce Legacy Bias? In *TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 2016.

[15] P. Kabbash, W. Buxton, and A. Sellen. *Two-handed input in a compound task*. ACM, New York, New York, USA, 1994.

[16] K. Kin, M. Agrawala, and T. DeRose. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *Proceedings of Graphics Interface 2009*, GI '09, pages 119–124. CIPS, 2009.

[17] A. Köpsel and N. Bubalo. Benefiting from legacy bias. *interactions*, 22(5):44–47, 2015.

[18] M. Kozhevnikov and R. P. Dhond. Understanding immersivity: image generation and transformation processes in 3D immersive environments. *Frontiers in psychology*, 2012.

[19] B. Lahey, A. Girouard, W. Burleson, and R. Vertegaal. PaperPhone: understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011.

[20] D. Mendes, F. Fonseca, B. Araujo, A. Ferreira, and J. Jorge. Mid-air interactions above stereoscopic interactive tables. In *IEEE Symposium on 3D User Interfaces*, pages 3–10, March 2014.

[21] M. Micire, M. Desai, A. Courtemanche, K. M. Tsui, and H. A. Yanco. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In *ITS '09: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM, 2009.

[22] M. R. Morris. Web on the wall: Insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, pages 95–104, New York, NY, USA, 2012. ACM.

[23] M. R. Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, m. c. schraefel, and J. O. Wobbrock. Reducing legacy bias in gesture elicitation studies. *interactions*, 21(3), 2014.

[24] M. R. Morris, J. O. Wobbrock, and A. D. Wilson. Understanding users' preferences for surface gestures. In *GI '10: Proceedings of Graphics Interface 2010*. CIPS, 2010.

[25] T. Moscovich and J. Hughes. Indirect mappings of multi-touch input using one and two hands. CHI '08, pages 1275–1284. ACM, 2008.

[26] D. Mller, J. Engel, and J. Dllner. Single-Pass Rendering of Day and Night Sky Phenomena. In M. Goesele, T. Grosch, H. Theisel, K. Toennies, and B. Preim, editors, *Vision, Modeling and Visualization*. The Eurographics Association, 2012.

[27] M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay. Mid-air pan-and-zoom on wall-sized displays. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011.

[28] M. Nielsen, M. Störring, and T. B. Moeslund. A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction. In *Proc. of Int. Gesture Workshop 2003*, 2003.

[29] D. A. Norman. Natural user interfaces are not natural. *interactions*, 2010.

[30] F. R. Ortega, F. Abyarjoo, A. Barreto, N. Rishe, and M. Adjouadi. *Interaction Design for 3D User Interfaces*. The World of Modern Input Devices for Research, Applications, and Game Development. CRC Press, 2016.

[31] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. *CHI EA '13: CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pages 955–960, 2013.

[32] N. Preuss, L. R. Harris, and F. W. Mast. Allocentric visual cues influence mental transformation of bodies. *Journal of vision*, 2013.

[33] J. Ruiz, Y. Li, and E. Lank. User-defined motion gestures for mobile interaction. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011.

[34] J. Ruiz and D. Vogel. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015.

[35] U. Schultheis, J. Jerald, F. Toledo, A. Yoganandan, and P. Mlyniec. Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3d tasks. In *IEEE Symposium on 3D User Interfaces*, pages 117–124, 2012.

[36] T. Seyed, C. Burns, M. Costa Sousa, F. Maurer, and A. Tang. Eliciting usable gestures for multi-display environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, pages 41–50, New York, NY, USA, 2012. ACM.

[37] R. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *The Philosophy of Mind: Classical Problems/contemporary Issues*, pages 218–221, 1992.

[38] S. Shepard and D. Metzler. Mental rotation: effects of dimensionality of objects and type of task. *Journal of Experimental Psychology*, 1988.

[39] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1297–1306, New York, NY, USA, 2012. ACM.

[40] B. Sousa Santos, P. Dias, A. Pimentel, J.-W. Baggerman, C. Ferreira, S. Silva, and J. Madeira. Head-mounted display versus desktop for 3D navigation in virtual reality: a user study. *Multimedia Tools and Applications*, 41(1):161–181, Aug. 2008.

[41] S. Stannus, A. Lucieer, and W.-T. Fu. Natural 7dof navigation & interaction in 3d geovisualisations. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, VRST '14, pages 229–230, New York, NY, USA, 2014. ACM.

[42] G. M. Troiano, E. W. Pedersen, and K. Hornbæk. User-defined gestures for elastic, deformable displays. In *AVI '14: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. ACM, 2014.

[43] A. Ulinski, C. Zanbaka, Z. Wartell, P. Goolkasian, and L. F. Hodges. Two handed selection techniques for volumetric data. In *IEEE Symposium on 3D User Interfaces*, March 2007.

[44] R.-D. Vatavu. User-defined gestures for free-hand tv control. In *Proceedings of the 10th European Conference on Interactive Tv and Video*, EuroiTV '12, pages 45–48, New York, NY, USA, 2012. ACM.

[45] R.-D. Vatavu. There's a world outside your tv: Exploring interactions beyond the physical tv screen. In *Proceedings of the 11th European Conference on Interactive TV and Video*, EuroITV '13, pages 143–152, New York, NY, USA, 2013. ACM.

[46] R. Warren and R. Holloway. Implementation of flying, scaling and grabbing in virtual worlds. In *Symposium on Interactive 3D Graphics*, pages 189–193. ACM, March 1992.

[47] M. Weiser. The computer for the 21st century. *Scientific American*, 272(3):78–89, 1995.

[48] D. Wigdor and D. Wixon. *Brave NUI World*. Designing Natural User Interfaces for Touch and Gesture. Elsevier, Apr. 2011.

[49] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009.

[50] L. Yu, P. Svetachov, P. Isenberg, M. H. Everts, and T. Isenberg. FI3D: Direct-Touch Interaction for the Exploration of 3D Scientific Visualization Spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1613–1622, 2010.