Experimental Analysis of Bare Hand Mid-air Mode-Switching Techniques in Virtual Reality

Hemant Bhaskar Surale
Cheriton School of Computer Science
University of Waterloo, Canada
hsurale@uwaterloo.ca

Fabrice Matulic
Preferred Networks Inc.
Japan
fmatulic@preferred.jp

Daniel Vogel
Cheriton School of Computer Science
University of Waterloo, Canada
dvogel@uwaterloo.ca

ABSTRACT

We present an empirical comparison of eleven bare hand, mid-air mode-switching techniques suitable for virtual reality in two experiments. The first evaluates seven techniques spanning dominant and non-dominant hand actions. Techniques represent common classes of actions selected by a methodical examination of 56 examples of prior art. The standard "subtraction method" protocol is adapted for 3D interfaces, with two baseline selection methods, bare hand pinch and device controller button. A second experiment with four techniques explores more subtle dominant-hand techniques and the effect of using a dominant hand device for selection. Results provide guidance to practitioners when choosing bare hand, mid-air mode-switching techniques, and for researchers when designing new mode-switching methods in VR.

KEYWORDS

interaction techniques, controlled experiments

ACM Reference Format:
Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental Analysis of Bare Hand Mid-air Mode-Switching Techniques in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300426

1 INTRODUCTION

Raskin defines a *mode* as a distinct setting within an interface where the same user input produces results different from those it would produce in other settings [68]. *Mode-switching* is simply the transition from one mode to another. Modes are common in all interfaces, including interfaces for Virtual Reality (VR) and Augmented Reality (AR). For example, a hand gesture-based 3D modelling application may have different modes for object creation, selection, and transformation. Depending on the mode, the movement of the hand is interpreted differently. Completing a task typically requires frequent mode-switching, so understanding the performance of different methods is important.

In VR, bare hand mid-air input is an alternative to device controllers. Techniques have been proposed for VR, AR, and related contexts using hands only (e.g. [61, 64, 93]) and hands combined with body postures (e.g. [12, 100]). Evaluations have focused on tasks like pointing (e.g. [99]), object manipulation (e.g. [69]), selection (e.g. [59]), and annotation [16], but no extensive comparisons of mode-switching techniques have been performed yet. Mode-switching techniques for mice [18], styli [48, 88], and touch [83] have been evaluated, but generalizing those results to 3D environments like VR is not straightforward.

We provide missing empirical evidence for the performance of bare hand mid-air mode-switching techniques in VR. Our focus is absolute, single-point input, suitable for the kind of direct object manipulations common in VR such as pointing at, grabbing and moving 3D elements in the virtual environment. To select techniques to evaluate, we examined bare hand mid-air interaction in different settings, then used three criteria to identify six classes of techniques suitable for mode-switching in VR. In two related experiments, we compare common input actions selected from each class using an adapted "subtraction method" protocol [18], used previously for 2D input.

The first experiment compares seven techniques, with a dominant-hand pinch as the fundamental manipulation trigger. The mode-switching techniques include three dominant hand postures: a fist, an open palm and pointing the index finger; and four non-dominant hand postures: a fist, an open palm, bringing the hand into the field-of-view and touching the head. As a comparison baseline, the button of a device controller held in the non-dominant hand was also

ACM ISBN 978-1-4503-5970-2/19/05.
https://doi.org/10.1145/3290605.3300426

included in the tests. A second experiment explores questions emerging from the results. The effect of more subtle dominant hand techniques is examined by testing pinching with wrist rotation, and pinching with different fingers; the effect of pinching as a manipulation trigger is compared with a controller held in the dominant hand.

Our empirically-derived insights can inform the design of VR applications using bare hand mode switching:

1. Dominant techniques using large motions are error prone and less preferred, but a more subtle variation of pinch is comparable to the fastest non-dominant techniques.
2. With the exception of a few dominant techniques, mode-switching times are comparable to most touch methods.
3. Using a dominant pinch as a manipulation trigger is comparable to using a device controller button.
4. All techniques from fastest to slowest: Non-Dominant (ND) device and Dominant (D) middle finger pinch; ND palm orientation and ND fist; ND head touch, D pinch orientation and D palm; ND field-of-view and D fist; D point.

2 BACKGROUND AND RELATED WORK

We define *bare hand* as input performed entirely by a hand posture or movement, without any device. Note that this definition does not specify the sensing method, so early work using instrumented gloves for hand tracking are considered bare hand for our purposes. We define *mid-air* as input conducted without contacting a non-body surface. In most cases, this means input performed in the space around the body. For on-body contacts, the sensing method is unimportant as long as the technique is conceptually a body contact (e.g. touching the head), and not using a device attached to the body (e.g. tapping on a smartwatch).

Formal Mode-Switching Evaluations. Mode-switching has been more commonly studied in 2D interfaces. Dillon et al. [18] introduced a formal “subtraction method” for comparing mouse mode-switching, which was adopted by Li et al. [48] and Tu et al. [88], who each compared five mode-switching techniques, and Surale et al. [83], who compared six techniques. Using the non-dominant hand for stylus mode-switching has been studied in detail by Ruiz et al. [71] who developed a temporal model, Lank et al. [44] who showed concurrent mode-switching is fastest, and Ruiz and Lank [72] who explored related aspects with multiple modes.

Related Evaluations of Interaction Techniques. For mid-air bare hand input, mode-switching has only been indirectly evaluated as part of larger interaction technique studies. In the context of large displays, Vogel and Balakrishnan [99] compare a relative pointing technique, which uses a fist mode-switch to "clutch", to a ray-cast technique without any mode-switch, but the mode-switch itself is not compared. Similar examples in large display research include Haque et al. [26], Polacek et al. [65], Jota et al. [33], and Katsuragawa et al. [38]. In the context of VR and related 3D contexts, Poupyrev et al. [66, 67] evaluated object pointing, manipulation, and selection techniques, Teather and Stuerzlinger compared pointing techniques [86], and Vanacken et al. [91], Grossman et al. [25], and Looser et al. [52] all examine bare-hand selection techniques. In most cases, these techniques have some explicit activation and deactivation of a mode, but mode-switching performance is not evaluated in isolation.

We are unaware of work comparing mode-switching techniques for bare hand mid-air input in VR using the formal subtraction method used for mouse, pen, and touch.

3 MODE-SWITCHING TECHNIQUES

To identify bare hand mid-air interaction techniques suitable for VR mode-switching, we examined research and commercial systems to create a list of candidate techniques.

Our examination included general surveys of 3D interaction from Argelaguet and Andujar [3], Jung et al. [35], Poupyrev et al. [67], Bowman et al. [13], Aigner et al. [2] and Groenewald et al. [24]. We also examined the results of elicitation studies for bare hand mid-air interaction with large displays [51, 73, 92, 103], general ubiquitous computing [15] and AR [64]. Finally, we found examples of bare hand mid-air techniques in many papers on interaction techniques (e.g. [80, 99]), new sensors (e.g. [40]), and commercial devices (e.g. Leap Motion [45], Kinect [42], Myo [57]). For this work we did not consider other non-hand input that may also be suitable for mode-switching, such as feet (e.g. [95]), voice (e.g. [11]), gaze (e.g. [62]), or exocentric interaction (e.g. [82]).

From our initial list of bare hand mid-air interaction techniques gathered from the literature we extracted a subset of candidates that we considered s suitable for mode-switching based on three filtering criteria.

Actions Suitable for Mode-Switching

We identified 40 different mid-air bare hand actions in 56 publications or device manuals. We initially considered actions suitable for dominant or non-dominant hands, even if the source considers a specific hand. To filter those actions to a subset suitable for mode-switching techniques, we created three criteria based on observable mechanics of hand or finger actions. These criteria are not explicitly about performance, because no previous work specifically evaluated mode-switching, nor are they about learnability, comfort, and social acceptability, because these aspects are often not reported. The filtering criteria are:

Independent — A mode-switch action should be fast to recognize and independent of previous tracking states, meaning it should not rely on time-based actions such as a specific

Table 1: Bare-hand mid-air interaction techniques suitable for mode-switching in VR based on examination of research papers and commercial systems. Similar *Actions* are grouped into *Classes*. The number of sources (*N*) is an approximate indication of popularity. Boldface actions are those tested in our experiments.

Class	N	Actions (with sources)
Pinch Finger(s)	29	**thumb touches index** [9, 15, 19, 29, 36, 40, 46, 56, 61, 63, 64, 77, 92, 100, 106]; thumb touches all fingers [6, 15, 34, 61, 64, 93, 94, 104]; thumb touches side of hand [50, 90, 99, 106]; thumb touches middle [15, 56, 77]; thumb touches index and middle [34, 77]; thumb touches ring [15, 56]; thumb touches pinky [15, 56]; thumb touches index, middle, and ring [15]; thumb touches ring and pinky [15]; thumb touches three fingers [64]
Extend Finger(s)	31	**extend index** [19, 20, 40, 45, 50, 51, 61, 64, 70, 80, 90, 94, 99, 101, 106]; extend thumb [30, 43, 51, 61, 64, 92, 93, 104, 106]; extend thumb-index-middle [20, 50]; clench index [90]; two hand point [64, 70, 80]; point with dwell [90]
Close Hand	18	**make fist** [15, 17, 20, 28, 30, 34, 40, 56, 61, 64, 69, 78, 80, 89, 90, 99, 101, 104]; make partial fist [19, 99]
Open Hand	62	**open hand with oriented palm** in/out [1, 7, 17, 19–22, 27–29, 37, 41, 42, 47, 51, 58, 59, 61, 64, 78, 81, 93, 94, 96–98, 101, 103–106], up/down [19, 28, 30, 41, 43, 59, 61, 70, 94, 96, 105, 106], right/left [17, 20, 22, 27, 30, 41, 94, 97, 103, 106]; open hand [1, 40, 80, 93, 98, 99]; open hand with finger(s) bent [90, 99]
Raise Hand (ND)	6	**hand raised into field-of-view** [41, 84, 93, 96]; raised above shoulder [81, 106]
Touch Body (ND)	6	**finger(s) touch head** [27, 100, 106], behind ear [51]; mouth [19]; hand touches waist [100]

movements. Conceptually, this means the technique can be recognized in a single sensor time frame (in implementations, it may actually be a few frames to compensate for noise). Examples of independent actions include pinching the thumb and finger together, making a fist, or raising an arm. Examples of non-independent actions include dwelling, gestures like drawing an ‘X’, or mimicking knocking.

Kinesthetic — The action should enable the mode to be maintained by the user, not the system. This means the posture, position, or gesture action can be “held” as long as the mode is needed, and the mode ends when no longer held. This creates a *kinesthetic quasimode* [68], known to reduce mode errors [76]. Examples of kinesthetic actions include pinching, making a fist, or a repeated gesture, where changing the posture or stopping the gesture releases the mode.

Unconstrained — The action can be executed with the dominant hand at any position and, in the case of a kinesthetic action, the dominant hand can easily reach any position for subsequent operations. Unconstrained actions include any non-dominant hand action that does not impede the dominant hand, and dominant hand actions such as pinching or making a fist. Examples of constrained actions are placing the dominant hand on the head, or pointing the dominant index finger towards the body, since those actions physically constrain the motion range of the arm.

We considered candidate actions suitable for mode-switching techniques when they satisfied all three criteria. This narrowed the list down to 29 actions found in 53 papers, listed in Table 1. Analogous actions are grouped into 6 classes. For example, all actions that involve a pinch of some kind are grouped into the general “Pinch Finger(s)” class. Most actions can be used for mode-switching with either hand, except those in the “Raise Hand” and “Touch Body” classes that are only suitable for non-dominant usage.

Selected Techniques for Evaluation

We selected the most popular action from each class to evaluate (boldface in Table 1). Five of these form eight mode-switching techniques because some actions are performed with both dominant and non-dominant hand.

With bare hand mid-air input, a mechanism is required for the user to indicate when their hand is just moving, or if it is performing an operation such as drawing a line. Buxton’s *Simple 2-State Transaction Model* [14] calls these states “tracking” and “dragging”. With a VR device controller or a mouse, this is achieved by pressing a button. For our bare hand mid-air system, we use a thumb-index pinch since it is a popular action, and it uses a subtle movement that can be easily sensed. We believe using a pinch is the most obvious choice, but other possibilities include tapping the palm with the thumb [15] and a partial finger bend or “Airtap” [99].

Non-Dominant Techniques. The non-dominant hand controls the mode and the dominant hand manipulates using the thumb-index pinch. The techniques are:

- *Non-Dominant Fist* (ND-FIST) — The mode is active when the non-dominant hand is clenched, and released when the hand relaxes so one or more fingers begin to open.
- *Non-Dominant Palm* (ND-PALM) — The mode is active when all fingers of the non-dominant hand are extended, roughly pointing up, with the palm facing to the right (assuming the left hand is non-dominant). The mode is released when

Figure 1: Selected Mode-switching techniques: (a) non dominant fist (ND-FIST); (b) non dominant palm (ND-PALM); (c) hand in field of view (ND-FOV); (d) touch head (ND-HEAD); (e) dominant fist (D-FIST); (f) dominant palm (D-PALM); (g) point (D-POINT) .

the palm orientation changes significantly or one or more fingers are no longer extended.

- *Hand Moved into Field of View* (ND-FOV) — The mode is active when the non-dominant hand is moved into the user's field of view, and released when it moves outside.
- *Touch Head* (ND-HEAD) — The mode is active when the non-dominant hand touches the side of the head (HMD).

Dominant Hand (D) Techniques. The dominant hand controls the mode by forming a posture, and a thumb pinch against the side of the index finger engages and disengages the "dragging" state for manipulation. The techniques are:

- *Dominant Fist* (D-FIST) — The mode is active when the dominant hand is clenched, and the mode is released when the hand relaxes, i.e. when one or more fingers begin to open.
- *Dominant Palm* (D-PALM) — The mode is active when all the fingers are extended, roughly pointing up, and the palm is facing away.
- *Point* (D-POINT) — The mode is active when the index finger is extended and all remaining fingers are closed.

Technique Sensing

To track the user's hands in the VR environment, we use a LEAP motion mounted on the front of the HMD [4, 49, 75]. We found the LEAP reliable for rendering the hand and tracking hand position, but the built-in posture recognizer often misclassified dominant fist, palm, and pinch postures during rapid mode switching, and detection of pinch actions to trigger manipulation was imprecise. To address these issues, we developed a user-calibrated classifier to discriminate between problematic postures, and added a force-sensitive resistor (FSR) to detect pinches and head touches.

User-calibrated Posture Classifier. In the experiment, we only need to independently discriminate between palm and pinch, or between fist and pinch. We train a simple classifier for each case, calibrated to each user. We describe the method using fist and pinch, but both are similar. A user draws 15 lines each with the pinch, palm, and fist postures. At each frame, a ten-dimensional feature vector consisting of fingertip positions and distances relative to the centre of the palm is recorded. The median and standard deviation (SD) in each dimension is calculated for the 20 frames before and after the start of each line. The classifier selects differentiating dimensions in which the fist median plus two SD is less than the pinch median minus two SD. In the rare case that two differentiating dimensions are not found, training is repeated with the user instructed to form the postures more clearly. Otherwise, each selected dimension is assigned a threshold half-way between the two medians. Once trained, a posture was considered a pinch when two or more selected dimensions exceeded the threshold. A 6-person pilot test found this method almost 99% accurate. During the experiment, this calibration process was conducted immediately before the first block testing dominant fist or palm.

Pinch and Touch Detection. An FSR taped to the distal phalanx of the dominant thumb is used to detect thumb-index pinches. The FSR is 7 mm in diameter and 0.3 mm thick with a sensing range of 0 to 175 psi. To detect a touch to the head, a larger FSR (25.4 mm diameter, 0.21 mm thick) is taped to the non-dominant side of the HMD. We used a threshold of 17 psi for both FSRs to detect a light pinch or touch. Thin wires from each FSR were connected to an Arduino Nano (Atmega328) mounted off the body. We verified that the sensors and wires did not impede the user's movements.

System Latency and Performance. Our code and 3D scene were simple and we used a high-end computer, so the Unity application ran at an optimal 90 FPS to supply the 90Hz HMD. The Leap motion provided a stream of hand postures at 110Hz, the Arduino updated pressure values approximately every 2ms, and the controllers were tracked at 250Hz to 1kHz. Our posture classifier did not use any temporal filtering. Since cycle duration is measured between two input times, the effective latency is 11ms at 90 FPS.

4 EXPERIMENT 1

The goal is to empirically compare mode-switching performance of the seven techniques listed above. We adapt the standard "subtraction method" [18, 48, 83, 88] protocol to VR. This determines the precise cost of mode-switching by subtracting the time to perform tasks using a single mode, from those when alternating between two modes.

Participants and Apparatus

16 participants (mean age 28.5 SD = 6.3, 5 women, all right-handed) were recruited. 7 had experience using a VR device. Remuneration was $10.

An HTC Vive VR head-mounted display (HMD), with a resolution of 1080 × 1200 px per eye, 90Hz refresh rate, and 110° FOV was used. Focal length was initially set to 63 mm, but participants were given the possibility to adjust it. A high-end Windows 10 machine (3.6GHz Intel i7 CPU, GeForce GTX 1080 GPU) ran the experiment application written in Unity 5.5.3f1. A LEAP motion and two FSRs were attached as described above.

Task

We adapt the 2D task used in previous mode-switching investigations [18, 48, 83] to a 3D task for VR. Considering that a common class of VR consumer applications are for sketching, painting, and 3D modeling (e.g. [8, 10, 23, 55, 60, 60, 79, 87]), with much prior research in these areas [5, 31, 39, 74, 85], we use 3D line drawing as our fundamental task. Note that this is an abstraction of many 3D tasks, such as creating objects other than lines (e.g. cubes, spheres), transforming an object (e.g. moving, scaling), or panning a world scene. Regardless, for the purposes of our experiment, the explicit mode change is more important than the gross movement of the hand during the task. All tasks were performed in a standing position.

Line Drawing. A series of five aligned pairs of spheres is presented to the user in the VR world (Figure 2). The spheres in a pair need to be connected by drawing a 3D line between them in the proper mode. Each sphere has diameter 100 mm, the distance between spheres in a pair is 50 mm, and pairs are stacked with 82 mm overlap. The position is calibrated so the topmost pair is at the participant's shoulder level. Sizes and distances among the pairs are chosen so that participants can easily reach each sphere without stepping, regardless of pair orientation or movement direction. We chose the centre of the palm as the line "ink" anchor as it is more stable than the fingers, which are used for mode-switching.

For each pair of spheres, the line drawing task can be seen as a four-step process. First, place the dominant hand inside the starting sphere (a change of opacity indicates the centre of the palm is inside). Second, engage line drawing using the pinch trigger. Third, move the hand to draw the line until it is inside the second sphere. Lastly, release the pinch to disengage line drawing. Subtle audio feedback ('tick') indicates line drawing engagement and disengagement. If engagement or disengagement occurs outside the sphere, an error is logged, a buzz sounds, and the participant has to redraw the line. This is repeated for all five pairs of spheres. The current pair is indicated using colour saturation, and the required direction of the line drawing is indicated through transparency (most opaque sphere to semitransparent sphere).

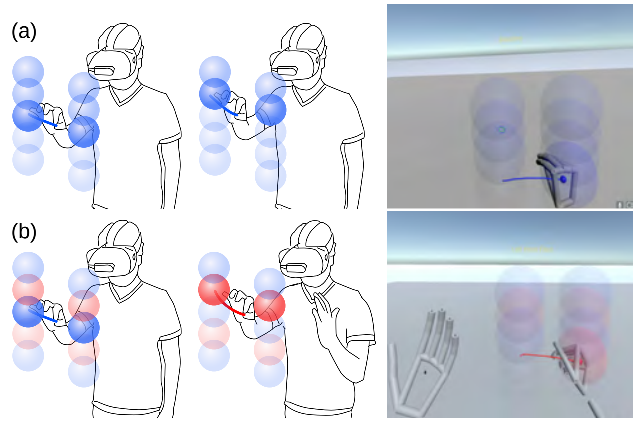

Figure 2: Line drawing task: (a) baseline; (b) compound.

Baseline and Compound Task Variations. There are two task variations. In the *baseline* task, five pairs of blue spheres are shown and the participant draws lines using only the dominant hand pinch. In the *compound* task, the five pairs alternate between blue and red colours, with the first pair being blue. The participant must draw lines to connect a pair of blue coloured spheres using a pinch, and red paired spheres with the specified mode switching technique. To switch modes, the participant must have formed the current mode-switching technique posture at the moment line drawing is engaged. As visual feedback, "moded" lines drawn with the mode-switching technique are red, and "unmoded" lines are blue. These red or blue trails function only as an abstract representation of two different modes. The colour of a small sphere rendered in the middle of the palm also indicates the mode. If there is a mode-switch error, a buzz sounds and the participant has to redraw the current line.

Design and Procedure

The experiment is a within subjects design. Mode-switching TECHNIQUE is the primary factor, with levels corresponding to the seven techniques described above (ND-FIST, ND-PALM, ND-FOV, ND-HEAD, N-FIST, D-PALM, D-POINT) and an eighth technique using a standard HTC controller with a button held in the non-dominant hand (ND-DEVICE). This functions as a non-bare hand comparison baseline since the mode is switched by holding the button. Like all other non-dominant hand techniques, the dominant hand draws the line using a pinch. Our block design deviates slightly from previous mode-switching studies [48, 83] in the following ways.

Pilot Study to Refine Design. Previous mode-switching studies used 9 blocks: 5 baseline (B) task blocks separated by 4 compound (C) task blocks (i.e. BCBCBCBCB). Considering that we wished to test 8 techniques (previous mode-switching studies compared 5 or 6), and mid-air gestures are more fatiguing than pen or touch interactions, we conducted a 4-person pilot test to see if some blocks could possibly be removed without impacting the results too much. Each session took more than 90 minutes. Examining a graph of task times by block for each technique gave no indication of pronounced systematic differences between any of the baseline blocks, and we did not find any significant interaction for TECHNIQUE × BLOCK. As a result, we reduce the number of baseline blocks from 5 to 2 by keeping only the first and the last baseline blocks (i.e. BCCCCB). The pilot study also tested 6 directions for the drawing task: left and right horizontal, up and down vertical, and in and out along the depth direction. A graph of time by task direction did not reveal any major differences, and we found no significant effect, so we reduced the directions to 3 for the main study: horizontal left, vertical down, and depth inward. With the optimized experimental design, all eight techniques can be tested in less than an hour with less fatigue.

Final Design. To minimize order effects, TECHNIQUE was counter-balanced using a 8 × 8 Latin Square. The session began with 5 baseline blocks for training. Then, for each TECHNIQUE, there was a 3 min practice period followed by 6 BLOCKS of tasks. The starting and ending block were *baseline* tasks, remaining blocks were *compound* tasks (i.e. BCCCCB). In each block, the participant had to complete 5 line drawing tasks in 3 directions (left-to-right, up-to-down, out-to-in) in random order. Breaks were encouraged between blocks. After the experiment, participants provided subjective ratings for each technique.

In sum there were: 8 TECHNIQUES (7 mid-air bare hand, 1 device) × 6 BLOCKS (2 *baseline*, 4 *compound*) × 3 directions × 5 line drawings = 720 line drawings per participant.

Dependent Measures

Mode Switching Time. The line drawing time is the duration starting from line engagement in the starting sphere until line disengagement in the ending sphere. Naively, one might directly compare line drawing times in a baseline block to a compound block (where a mode switch was required). However, all line drawings share a common overhead of moving from the ending sphere of the previous line to the starting sphere of the current line. Therefore, we use the "subtraction method" [18, 83] to precisely isolate mode-switching time.

In a set of line drawing tasks, there are three "cycles". The first is from the moment spheres are visible until the pinch trigger is released after drawing the line between the first pair of spheres. The second cycle begins immediately after, and ends when the pinch trigger is released after drawing the line between the third pair of spheres. The third cycle begins immediately after, ending when the pinch trigger is released after drawing the line between the last pair of spheres. During the compound task, the second and third cycles are *full cycles*. Each captures a complete mode-switch operation: the participant switches into a mode using the specified technique, draws a line connecting spheres, switches out of the mode, draws a line connecting another set of spheres, and disengages line drawing mode. The first cycle only guarantees the baseline mode is active before the second cycle. In each block, there are 6 full cycles (2 per direction). For each technique, each participant completes 24 full cycles with mode switching (in 4 compound task blocks) and 12 full cycles with no mode switching (in 2 baseline task blocks).

The subtraction method isolates mode switching time using mean times from cycles. For each block, the mean time for the second and third cycle is calculated using error-free cycles. The mode-switching time is calculated by subtracting the mean full cycle time of the first and the last baseline blocks from the mean full cycle time of a compound block. In total, this provides 8 *mode-switch time* measurements per participant, per technique (2 per block).

Errors and Error Rates. Like Li et al. and Surale et al., we identify three error types. A *start error* occurs if the line drawing is initiated outside the active starting sphere. An *end error* occurs if the line did not connect the two active spheres in the correct direction. A *mode error* occurs when the wrong mode is used to connect the spheres in compound tasks (e.g. connecting blue spheres with a red line). We further distinguish between mode-in and mode-out errors. A *mode-in error* occurs when the participant fails to transition from baseline to the specified mode-switching technique. A *mode-out error* is when the participant fails to transition from the specified mode-switching technique to baseline. Each of these error types are recorded per trial as 1 if the error occurred, and 0 otherwise. The mean value across trials produces an error rate.

Results

We examined error-free *line drawing times* to identify outliers more than 3 standard deviations from the mean for each TASK. This removed 1% of the line drawing trials. Using the remaining error-free full cycles, we used the subtraction method explained above to calculate *mode-switch times*. Visual inspection of the mode-switch time distribution suggested non-normality, confirmed by a Shapiro-Wilk test. To compensate, all mode-switching time data points were log transformed. There were 29 data points with slightly negative mode-switching times, primarily for the fastest techniques

ND-PALM and ND-DEVICE. To compensate, we added 440 ms to all times to guarantee positive values required by the log transform. Note that log transformed data is only used for statistical tests, all reported times are actual measured values.

Analysis Method. For mode-switching time and error rates, we performed a TECHNIQUE × BLOCK repeated measures ANOVA. Pairwise t-tests with Bonferroni corrections are used. For interaction effects, we restrict pairwise tests to comparing means across factor dimensions independently. When the sphericity condition is violated, degrees of freedom were adjusted using Greenhouse-Geisser ($\epsilon < 0.75$) or Huynh-Feldt ($\epsilon \geq 0.75$) corrections.

Learning Effects. To determine if performance changed during the four compound blocks, we tested for effects of TECHNIQUE × BLOCK on *mode-switch time* and *combined error rate* (one measure capturing whether any error occurred). Also, to determine if performance changed between the start and end baseline blocks, we tested effects of TECHNIQUE ×BLOCK on *cycle time* of baseline blocks only. We found no statistically significant interactions involving block in any of these test, indicating no or minimal learning effects. This matches the performance stability noted by Li et al. [48] and Surale et al. [83] All blocks are used in subsequent analyses, with BLOCK used as a repeated measure.

Mode Switch Time. Most non-dominant techniques were faster than dominant ones, with the device controller being the fastest (Figure 3). There was a significant main effect of TECHNIQUE ($F_{7,105}$ = 4.93, p < .0001, η^2_G = .23). Post hoc tests found ND-DEVICE faster than all dominant techniques (D-FIST, D-POINT, and D-PALM) (all p < .001), and two non-dominant techniques (ND-HEAD, ND-FOV) (both $p < 0.05$). ND-PALM was faster than D-PALM (p < .05), D-FIST (p < .0001), D-POINT (p < .001), and ND-FOV (p < .01). Furthermore, ND-FIST was faster than D-FIST and D-POINT (both p < .05). No significant differences were among the dominant hand techniques. Ranking TECHNIQUES from fastest to slowest measured mode-switch time: ND-DEVICE (331 ms), ND-PALM (459 ms), ND-FIST (513 ms), ND-HEAD (528 ms), D-PALM (729 ms), ND-FOV (793 ms), D-FIST (846 ms), and D-POINT (955 ms).

Overall Error Rate. Overall error rates were between 5.3% and 18.3%, with dominant techniques more error-prone than non-dominant techniques. Non-dominant hand mid-air techniques were not significantly different from using a device controller. There is a significant main effect of TECHNIQUE on *overall error rate* ($F_{7,105}$ = 5.95, p < .0001, η^2_G = .28). Post-hoc tests revealed that the rate for ND-DEVICE (5.3%) was lower than that of D-POINT (18.3%), D-PALM (13.6%), and D-FIST (15.3%) (all p < .001). The D-POINT error rate was higher than that of ND-HEAD (7.8%), ND-PALM (8.0%), ND-FOV (9.2%) (all p < .001), and ND-FIST (9.5%) (p < .05). Finally, D-FIST also had a higher error rate than ND-FOV, ND-HEAD, and ND-PALM (all p < .05). Note that the overall error rate for the baseline task is 8%, so rates between 5.3% to 18.3% with mode-switching are quite similar. Error rates in this range are in line with reported error rates for mid-air interactions [53, 54].

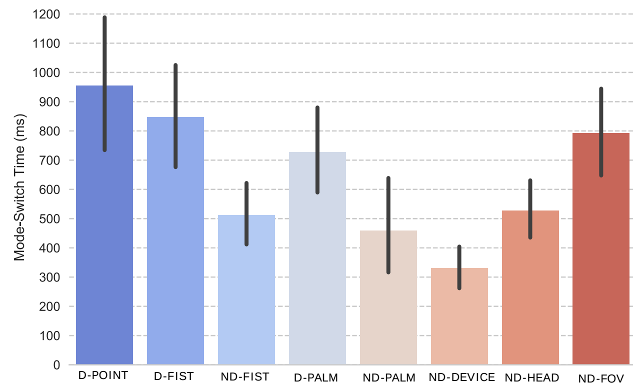

Figure 3: *Mode-switch time* by TECHNIQUE (error bars in all graphs are 95% CI).

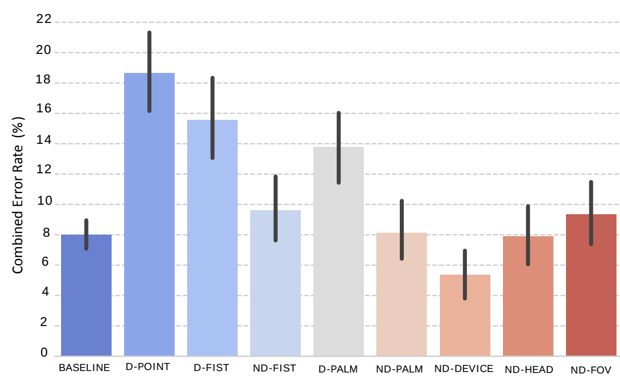

Figure 4: *Overall error rate* by TECHNIQUE.

Start and End Error Rates. start error rate was the largest contributor to *overall error rate* (Figure 5). Measured rates were between 3.7% and 14.7%, with dominant techniques more error prone. A main effect of TECHNIQUE on *start error rate* ($F_{7,105}$ = 5.26, p < .0001, η^2_G = .26) with post hoc tests showed ND-DEVICE (3.7%) was more robust than D-FIST (12%), D-POINT (14.7%) (both p < .001), and D-PALM (9%) (p < .05). Error rates of ND-FIST (7.6%) (p < .05), ND-FOV (7.4%), ND-HEAD (6.6%), and ND-PALM (6.2%) (p < .001) were significantly lower than D-POINT (14.7%), and the ND-PALM rate was lower than D-FIST (all p < .001). For *end error rate*, we did not find significant effect of TECHNIQUE, BLOCK, or their interaction. Measured rates were between 2.2% and 6.1%.

Mode-In and Mode-Out Error Rates. For *mode-in error rate*, Non-dominant hand techniques were comparable to using a

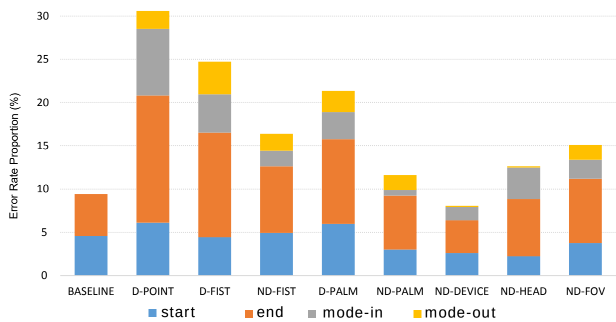

Figure 5: Proportion of specific error rates by TECHNIQUE. Note multiple error types can occur in a cycle.

controller, with dominant fist and palm slightly more error-prone. Dominant pointing had high mode-in errors. All rates were between 0.6% and 7.6%. There is a main effect of TECHNIQUE on *mode-in error rate* ($F_{3.79,56.92} = 6.10$, $p < .0001$, $\eta_G^2 = .29$). Post hoc tests showed D-POINT had a higher rate (7.6%) than all other techniques: D-PALM (3.1%), ND-DEVICE (1.5%), ND-FIST (1.8%), ND-FOV (2.2%), ND-HEAD (3.6%), and ND-PALM (0.6%) (all $p < .05$). Moreover, the ND-PALM rate was lower than D-FIST (4.4%), D-PALM, and ND-HEAD (all $p < .05$).

Except D-FIST, most of the techniques had *mode-out error rates* below 2.4%, suggesting that mode disengagement was a minor contributor toward *overall error rate*. Measured rates were between 0.1% and 3.7%. There is a main effect of TECHNIQUE on *mode-out error rate* ($F_{4.13,62} = 3.28$, $p < .01$, $\eta_G^2 = .18$). Post hoc tests showed ND-DEVICE (0.01%) and ND-HEAD (0.01%) had a lower error rate than ND-FIST (1.9%), D-FIST (3.7%), and D-PALM (2.4%) (all $p < .05$).

Subjective Ratings. After the experiment, we asked participants to rate each technique with respect to six aspects: ease-of-learning, ease-of-use, accuracy, speed, eye fatigue and hand fatigue. A 5-point continuous scale was used, with 1 being the worst score (e.g. low accuracy, hard to learn, very fatiguing) and 5 being the best (e.g. high accuracy, easy to learn, not fatiguing).

Table 2 (a) summarizes the ratings. The distribution of ratings was non-normal due to high negative skewness, so values were transformed using Aligned Rank Transform [102]. ANOVAs performed on transformed data revealed significant main effects of TECHNIQUE on all the aspects except eye fatigue. The main results of pairwise comparisons between TECHNIQUE for each rating are:

- For **hand fatigue**, D-PALM ($p = .049$) and ND-FOV ($p < .03$) were perceived as more fatiguing than ND-DEVICE ($p < .05$). However, for the remaining techniques, reported fatigue levels were not significantly different. This indicates that fatigue levels for the overall experiment were low.
- For *speed* and *accuracy*, D-POINT and D-PALM were rated significantly slower and less accurate than ND-DEVICE ($p < .01$). For the remaining techniques, there was no significant difference in terms of speed. ND-DEVICE was perceived as being more accurate than all the dominant techniques ($p < .001$) and some non-dominant techniques, ND-FIST and ND-FOV ($p < .05$). The remaining techniques, ND-HEAD and ND-PALM were rated as being more accurate than D-POINT and D-PALM ($p < .05$).
- For *ease-of-use* and *ease-of-learning*, D-PALM was rated worse than all non-dominant techniques ($p < .05$), followed by D-POINT, which was rated worse than all non-dominant techniques ($p < .05$) except ND-FOV. D-FIST was perceived as being harder to learn than ND-DEVICE and ND-PALM. ND-DEVICE was easier to use compared to all the dominant hand techniques ($p < .01$).

Summary

Two non-dominant techniques, forming a fist or palm, were not significantly different from using a controller device, and had error rates comparable to the un-moded baseline task. Yet, almost half of the participants picked the controller as most preferred. To understand how the pinch engagement trigger affects mode-switching, the next experiment evaluates a dominant device controller as a trigger.

Overall, non-dominant techniques are generally faster and less error prone than dominant ones. However, using only a dominant hand would free the other hand, and in theory, this should reduce fatigue. So why did dominant techniques perform relatively poorly? Participant comments suggest some confusion: *"Difficult to change the modes using the same hand. It is frustrating when it recognizes hand inaccurately"* [P3], *"I feel Index, Fist, and Pinch are all same"* [P7]. When choosing fist, palm, and point for dominant techniques, we felt these would reduce confusion since they are very different from the un-moded pinch trigger. Instead, using very different actions seem to have increased confusion, and perhaps introduced a time penalty for larger finger movements required to switch between pinching and the mode-switch action. We explore this in the next experiment by testing subtle variations of a pinch for mode-switching actions.

5 EXPERIMENT 2

This experiment has two goals. First, test more subtle dominant hand mode-switching techniques to see if actions more similar to a pinch trigger might perform better. Second, test the effect of using a device controller as the manipulation trigger. The apparatus, quantitative measures and experimental protocol are the same as in Experiment 1.

Participants

We recruited 12 participants (mean age 29.6 SD = 3.37, 5 women, 1 left-handed). 6 participants had experience using VR device. Remuneration was $10.

Techniques

We tested two sets of techniques. The first set consists of two subtle variations of a dominant hand pinch as a mode-switching action. We use a *pinch baseline* in this set, meaning a pinch is used to engage line drawing as in Experiment 1. The techniques are:

- *Oriented Pinch* (D-ORIENT) — The mode is active when the wrist is rotated clockwise (from the user's perspective) more than 45°. Manipulation is engaged and disengaged using the thumb-to-index pinch.
- *Middle Finger Pinch* (D-MIDDLE) — The mode is active when the middle finger and thumb are pinching. This also acts as simultaneous engagement of the manipulation trigger. Two 7 mm diameter FSRs were taped to the tips of the index and middle fingers for precise detection.

For the second set, we re-use two non-dominant techniques from Experiment 1, but this time with *controller baseline*, where selection is triggered by the button of a device controller held in the dominant hand. This enables us to compare drawing a line with and without holding a physical controller. An asterisk post-fix denotes these are versions of the same techniques, but with a controller for selection.

- *Non-Dominant Palm* (ND-PALM∗) — The mode is active when all fingers of the non-dominant hand are extended, roughly pointing up, with the palm facing right.
- *Non-Dominant Controller* (ND-DEVICE∗) — The mode is active when the button of the non-dominant hand controller is pressed. Note that both hands hold controllers.

Design and Procedure

The design and procedure are similar to the first experiment. The mode switching TECHNIQUE is a within-subjects factor with levels corresponding to the four mode-switching techniques. The TECHNIQUE order was counter-balanced using a 4 × 4 Latin Square. In sum: 4 TECHNIQUES × 6 BLOCKS (2 *baseline*, *4 compound*) × 3 directions × 5 line drawings = 360 line drawings per participant.

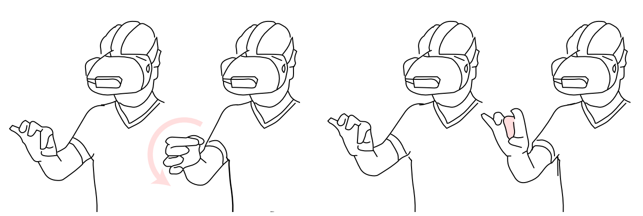

Figure 6: Mode switching techniques evaluated in experiment 2. (a) orientated pinch (D-ORIENT); (b) middle finger pinch (D-MIDDLE)

Results

Data Pre-Processing. The same methods were used as in Experiment 1. Less than 1% of the line drawing trials were removed as outliers, and mode-switch times were log transformed to correct non-normality. This resulted in 29 negative values so 275 ms was added to all times to obtain positive values. Again, log transformed data is used only for statistical analyses. Reported times are measured values.

Learning Effects. To determine if performance changed during the four compound blocks, we tested for effects of TECHNIQUE ×BLOCK on *mode-switch time* and *combined error rate*. We found no statistically significant interaction indicating no learning effect across blocks, so all blocks are used in the subsequent analysis.

Mode Switch Time. Middle finger pinch, a more subtle dominant technique, was among the fastest techniques. There was a significant main effect of TECHNIQUE ($F_{3,33}$ = 3.96, $p < .05$, $\eta^2_G = .26$). Post hoc tests found that ND-DEVICE∗ was faster than ND-PALM∗ and D-ORIENT ($p < .05$). Furthermore, D-MIDDLE is faster than D-ORIENT ($p < .001$). However, we did not find significant differences between ND-DEVICE∗ and D-MIDDLE. Ranking TECHNIQUES from fastest to slowest mode-switch time: ND-DEVICE∗ (226 ms) or D-MIDDLE (233 ms), ND-PALM∗ (467 ms), and D-ORIENT (669 ms).

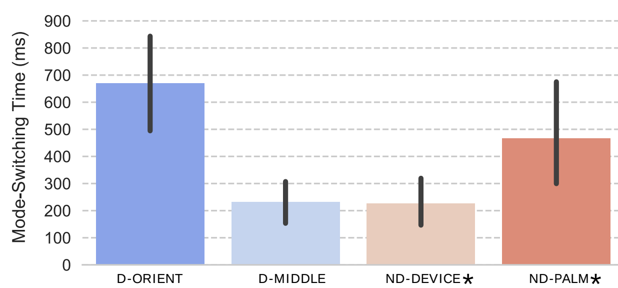

Figure 7: Mode-switch times by TECHNIQUE.

Error Rates. Overall error rate for the pinch and controller baseline tasks are below 5% (3.9% to 4.5%), and between 5% to 9.5% for mode-switching techniques (Figure 8). All techniques have comparable error rates, with the exception of middle finger pinch, which had a higher end error rate than non-dominant palm (Figure 9). There is no significant effect of TECHNIQUE on any type of error, except for *end error rate*. We found a main effect of TECHNIQUE on *end error rate* ($F_{3,33}$ = 3.52, $p < .05$, $\eta^2_G = .26$). Post hoc tests show that D-MIDDLE (5.0%) is higher than ND-PALM∗ (0.6%) ($p < .05$).

Table 2: Subjective ratings for both Experiments (mean ± SEM).

	(a) Experiment 1								(b) Experiment 2			
	D-POINT	D-FIST	ND-FIST	D-PALM	ND-PALM	ND-DEVICE	ND-HEAD	ND-FOV	D-ORIENT	D-MIDDLE	ND-DEVICE∗	ND-PALM∗
Accuracy	3.1 ±.3	3.7 ±.3	3.9 ±.2	3.2 ±.2	4.4 ±.2	**4.9 ±.1**	4.4 ±.2	3.9 ±.3	4.2 ±.2	4.3 ±.3	**4.5 ±.2**	3.8 ±.3
Learning	3.1 ±.4	3.7 ±.3	4.3 ±.3	3.2 ±.3	4.7 ±.1	**5.0 ±.0**	4.4 ±.2	4.1 ±.3	4.4 ±.3	**4.3 ±.4**	**4.3 ±.3**	3.8 ±.2
Ease of Use	3.2 ±.3	3.5 ±.3	3.9 ±.3	2.9 ±.3	4.7 ±.2	**4.8 ±.2**	4.1 ±.3	4.0 ±.3	4.1 ±.2	**4.3 ±.4**	4.2 ±.3	3.7 ±.2
Eye Fatigue	**4.6 ±.3**	4.8 ±.2	4.9 ±.1	4.9 ±.1	4.9 ±.1	4.9 ±.1	4.9 ±.1	**4.6 ±.3**	**4.8 ±.1**	4.9 ±.1	4.9 ±.1	4.9 ±.1
Hand Fatigue	**3.9 ±.4**	4.2 ±.3	4.2 ±.3	**3.9 ±.4**	4.4 ±.2	4.9 ±.1	4.4 ±.3	**3.9 ±.4**	4.4 ±.3	4.8 ±.1	4.6 ±.3	**4.3 ±.3**
Speed	3.4 ±.3	4.1 ±.2	4.2 ±.2	3.6 ±.3	4.5 ±.2	**4.8 ±.2**	4.3 ±.2	3.8 ±.3	4.2 ±.3	**4.6 ±.3**	4.4 ±.2	3.8 ±.3
Combined	3.6 ±.1	4.0 ±.1	4.2 ±.1	3.6 ±.1	4.6 ±.1	**4.9 ±.0**	4.4 ±.1	4.1 ±.1	4.4 ±.1	**4.6 ±.1**	4.5 ±.1	4.1 ±.1

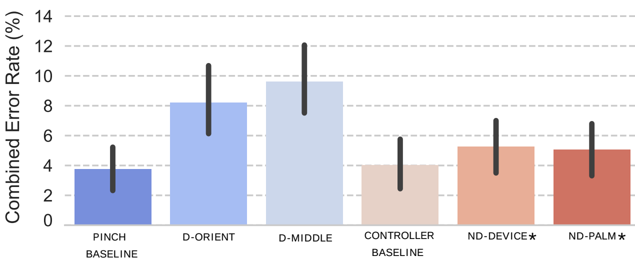

Figure 8: *Overall error rate* by TECHNIQUE. Baseline techniques have *start* and *end error rates*.

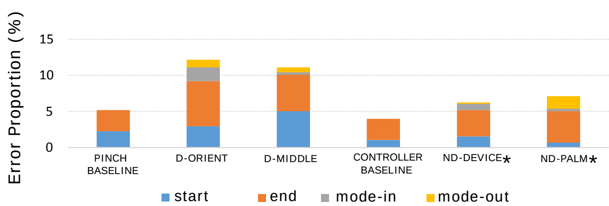

Figure 9: Proportion of specific error rates by TECHNIQUE.

Subjective Ratings. Similar procedures were followed to collect and process subjecting ratings. However, no statistically significant differences were found for any aspect, suggesting participants may have perceived all four techniques equally. Table 2 (b) summarizes those ratings. For overall preference, 42% of the participants liked D-MIDDLE and ND-DEVICE∗ and 42% disliked ND-PALM∗.

6 DISCUSSION

We discuss findings considering both experiments, and their potential impact on the design of VR interfaces.

Use non-dominant actions for accuracy-focused tasks and dominant actions for longer tasks. When considering accuracy only, non-dominant techniques such as raising a palm or touching the head are superior. If dominant hand mode switching is required, a subtle variation on the selection trigger, such as the middle finger pinch, is recommended. Slower and more error prone dominant postures, such as fist or palm, may be suitable for infrequent mode switching and tasks with less critical accuracy demands, such as panning. Also consider that while the non-dominant techniques are precise and fast, fatigue may be an issue for longer periods of use [32]. A mix of dominant and non-dominant mode-switching might prove to be most effective.

Avoid the pointing posture for frequent mode-switching. Dominant pointing was the slowest technique, and led to the most errors. This is surprising, since it is a commonly used mid-air bare hand action. Using it for raycast pointing is likely suitable [99], but other pointing techniques should be considered if frequent mode switching is expected.

Subtle dominant hand techniques are a promising alternative to non-dominant controller technique. Experiment 1 showed that the device controller technique was least error prone and fastest, but Experiment 2 shows that dominant-hand techniques can reach the performance levels of non-dominant ones when made subtle, as demonstrated by using the middle finger pinch as a simultaneous mode-switch and manipulation trigger. While we found significant differences between using a device and all dominant techniques in Experiment 1, mode-switching time for middle finger pinch (233 ms) was not significantly different from the device controller (331 ms) in Experiment 2. Considering the device controller was used for both mode-switching and manipulation trigger, this is quite remarkable. For overall errors, middle-finger pinch was comparable for all error rates, except the end error rate. However, the overall error rate for middle finger (9.5%) may be evidence of some potential increase compared to the very low rates when using a device controller. Nevertheless, subtle dominant hand techniques were perceived to be less fatiguing. So, in the future, more variations of subtle gestures need to be tested, for instance, single handed microgestures [15], which could be useful for long hours of use.

Switching between very different dominant postures may be confusing. For dominant hand techniques, we noted that higher error rates may have had more to do with motor-control confusion caused by rapidly switching between the baseline pinch and a qualitatively different posture such as fist, point, or hand. We believe the similarity to the basic pinch for the oriented pinch and middle finger pinch in Experiment 2 alleviated this confusion.

A pinch is a practical alternative to a controller button We saw little practical difference between using a pinch trigger compared to a controller button. When the non-dominant mode-switching action is held constant, any difference between a pinch or a controller button for line drawing trigger seems negligible. Comparing across experiments, t-tests found no significant differences in mode-switch time between ND-DEVICE and ND-DEVICE$\ast$ ($t(97.2) = 1.7941$, $p = .07$), or between ND-PALM and ND-PALM$\ast$ ($t(99.2) = -0.0656$, $p = .94$). Comparing only line drawing times during baseline tasks for these same two pairs further demonstrates a similarity. The times are actually significantly faster when using a pinch trigger compared to a device trigger ($t(1775.5) = -4.20$, $p < .001$), but the effect size is as little as 41ms between ND-DEVICE and ND-DEVICE$\ast$. Finally, we observed no differences between overall baseline task time in Experiment 2 when pinching for a trigger, or when using a device button ($t(102.3) = 1.39$, $p = 0.16$). This is an encouraging result showing that a dominant pinch action for selection may be as effective as a device controller button.

7 CONCLUSION

We presented an analysis of bare hand mid-air mode-switching techniques for VR. Techniques in Experiment 1 were selected using a principled review of related work, and techniques in Experiment 2 were selected to investigate specific questions raised by the first experiment. We found non-dominant techniques to be fast and accurate compared to most dominant techniques, but a dominant middle finger pinch shows comparable performance. Most dominant hand techniques, including popular actions like fist and point, also incurred high error rates likely due to confusion with the unmoded pinch manipulation trigger. Our findings can assist designers in making informed decisions when mapping techniques to mode-switching actions for VR applications. Our results may generalize to other 3D interfaces using absolute positioning and direct manipulation, like AR. We hope these results prove as useful to the VR community as previous mode-switching studies have been for pen and touch.

8 ACKNOWLEDGEMENTS

This work was made possible by NSERC Discovery Grant #2018-05187, the Canada Foundation for Innovation Infrastructure Fund “Facility for Fully Interactive Physio-digital Spaces” (#33151), and Ontario Early Researcher Award #ER16-12-184.

REFERENCES

[1] Vamsi Kiran Adhikarla, Pawel Wozniak, Attila Barsi, Dave Singhal, Péter Tamás Kovács, and Tibor Balogh. 2014. Freehand interaction with large-scale 3D map data. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2014.* IEEE, 1–4.

[2] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindbauer, Alexandra Ion, Shengdong Zhao, and JTKV Koh. 2012. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *Microsoft Research TechReport MSR-TR-2012-111* 2 (2012).

[3] Ferran Argelaguet and Carlos Andújar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37 (2013), 121–136.

[4] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *Virtual Reality (VR), 2016 IEEE.* IEEE, 3–10.

[5] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George W Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR.. In *CHI.* 5643–5654.

[6] Teo Babic, Harald Reiterer, and Michael Haller. 2017. GestureDrawer: one-handed interaction technique for spatial user-defined imaginary interfaces. In *Proceedings of the 5th Symposium on Spatial User Interaction.* ACM, 128–137.

[7] Gilles Bailly, Robert Walter, Jörg Müller, Tongyan Ning, and Eric Lecolinet. 2011. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In *IFIP Conference on Human-Computer Interaction.* Springer, 248–262.

[8] BasementVR 2017. BasementVR. (2017). http://store.steampowered.com/app/561390/BasementVR/.

[9] Hrvoje Benko and Andrew D Wilson. 2010. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces.* ACM, 19–28.

[10] Blocks 2017. Google Blocks. (2017). https://vr.google.com/blocks/.

[11] Richard A. Bolt. 1980. *"Put-that-there": Voice and Gesture at the Graphics Interface.* ACM, New York, NY, USA, 262–270.

[12] Benoît Bossavit, Asier Marzo, Oscar Ardaiz, and Alfredo Pina. 2013. Hierarchical menu selection with a body-centered remote interface. *Interacting with Computers* 26, 5 (2013), 389–402.

[13] Doug A Bowman and Larry F Hodges. 1999. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing* 10, 1 (1999), 37–53.

[14] William Buxton. 1990. A three-state model of graphical input. In *Human-computer interaction-INTERACT*, Vol. 90. 449–456.

[15] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User elicitation on single-hand microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 3403–3414.

[16] Yun Suk Chang, Benjamin Nuernberger, Bo Luan, and Tobias Höllerer. 2017. Evaluating gesture-based augmented reality annotation. In *3D User Interfaces (3DUI), 2017 IEEE Symposium on.* IEEE, 182–185.

[17] Paolo Cremonesi, Antonella Di Rienzo, and Franca Garzotto. 2015. Personalized interactive public screens. In *Workshop on Interacting with Smart Objects 2015.* 10.

[18] R. F. Dillon, Jeff D. Edey, and Jo W. Tombaugh. 1990. Measuring the true cost of command selection: techniques and results. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90* April (1990), 19–26. https://doi.org/10.1145/97243.97247

[19] Haiwei Dong, Ali Danesh, Nadia Figueroa, and Abdulmotaleb El Saddik. 2015. An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. *IEEE Access* 3 (2015), 543–555.

[20] Dustin Freeman, Ramadevi Vennelakanti, and Sriganesh Madhvanath. 2012. Freehand pose-based gestural interaction: Studies and implications for interface design. In *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on.* IEEE, 1–6.

[21] Franca Garzotto and Matteo Valoriani. 2013. Touchless gestural interaction with small displays: a case study. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI.* ACM, 26.

[22] Ivano Gatto and Fabio Pittarello. 2012. Prototyping a gestural interface for selecting and buying goods in a public environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces.* ACM, 784–785.

[23] Gravitysketch 2017. Gravitysketch. (2017). https://www.gravitysketch.com/vr/.

[24] Celeste Groenewald, Craig Anslow, Junayed Islam, Chris Rooney, Peter Passmore, and William Wong. 2016. Understanding 3D mid-air hand gestures with interactive surfaces and displays: a systematic literature review. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion!* BCS Learning & Development Ltd., 43.

[25] Tovi Grossman and Ravin Balakrishnan. 2006. The design and evaluation of selection techniques for 3D volumetric displays. In *Proceedings of the 19th annual ACM symposium on User interface software and technology.* ACM, 3–12.

[26] Faizan Haque, Mathieu Nancel, and Daniel Vogel. 2015. Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 3653–3656.

[27] Eiji Hayashi, Manuel Maas, and Jason I Hong. 2014. Wave to me: user identification using body lengths and natural gestures. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems.* ACM, 3453–3462.

[28] Luke Hespanhol, Martin Tomitsch, Kazjon Grace, Anthony Collins, and Judy Kay. 2012. Investigating Intuitiveness and Effectiveness of Gestures for Free Spatial Interaction with Large Displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12).* ACM, New York, NY, USA, 6:1–6:6.

[29] Hololense 2017. Hololense Gestures. (2017). https://developer.microsoft.com/en-us/windows/mixed-reality/gestures.

[30] Mohammadreza Hosseini, Pascal Vallotton, Tomasz Bednarz, and Arcot Sowmya. 2013. A study of touchless versus touch-based interactions with bacterial biofilm images. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry.* ACM, 285–290.

[31] Bret Jackson and Daniel F Keefe. 2016. Lift-off: Using reference imagery and freehand sketching to create 3d models in vr. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1442–1451.

[32] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).* ACM, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523

[33] Ricardo Jota, Miguel A Nacenta, Joaquim A Jorge, Sheelagh Carpendale, and Saul Greenberg. 2010. A comparison of ray pointing techniques for very large displays. In *Proceedings of Graphics Interface 2010.* Canadian Information Processing Society, 269–276.

[34] Alvin Jude, G Michael Poor, and Darren Guinness. 2016. Grasp, Grab or Pinch? Identifying User Preference for In-Air Gestural Manipulation. In *Proceedings of the 2016 Symposium on Spatial User Interaction.* ACM, 219–219.

[35] Jaemoon Jung, Hanjun Park, Dongwook Hwang, Minseok Son, Donghyun Beck, Juhee Park, and Woojin Park. 2014. A review on interaction techniques in virtual environments. In *Proc. 2014 International Conference on Industrial Engineering and Operations Management.* 1582–1590.

[36] Thomas Jung and Patrick Bauer. 2017. Constraint-based modeling technique for mid-air interaction. In *Proceedings of the 5th Symposium on Spatial User Interaction.* ACM, 157–157.

[37] Maria Karam et al. 2005. A study on the use of semaphoric gestures to support secondary task interactions. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems.* ACM, 1961–1964.

[38] Keiko Katsuragawa, Krzysztof Pietroszek, James R Wallace, and Edward Lank. 2016. Watchpoint: Freehand pointing with a smartwatch in a ubiquitous display environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces.* ACM, 128–135.

[39] Daniel F Keefe, Daniel Acevedo Feliz, Tomer Moscovich, David H Laidlaw, and Joseph J LaViola Jr. 2001. CavePainting: a fully immersive 3D artistic medium and interactive experience. In *Proceedings of the 2001 symposium on Interactive 3D graphics.* ACM, 85–93.

[40] David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology.* ACM, 167–176.

[41] Yeji Kim, Sohyun Sim, Seoungjae Cho, Woon-woo Lee, Young-Sik Jeong, Kyungeun Cho, and Kyhyun Um. 2014. Intuitive NUI for controlling virtual objects based on hand movements. In *Future Information Technology.* Springer, 457–461.

[42] Kinect 2017. Kinect gestures. (2017). https://support.xbox.com/en-GB/xbox-360/accessories/body-controller.

[43] Arun Kulshreshth and Joseph J LaViola Jr. 2014. Exploring the usefulness of finger-based 3D gesture menu selection. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems.* ACM, 1093–1102.

[44] Edward Lank, Jaime Ruiz, and William Cowan. 2006. Concurrent Bimanual Stylus Interaction: A Study of Non-preferred Hand Mode Manipulation. In *Proceedings of Graphics Interface 2006 (GI '06).* Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 17–24. http://dl.acm.org/citation.cfm?id=1143079.1143083

[45] Leap 2017. Leap gestures. (2017). https://developer.leapmotion.com/documentation/csharp/devguide/Leap_Gestures.html.

[46] Joon Hyub Lee, Sang-Gyun An, Yongkwan Kim, and Seok-Hyung Bae. 2018. Projective Windows: Bringing Windows in Space to the Fingertip. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM, 218.

[47] Sang-Su Lee, Jeonghun Chae, Hyunjeong Kim, Youn-kyung Lim, and Kun-pyo Lee. 2013. Towards more natural digital content manipulation via user freehand gestural interaction in a living room. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing.* ACM, 617–626.

[48] Yang Li, Ken Hinckley, Zhiwei Guan, and James A. Landay. 2005. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05.* ACM Press, New York, New York, USA, 461. https://doi.org/10.1145/1054972.1055036

[49] Lorraine Lin and Sophie Jörg. 2016. Need a hand?: how appearance affects the virtual hand illusion. In *Proceedings of the ACM Symposium on Applied Perception.* ACM, 69–76.

[50] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle arms-down mid-air interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology.* ACM, 63–71.

[51] Andreas Löcken, Tobias Hesselmann, Martin Pielot, Niels Henze, and Susanne Boll. 2012. User-centred process for the definition of free-hand gestures applied to controlling music playback. *Multimedia systems* 18, 1 (2012), 15–31.

[52] Julian Looser, Mark Billinghurst, Raphaël Grasset, and Andy Cockburn. 2007. An evaluation of virtual lenses for object selection in augmented reality. In *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*. ACM, 203–210.

[53] Paul Lubos, Gerd Bruder, Oscar Ariza, and Frank Steinicke. 2016. Touching the Sphere: Leveraging Joint-Centered Kinespheres for Spatial User Interaction. In *Proceedings of the 2016 Symposium on Spatial User Interaction*. ACM, 13–22.

[54] Paul Lubos, Gerd Bruder, and Frank Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In *3D User Interfaces (3DUI), 2014 IEEE Symposium on*. IEEE, 11–18.

[55] Masterpiece 2017. MasterpieceVR. (2017). http://www.store.steampowered.com/app/504650/MasterpieceVR/.

[56] Jess McIntosh, Asier Marzo, and Mike Fraser. 2017. SensIR: Detecting Hand Gestures with a Wearable Bracelet using Infrared Transmission and Reflection. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, 593–597.

[57] Myo 2017. Myo gestures. (2017). https://support.getmyo.com/hc/en-us/articles/202647853-What-gestures-does-the-Myo-armband-recognize-.

[58] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 177–186.

[59] Tao Ni, Doug A Bowman, Chris North, and Ryan P McMahan. 2011. Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *International Journal of Human-Computer Studies* 69, 9 (2011), 551–562.

[60] Paintlabvr 2017. Paintlab. (2017). https://www.paintlabvr.com/.

[61] Anna Pereira, Juan P Wachs, Kunwoo Park, and David Rempel. 2015. A user-developed 3-D hand gesture set for human–computer interaction. *Human factors* 57, 4 (2015), 607–621.

[62] Thies Pfeiffer. 2008. Towards gaze interaction in immersive virtual reality: Evaluation of a monocular eye tracking set-up. In *Virtuelle und Erweiterte Realität-Fünfter Workshop der GI-Fachgruppe VR/AR*. 81–92.

[63] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*. ACM, 99–108.

[64] Thammathip Piumsomboon, Adrian J. Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 282–299.

[65] Ondrej Polacek, Martin Klima, Adam J Sporka, Pavel Zak, Michal Hradis, Pavel Zemcik, and Vaclav Prochazka. 2012. A comparative study on distant free-hand pointing. In *Proceedings of the 10th European conference on Interactive tv and video*. ACM, 139–142.

[66] Ivan Poupyrev and Tadao Ichikawa. 1999. Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing* 10, 1 (1999), 19–35.

[67] Ivan Poupyrev, T Ichikawa, S Weghorst, and M Billinghurst. 1998. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, Vol. 17. Wiley Online Library, 41–52.

[68] Jef Raskin. 2000. *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional, 55–59.

[69] Guido Maria Re and Monica Bordegoni. 2014. A Natural User Interface for Navigating in Organized 3D Virtual Contents. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 93–104.

[70] Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. 2014. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 4077–4086.

[71] Jaime Ruiz, Andrea Bunt, and Edward Lank. 2008. A Model of Non-preferred Hand Mode Switching. In *Proceedings of Graphics Interface 2008 (GI '08)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 49–56. http://dl.acm.org/citation.cfm?id=1375714.1375724

[72] Jaime Ruiz and Edward Lank. 2007. A study on the scalability of non-preferred hand mode manipulation. In *Proceedings of the 9th international conference on Multimodal interfaces*. ACM, 170–177.

[73] Jaime Ruiz and Daniel Vogel. 2015. Soft-constraints to reduce legacy and performance bias to elicit whole-body gestures with low arm fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3347–3350.

[74] Steven Schkolne, Michael Pruett, and Peter Schröder. 2001. Surface drawing: creating organic 3D shapes with the hand and tangible tools. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 261–268.

[75] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. These are not my hands!: Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1577–1582.

[76] Abigail J Sellen, Gordon P Kurtenbach, and William AS Buxton. 1992. The prevention of mode errors through sensory feedback. *Human-computer interaction* 7, 2 (1992), 141–164.

[77] Kyeongeun Seo and Hyeonjoong Cho. 2014. AirPincher: A Handheld Device for Recognizing Delicate Mid-air Hand Gestures. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST'14 Adjunct)*. ACM, New York, NY, USA, 83–84. https://doi.org/10.1145/2658779.2658787

[78] Junchen Shen, Yanlin Luo, Xingce Wang, Zhongke Wu, and Mingquan Zhou. 2014. Gpu-based realtime hand gesture interaction and rendering for volume datasets using leap motion. In *Cyberworlds (CW), 2014 International Conference on*. IEEE, 85–92.

[79] Sketchfab 2017. Sketchfab. (2017). https://sketchfab.com/virtual-reality.

[80] Peng Song, Wooi Boon Goh, William Hutama, Chi-Wing Fu, and Xiaopei Liu. 2012. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1297–1306.

[81] Sophie Stellmach, Markus Jüttner, Christian Nywelt, Jens Schneider, and Raimund Dachselt. 2012. Investigating Freehand Pan and Zoom.. In *Mensch & computer*. 303–312.

[82] Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 265–272.

[83] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2017. Experimental Analysis of Mode Switching Techniques in Touch-based User Interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3267–3280.

[84] Justin H Tan, Cherng Chao, Mazen Zawaideh, Anne C Roberts, and Thomas B Kinney. 2013. Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures. *Radiographics* 33, 2 (2013), E61–E70.

[85] Shun'ichi Tano, Shinya Yamamoto, Junko Ichino, Tomonori Hashiyama, and Mitsuru Iwata. 2013. Truly Useful 3D Drawing System for Professional Designer by "Life-sized and Operable" Feature and New Interaction. In *IFIP Conference on Human-Computer Interaction*. Springer, 37–55.

[86] Robert J Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*. IEEE, 87–94.

[87] Tiltbrush 2017. Tiltbrush. (2017). https://www.tiltbrush.com/.

[88] Huawei Tu, Xing-Dong Yang, Feng Wang, Feng Tian, and Xiangshi Ren. 2012. Mode switching techniques through pen and device profiles. In *Proceedings of the 10th asia pacific conference on Computer human interaction - APCHI '12*. ACM Press, New York, New York, USA, 169. https://doi.org/10.1145/2350046.2350081

[89] Apivan Tuntakurn, Saowapak S Thongvigitmanee, Vera Sa-Ing, Shoichi Hasegawa, and Stanislav S Makhanov. 2013. Natural interactive 3D medical image viewer based on finger and arm gestures. In *Biomedical Engineering International Conference (BMEiCON), 2013 6th*. IEEE, 1–5.

[90] Florian van de Camp, Alexander Schick, and Rainer Stiefelhagen. 2013. How to click in mid-air. In *International Conference on Distributed, Ambient, and Pervasive Interactions*. Springer, 78–86.

[91] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2009. Multimodal selection techniques for dense and occluded 3d virtual environments. *International Journal of Human-Computer Studies* 67, 3 (2009), 237–255.

[92] Radu-Daniel Vatavu. 2012. User-defined gestures for free-hand TV control. In *Proceedings of the 10th European conference on Interactive tv and video*. ACM, 45–48.

[93] Radu-Daniel Vatavu. 2013. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *Journal of Ambient Intelligence and Smart Environments* 5, 2 (2013), 187–211.

[94] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap gestures for TV: insights from an elicitation study. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 131–138.

[95] Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 21.

[96] Amilcar Meneses Viveros and Erika Hernández Rubio. 2013. Kinect©, as interaction device with a tiled display. In *International Conference on Human-Computer Interaction*. Springer, 301–311.

[97] Bahtijar Vogel, Oskar Pettersson, Arianit Kurti, and Andreas Simonsson Huck. 2012. Utilizing gesture based interaction for supporting collaborative explorations of visualizations in TEL. In *Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on*. IEEE, 177–181.

[98] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 137–146.

[99] Daniel Vogel and Ravin Balakrishnan. 2005. Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. ACM, New York, NY, USA, 33–42. https://doi.org/10.1145/1095034.1095041

[100] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 841–850.

[101] Robert Walter, Gilles Bailly, Nina Valkanova, and Jörg Müller. 2014. Cuenesics: using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 299–308.

[102] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 143–146.

[103] Huiyue Wu and Jianmin Wang. 2012. User-defined body gestures for TV-based applications. In *Digital Home (ICDH), 2012 Fourth International Conference on*. IEEE, 415–420.

[104] Ionuţ-Alexandru Zaiţi, Ştefan-Gheorghe Pentiuc, and Radu-Daniel Vatavu. 2015. On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. *Personal and Ubiquitous Computing* 19, 5-6 (2015), 821–838.

[105] Mario Martínez Zarzuela, Francisco J Díaz Pernas, Sergio Martín Calzón, David González Ortega, and Miriam Antón Rodríguez. 2013. Educational tourism through a virtual reality platform. *Procedia Computer Science* 25 (2013), 382–388.

[106] Jamie Zigelbaum, Alan Browning, Daniel Leithinger, Olivier Bau, and Hiroshi Ishii. 2010. G-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*. ACM, 261–264.