Comparing Pinch and Point Poses for Single-Stroke Drawing in Virtual Reality

Tushar Billakanti and Jay Henderson

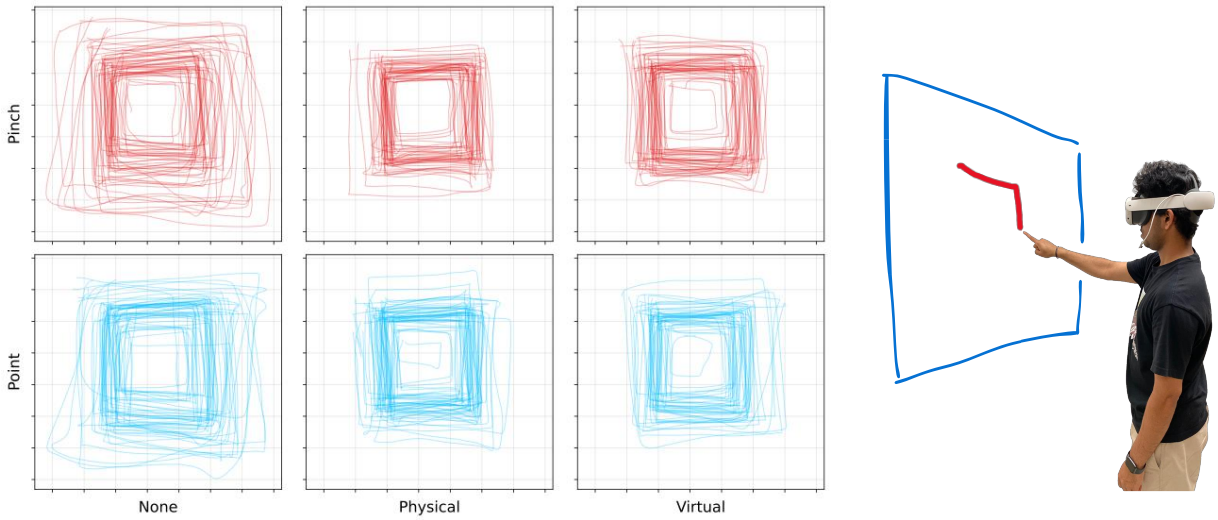


Figure 1: Square strokes produced in experiment 2 across all participants, poses, and surfaces, without rescaling.

Abstract—
This paper presents an investigation of bare-hand input for drawing single strokes in extended reality. Through two user studies, we compared two hand poses: pinch and point, across three surface types: a physical surface, a virtual surface, and no surface (free space). We evaluated the participant-produced strokes based on geometric and algebraic properties, performance, and subjective experience. Despite the common use of point poses for direct interaction with 2D content, our results showed that the pinch pose outperformed the point pose across most metrics. We discuss the strengths and limitations of each hand pose for bare-hand stroke input and provide insights into the role of surfaces in drawing tasks within extended reality.

Index Terms—Drawing, Barehand Input, Virtual Reality.

1 Introduction

Advancements in optical hand tracking have made commercial virtual reality (VR) and extended reality (XR) devices capable of untethered bare-hand input. This shift from controller-based input to bare-hand input offers advantages in flexibility and portability, suited for ubiquitous use, and can facilitate interactions similar to those on capacitive touch screens. Despite the advantages, a challenge remains of how to provide precise input when interacting without physically touching a device. One area where precision is important is in drawing tasks, which rely on hand position and fine motor control to create accurate strokes.

When no controller is used, bare-hand drawing in VR is typically accomplished in one of two ways. First, the user could perform a hand movement or gesture to indicate the start and end of a stroke while remaining in mid-air. Second, the user could begin and end their stroke by intersecting their hand with a virtual object or surface within the VR environment. While the latter provides an explicit delimiter, that is, either intersecting with the object or not, the former depends on a gesture.

Two gestures have emerged as particularly effective for producing precise input in VR environments. The first is the *pinch* gesture [23, 33, 41, 49], where the user touches their index finger and thumb together on the same hand. This gesture is the default action for indicating intent on many commercial XR devices, such as the Apple Vision Pro [1] or Meta Quest [37]. Suggested by prior work, the second gesture involves extending the index finger to create a *point* pose, with the thumb either extended or curled toward the rest of the hand to indicate intent [5,21,30,45]. In the case of drawing, each of these gestures would correspond to a specific hand pose, *pinch* or *point*, that is maintained during interaction. While the pinch gesture is effective for indicating intent, it is not how we typically interact with existing direct input devices, like touchscreens, using our bare hands. On touchscreens, we usually use a single extended finger to perform actions like drawing, tapping buttons, or manipulating content. Therefore, we question whether the point pose, where the index finger is extended, is better suited for drawing tasks in VR.

In addition to hand poses, the presence or absence of a surface has been shown to impact the accuracy of input in VR [28, 47]. In our work, we consider three surface conditions: physical surfaces, virtual surfaces, and no surface. Physical surfaces have the advantage of both visual and passive haptic feedback, whereas virtual surfaces only have the former. Drawing in mid-air without any surface removes both forms of feedback, which has been shown to degrade drawing precision [3]. Therefore, we consider surface as a contrasting condition for comparing *pinch* and *point* poses. Considering a single extended finger is the default interaction method for direct input with physical surfaces (i.e. touchscreens), we question whether *point* may be more

• *Tushar Billakanti and Jay Henderson are with Memorial University of Newfoundland. E-mail: tbillakanti | jayhend @mun.ca*

practical than *pinch* when a surface is present (either physical or virtual) or absent.

To investigate these questions, we follow an experimental protocol inspired by Tu et al. [43], who conducted an empirical user study evaluating finger and pen strokes using a set of algebraic and geometric properties, as well as subjective feedback. We focus our study on single strokes, or unistrokes, to isolate the properties of individual strokes rather than those of an entire multi-stroke drawing. Since pinch and point gestures determine when drawing begins and ends, an important aspect of stroke accuracy is how reliably they indicate the user's intent to start and stop drawing. When the gesture is detected too early or released too late, unintended or missing stroke segments can appear at the beginning or end of the stroke.

Our formal research questions are as follows:

RQ1: How do pinch and point hand poses compare for drawing single strokes in VR?

RQ2: How does the presence of a virtual, physical, or no surface compare for single-stroke drawing?

RQ3: How do surfaces and poses interact with one another for single-stroke drawing?

2 Related Work

In the context of drawing, our work considers single-stroke drawings, also known as unistrokes. A unistroke is defined as a single, continuous stroke, with a starting point, provided by some input mechanism (e.g. finger, hand, pen or mouse). Unistrokes have the unique property of being classified as both a gesture and a drawing. Thus, we position our work at the intersection of gestures, bare-hand input, and drawing in VR.

2.1 Unistroke Gestures

Much of the unistroke gesture work over the past two decades has focused on pen- or finger-based input on digitized surfaces [9, 10, 43, 51]. To compare pen- and finger-based input, Tu et al. conducted an empirical study varying where participants were asked to copy a set of strokes with varied complexity and size [43]. In a user study (N=15), they asked participants to draw 12 gestures, reproduced in Figure 4, comparing them on 9 empirical metrics and subjective feedback. The metrics included algebraic property features, local geometric shape measures, and global geometric shape measures. They found finger-drawn strokes to be larger and faster than pen-drawn, but that finger-drawn resulted in more deviation in intersection points and in the shape of corners. However, the overall shape accuracy (which they measured by proportional shape distance) was similar across both. Participants perceived their pen-drawn strokes to be more accurate.

2.2 Bare-hand Input

One fundamental difference between producing strokes on digitized surfaces versus in mid-air in VR, is the lack of an explicit delimiter to indicate intent for performing an action [20]. Many methods of indicating intent in bare-hand mid-air interactions have been proposed. Early works have proposed the use of a *pinch* gesture, where the user touches their index finger and thumb together, due to its ease of recognition and natural kinesthetic feedback [4, 49]. Balakrishnan and Mackenzie found the pinch grip to result in increased throughput when compared to an unsupported index finger for selection. They attribute this in part due to the similarities with pinch and holding or manipulating a pen [4]. This *pinch pose* has since been leveraged for strokes including word-gestures [33], directional gestures for in-car infotainment systems [18] and drawing or sketching in virtual reality [11].

In a point-and-click task on a large distant display, Vogel and Balakrishnan [45] propose *AirTap*, which is a "down and up" gesture of the index finger and *ThumbTrigger* [17, 30] which is an "in and out" gesture of the thumb, for both of which the hand is in an *ambiguous pose* or somewhere in-between open and closed (fist). Recent works expand upon the *AirTap* while using a *point* pose, or extending the index finger, before doing the "in and out" thumb gesture [5, 21]; however, Banerjee et al. [5] restrict to thumb tapping the side of the middle finger rather than the index finger. Surale et al. [41] conducted a series of studies investigating bimanual and unimanual bare-hand mode-switching techniques in VR. While acknowledging point poses as suitable for raycasting [45], their results indicate point poses should be avoided for frequent mode-switching and that pinch poses were well suited to simulate a button press. In a usability study, Jude et al. [23] found grasp poses are preferred for manipulation over pinching or making a fist.

2.3 Sketching and Drawing in VR

Previous work has studied how to facilitate accurate stroke drawing and sketching in VR. However, few of these works have considered drawing with the user's bare hand [11, 22, 33, 40, 44, 53], with most opting for physical devices like VR controllers [3, 6, 15, 27, 31, 32, 52] or specialized stylii [2, 3, 7, 13, 14, 26, 39, 46, 48]. Across these works, there has been a variety of different target strokes or drawings that participants were asked to produce, including straight lines [3, 7, 22, 25, 40, 53], curved lines [25], primitive 2D shapes (e.g. circles, squares, triangles) [3, 14, 22, 40, 46, 48], handwriting [7, 15, 27, 44], 3D cubic or curved geometries [6, 11, 32, 48], or low fidelity 3D objects (e.g. a building/house [2, 31], lantern [52], chair and flowerpot [31]).

For the purpose of our paper, we focus on literature that considered drawing with the user's bare hand. Chen et al. [11] compared helical strokes drawn in mid-air by a stylus, a VR controller, a hand pointing pose, and a hand pinching pose, using a foot switch to indicate intent to draw. They found the stylus and controller to be significantly more precise than the two bare-hand input techniques, but that point was more precise than pinch and pinch was fastest overall. Jiang et al. combine both pinch and point poses through *HandPainter* [22], a system that supports drawing on the non-dominant hand using a point pose and thumb trigger and drawing in mid-air using a pinch pose. When considering mid-air drawing, others have used the point pose [44, 53] or pinch pose [8, 33]. Cai et al. have specifically leveraged the pinch pose as a pen-holding posture to draw on virtual canvases and in 3D space, while using microgestures while holding the pen pose for mode switching [8]. Drawing on physical or virtual surfaces does not typically require a specific hand pose, but prior VR work has focused on the use of an extended index finger [40].

2.4 Haptic Feedback

Another issue with bare-hand interaction, particularly when interacting with virtual content, is the lack of haptic feedback to indicate contact to the user [20]. Prior works have shown that passive haptic feedback can support virtual object manipulation [28, 47], for menu selection [54], and target acquisition [12, 42]. Notably, Cheng et al. found that physical surfaces improved performance and comfort, reduced exertion and strain, and were preferred by users in target selection and docking tasks [12]. Liu et al. [29] similarly leveraged a physical surface for passive haptic feedback in a bare-hand VR whiteboard collaboration system, reporting improved writing accuracy and usability compared to a purely mid-air interaction technique.

Other research has explored the use of passive haptic feedback through surfaces to support drawing in virtual spaces, both on flat (or 2.5D) surfaces [2, 3, 13, 22, 40, 44] and volumetric surfaces [39, 46]. Despite lack of tactile feedback, researchers have also found 2D virtual surfaces to support spatial 3D drawings [3, 31]. Active haptic feedback including vibrotactile [14], pressure [14], and tension using a string [44] has been explored to support drawing strokes in virtual spaces.

3 Experiment 1

3.1 Participants

We recruited 18 participants (9 women and 9 men) with an average age of 22.72 years (*SD* of 3.71 years) from our university. 17 participants were right handed and 1 participant was left handed. Most participants self-reported being novice users of VR systems: 9 had never used VR before, 8 had used VR once or twice, and 1 occasionally used VR. No participants reported being regular users of VR or being very experienced with VR. All participants reported normal or corrected-to-normal vision. The study was reviewed by our institutional research ethics board and participants provided informed consent prior to beginning

the study. Participants were compensated with $20 CAD for their participation.

3.2 Apparatus

We used a Meta Quest 3 VR head-mounted display (HMD) with a resolution of 2064 x 2208 pixels per eye and a native 120 Hz refresh. The HMD has a 110° horizontal and 96° vertical effective field of view (FOV). The VIRTUAL environment (VE) was constructed in Unity 3D (v 2022.3.17f1). Development included the Meta XR All-in-one SDK v.66.0.0 [34], required for headset and hand tracking.

3.2.1 Hand Tracking

Hand tracking was accomplished through inside-out optical tracking on the Meta Quest 3, using the OVR Skeleton [36] for hand data segmentation. Our application facilitated 3D stroke drawing using two distinct hand poses: a pinch pose and a point pose, with participants free to use either hand depending on their reported handedness. The pinch pose was chosen because it is the standard gesture for indicating intent in bare-hand mixed reality interactions (e.g. it is the default on headsets like the Meta Quest [36] and Apple Vision Pro [1]) and is widely used in prior research [23, 33, 41, 45]. The point pose was inspired by prior works [5, 21, 30, 45], but also reflects the index finger as the prototypical end-effector for touchscreen (surface) interaction.

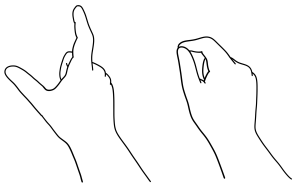


Figure 2: The PINCH gesture and pose held while drawing.

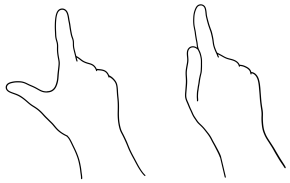


Figure 3: The POINT gesture and pose held while drawing.

3.2.2 Pinch Pose

The pinch pose was detected when the thumb and index finger touched, as in Figure 2. The pinch was detected through the OVRHand's built-in "pinch" gesture recognizer [36]. For NOSURFACE (free space) drawing, the drawing position was calculated as the mean position of the tips of the thumb and index finger. In the surface conditions (VIRTUAL or PHYSICAL), the system collected four points: the tip and intermediate joint of the index finger, along with the thumb tip and distal joint. These points formed two rays, which were averaged to create a single ray, similar to how a pen would be held in the hand. The drawing point was determined by the intersection of this ray with the surface and stabilized using a One Euro filter to reduce jitter.

3.2.3 Point Pose

This pose was recognized when the index finger was extended (index feature's curl was NOT CLOSED), while the other fingers (thumb, middle, ring, and pinky) were not extended (features' curl was NOT OPEN) [36]. This point pose was similar to previous work [5,21,30,45]. Participants were instructed to keep their index finger and thumb in an *L*-shape, then to fold their thumb inward (resting against the palm) to begin a stroke, shown in Figure 3. For NOSURFACE (free space), the drawing position was the index finger tip. In the surface conditions, the system collected two points, the index finger tip and the intermediate joint to create a ray. The drawing point was defined by where this ray intersected the canvas surface, and the position was stabilized with a One Euro filter.

3.2.4 Drawing Conditions

Our experiment varied both the drawing surface (PHYSICAL, VIRTUAL, or NOSURFACE) and hand pose (PINCH or POINT). In the NOSURFACE *condition*, drawing was initiated as soon as participants performed the designated hand pose (PINCH or POINT) and continued as long as they maintained the pose. Drawing stopped when the pose was not detected for 10 consecutive frames. For the *surface conditions* (both VIRTUAL and PHYSICAL), participants started drawing when their pose was detected and an invisible ellipsoid game object attached to the fingertip (or, in the pinch condition, to the midpoint between the thumb and index finger) collided with the surface. The ellipsoid extended to the length of the user's finger to ensure it maintained collision while intersecting with the VIRTUAL surface. Drawing stopped if either the pose was not detected for 10 frames, the ray no longer intersected with the surface for 10 frames, or both conditions occurred simultaneously. This method provided more robust control for surface-based drawing.

We intentionally did not use a motion capture (MoCap) system, such as OptiTrack, as wearing the IR markers interfered with hand movement, particularly when interacting with the PHYSICAL surface using a pinch pose. As well, we wanted to focus on using a widely accessible, commercially available VR system, like the Meta Quest 3, to ensure practicality and ease of use.

3.2.5 Surfaces

We used Meta's space setup system [35] coupled with world-locking feature to create scene anchors to set up the whiteboard as a PHYSICAL surface. The PHYSICAL surface's canvas used the wall art anchor synced to the real world position of a whiteboard in the lab. The VIRTUAL surface's canvas used another scene anchor to stay at a fixed position with respect to the real world. These canvases were white and semi-transparent (opacity of 0.6), so that participants could see their hand if it intersected the canvas. A black hand avatar was used for high contrast against the white surfaces.

3.2.6 3D Stroke Drawing

Since our experiment required drawing in 3D, rather than use Unity's 2D LineRenderer, we developed a custom procedural tube generator for rendering meshes in 3 dimensions. As the user moves their hand in VIRTUAL reality, points are added to the tube generator, which then dynamically creates meshes and updates the tube. The tube radius was 0.6cm and its shape was computed using Frenet-Serret frames to ensure smooth transitions along its path.

3.3 Procedure

The goal of our study was to evaluate the impact of surface type and hand pose on drawing performance in VR. Thus, the experiment consisted of six different conditions, combining two independent variables:
Surface Type: Participants were required to draw on either (a) a PHYSICAL surface, (b) a VIRTUAL surface, or (c) NOSURFACE.
Hand Pose: Participants utilized one of two hand poses during the task: (a) the pinch pose, where the thumb and index finger touched to simulate holding a pen, or (b) the point pose, where the index finger was extended with other fingers curled.

As our experiment was a within-subjects design, we counter balanced orderings of conditions across participants using a 6x6 Latin Square design. The study was conducted in VIRTUAL reality (VR), so after providing an explanation of the study, participants were fitted with a Meta Quest 3. Before each condition, the experimenter gave a brief demonstration of how the condition works. Each participant completed three blocks of each condition. After each block, participants were permitted a break for as long as required. Participants were standing for the duration of the experiment task, thus, in surface conditions, they could physically move closer or further away from the surface as desired. After each condition, participants exited VR and completed the NASA Task Load Index rating sheet [19]. Participants then returned to VR to continue the next condition. After the experiment was complete, participants filled out a brief open-ended questionnaire on their preferences and experience. In total, the study took approximately one hour to complete.

3.3.1 Experimental Task

The experimental task followed closely to Tu et al. [43]. We began with a target stroke appearing in front of the participant. The stroke was visible for five seconds, after which it disappeared. Participants were asked to complete the stroke drawings in the assigned condition

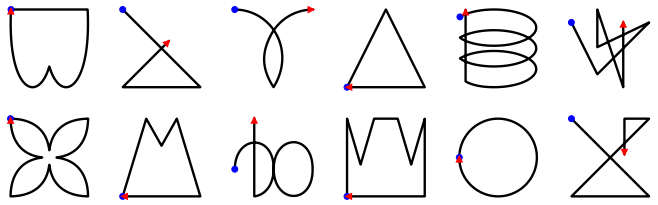


Figure 4: Target strokes in experiment 1, replicated from Tu et al. [43].

as accurately as possible. Once the stroke was completed, the system presented a brief pause for 2 seconds before the next stroke appeared. We opted against an always visible stroke and training task so that the drawing and panel locations would remain fixed in the virtual world and that the participant could remain stationary. Each block consisted of 12 target strokes, which were randomized in order of appearance.

The target stroke consisted of a blue circle as the starting point, with red arrows indicating the direction if necessary and a red arrow as the finishing point. The target strokes replicated that of Tu et al. [43] to capture varying levels of complexity with a diverse set of geometric properties. These strokes are presented in Figure 4.

3.4 Experimental Design

Our experiment used a 2×3 within-subjects design with the following factors and levels: **Hand pose** (PINCH or POINT) and **Surface** (PHYSICAL, VIRTUAL, and NOSURFACE).
In total, we collected:

18 participants $\times$
2 hand poses $\times$
3 surfaces $\times$
3 blocks $\times$
12 strokes =
3888 user drawn strokes

3.5 Results

3.5.1 Data Preparation

Data parsing and preparation was conducted using Python. To start, we removed any strokes with less than 3 points (i.e. 3D vectors representing the path of the drawing), as we deemed this insufficient for analysis. This consisted of 16 strokes, leaving 3872 valid trials. To compare 2D and 3D strokes, we standardized the 3D strokes by rotating and translating all paths to lie flat on the XY-plane. This was done by first fitting a plane to each stroke's 3D point cloud using Principal Component Analysis (PCA). We then rotated each point cloud so that its normal vector aligned with the z-axis. To ensure consistent alignment, the plane normal is oriented to always point upwards (positive z-direction). After rotation, the points were centered by translating them based on their centroid to ensure that all strokes were centered on the origin.

After aligning the strokes, we then manually annotated each stroke based on three metrics of correctness: *incomplete-distinguishable*, meaning the stroke represented the target stroke but the user either unintentionally began the tracking pose too late or ended it too early; *wrong-shape*, meaning the user drew the wrong shape, either not matching the stroke path or drawing the wrong stroke entirely; and *unrecognizable*, meaning there was no resemblance of an intentional shape drawn or the user scribbled out what they were drawing. If the stroke met any of these three categories, it was marked as an incorrect stroke, otherwise, it was marked as a correct stroke. One of the authors coded every stroke into these four categories, followed by a second author recoding the entire dataset independently as a reliability check. Out of 3872 strokes, 866 were marked as incorrect strokes (22.37%), leaving 3006 to be used in analysis requiring correct strokes (77.63%). Agreement between the two annotators was very high (Cohen's $\kappa = 0.982$, 26 disagreements out of 3872 strokes).

For the following analyses, we ran one- or two-way analyses of variance (ANOVAs) on aggregate (or log-transformed) data that met the normality assumption via Shapiro-Wilks. If this assumption was not met, we applied an aligned rank transform (ART) [24, 50] before performing an ANOVA.

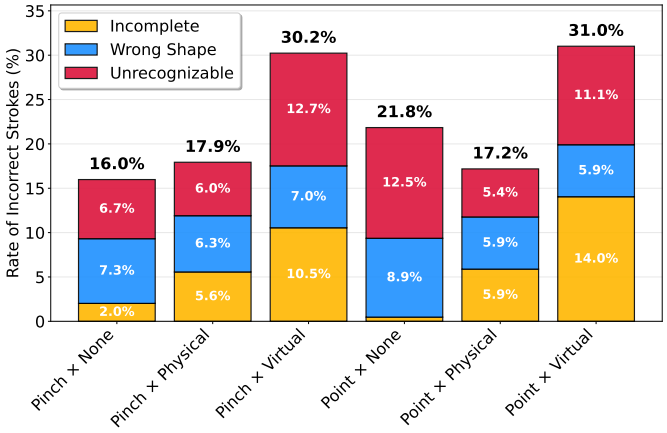


Figure 5: Rates of incorrect strokes and distributions across conditions (pose $\times$ surface) in experiment 1.

3.5.2 Incorrect Stroke Distribution

An ART-ANOVA revealed a significant main effect of surface type ($F_{2,85} = 12.30$, $p < 0.001$, $\eta^2_G = 0.193$) on incorrect stroke rate. Pose was not significant ($F_{1,85} = 1.35$, $p = 0.249$, $\eta^2_G = 0.013$) and the interaction between pose and surface was also non-significant ($F_{2,85} = 1.22$, $p = 0.299$, $\eta^2_G = 0.023$). Post-hoc pairwise comparisons using Tukey's method for the surface factor indicated that the VIRTUAL condition resulted in significantly higher incorrect stroke rates compared to both the NOSURFACE condition ($p = 0.001$) and the PHYSICAL condition ($p < 0.001$). No significant difference was found between the NOSURFACE and PHYSICAL surface conditions.

Overall incorrect stroke rates were lowest in the NOSURFACE conditions (15.98% for PINCH, 21.84% for POINT) and PHYSICAL conditions (17.93% for PINCH, 17.18% for POINT), while VIRTUAL conditions showed substantially higher incorrect stroke rates (30.23% for PINCH, 31.01% for POINT) (Figure 5). Incomplete strokes were most common in VIRTUAL conditions (10.54% for PINCH vs. 14.04% for POINT). Wrong shape strokes (user errors) remained relatively consistent across conditions (5.86–8.89%), while rates of unrecognizable strokes were highest in VIRTUAL conditions (12.71% for PINCH, 11.11% for POINT) and POINT with NOSURFACE (12.48%).

3.5.3 Start and End Anomaly Detection

When tagging incorrect strokes, we noticed that often despite drawing the requested stroke, many of the produced strokes had anomalies present on the start or end of its path. Anomalies at the start of the stroke were likely due to the user unintentionally beginning a stroke, as the system detected a PINCH or POINT pose. Anomalies at the end of a stroke were likely due to the user not stopping the pose when they had finished drawing, i.e. they got "stuck" in the drawing mode. The anomalies typically appeared as harsh corners near the beginning or end of a stroke. Examples of anomalies are shown in Figure 6.

To account for this, we took a two-step approach to detect and analyze these anomalies. After normalizing the produced stroke, we calculated an angle profile by obtaining the angle at each point, followed by finding the peaks of where largest changes had occurred in each stroke (minimum 30 degrees). Using the largest changes, we then used dynamic time warping (DTW) [16] to compare combinations of the peaks, to find the combination that resulted in the best match to the desired target. Since this was an approximation, after collecting the best combination, we manually confirmed each stroke and shifted the positions of start and end anomalies to a different angle peak if required.

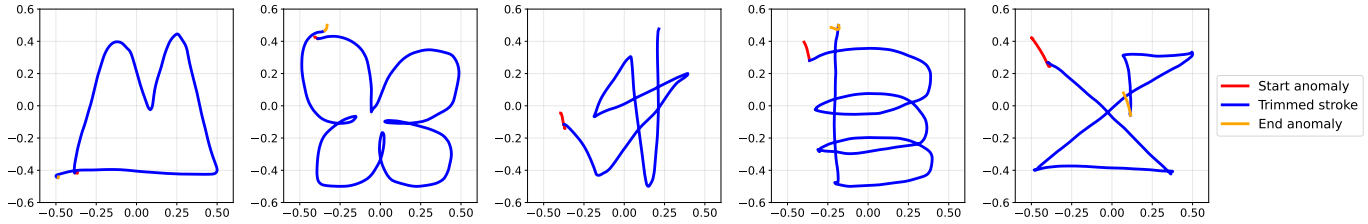


Figure 6: Anomaly examples from experiment 1.

We consider size of both start and end anomalies. Due to variations in stroke size, we analyze and report anomaly sizes as percentage of the entire stroke. Raw anomaly size and stroke size used in this calculation are in cumulative distance between points.

We ran a two-way ANOVA on log-transformed start anomaly percentages. Surface type and the pose $\times$ surface interaction violated sphericity assumptions, so Greenhouse-Geisser corrections were applied. The analysis revealed significant main effects for pose ($F_{1,17} = 11.12$, $p = 0.004$, $\eta_G^2 = 0.111$), surface type ($F_{2,34} = 5.08$, $p = 0.026$, $\eta_G^2 = 0.080$), and pose $\times$ surface interaction ($F_{2,34} = 9.22$, $p = 0.005$, $\eta_G^2 = 0.107$).

The main effect of pose indicated that POINT poses ($M = 2.05\%$, $SD = 2.51\%$) produced significantly larger start anomaly percentages than PINCH poses ($M = 0.88\%$, $SD = 0.63\%$). For surface type, the NOSURFACE condition ($M = 2.16\%$, $SD = 2.99\%$) showed higher start anomaly percentages compared to both PHYSICAL ($M = 1.14\%$, $SD = 0.88\%$) and VIRTUAL conditions ($M = 1.11\%$, $SD = 0.86\%$). Post-hoc pairwise comparisons with Tukey correction revealed no significant differences between surface conditions when averaged across pose types. However, examination of the interaction showed a significant difference between PINCH,NOSURFACE and POINT,NOSURFACE conditions ($p = 0.010$), indicating that POINT poses in the NOSURFACE condition produced significantly larger start anomaly percentages ($M = 3.55\%$, $SD = 3.73\%$) than PINCH poses in the same condition ($M = 0.76\%$, $SD = 0.66\%$). The remaining conditions showed relatively similar and lower mean start anomaly percentages between 0.93% and 1.32% (see Figure 7).

We also ran a two-way within-subjects ANOVA on log-transformed end anomaly percentages. The analysis revealed a significant main effect for pose ($F_{1,17} = 23.33$, $p < 0.001$, $\eta_G^2 = 0.228$), but not for surface type ($F_{2,34} = 3.14$, $p = 0.056$, $\eta_G^2 = 0.053$) or pose $\times$ surface interactions ($F_{2,34} = 2.45$, $p = 0.101$, $\eta_G^2 = 0.036$). The main effect of pose indicated that POINT poses ($M = 2.41\%$, $SD = 1.65\%$) produced significantly larger end anomaly percentages than PINCH poses ($M = 1.10\%$, $SD = 0.73\%$). Overall, the anomaly analyses revealed that POINT poses consistently produced significantly larger anomalies than PINCH poses at both stroke start and end positions, with the NOSURFACE condition being particularly problematic for start anomalies in POINT poses (see Figure 7).

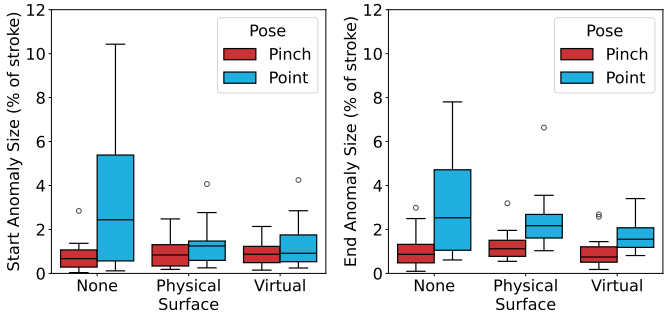


Figure 7: Box plots of start and end anomaly sizes across poses and surfaces in experiment 1.

3.5.4 Stroke Accuracy via Dynamic Time Warping

To measure stroke accuracy, DTW distances [16] between target templates and user produced strokes after removing start and end anomalies. Prior to calculating DTW distances, we resampled both stroke sequences to 200 points based on arc length to match their spatial sampling. We analyzed strokes in 2D (x-y coordinates only), as our PCA-based coordinate transformation ensured that stroke trajectories primarily lie in the XY-plane, with minimal z-component variation (representing the intended stroke).

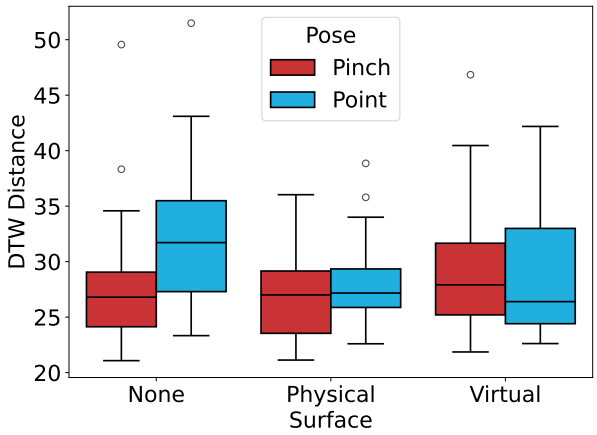


Figure 8: DTW distances from target strokes across poses and surfaces in experiment 1.

We ran an ART-ANOVA on DTW distances. The analysis revealed significant main effects for pose ($F_{1,85} = 11.08$, $p = 0.001$, $\eta_G^2 = 0.097$), surface type ($F_{2,85} = 4.76$, $p = 0.011$, $\eta_G^2 = 0.085$), and pose $\times$ surface interaction ($F_{2,85} = 9.21$, $p < 0.001$, $\eta_G^2 = 0.152$). The main effect of pose indicated that POINT poses ($M = 29.91$, $SD = 6.23$) produced significantly higher DTW distances than PINCH poses ($M = 28.29$, $SD = 5.98$). For surface type, the NOSURFACE condition ($M = 30.52$, $SD = 7.41$) showed higher DTW distances compared to both PHYSICAL ($M = 27.88$, $SD = 4.47$, $p = 0.010$) and VIRTUAL conditions ($M = 28.89$, $SD = 6.03$, $p = 0.083$). No significant difference was found between PHYSICAL and VIRTUAL surface conditions ($p = 0.689$).

However, examination of the interaction showed that the effect of pose was dependent on surface condition. Significant differences were found between POINT,NOSURFACE and PINCH,NOSURFACE conditions ($p < 0.001$), indicating that POINT poses in the NOSURFACE condition produced significantly higher DTW distances ($M = 32.86$, $SD = 7.38$) than PINCH poses in the same condition ($M = 28.17$, $SD = 6.87$) (Figure 8). Additionally, POINT poses showed significant improvement when performed on PHYSICAL ($M = 28.28$, $SD = 4.50$; $p = 0.010$) and VIRTUAL ($M = 28.57$, $SD = 5.69$; $p = 0.004$) surfaces compared to the NOSURFACE condition. The remaining surface conditions showed relatively similar DTW distances between poses, with PINCH poses maintaining consistent performance across all surface types.

3.5.5 Size

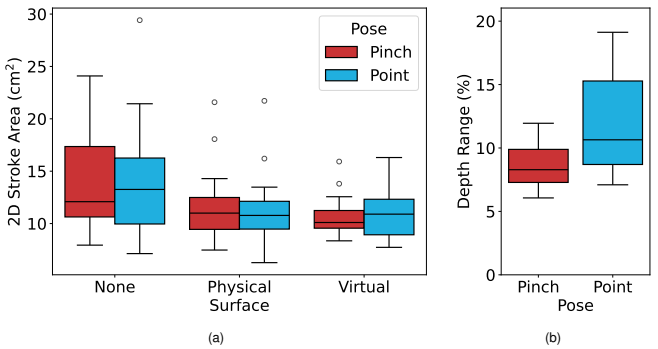


Figure 9: Dimensions of produced strokes in experiment 1 in terms of (a) Bounding box sizes across conditions (cm^2) and (b) Depth range across poses in NoSurface.

Next we considered the size of the stroke. We define size as the 2D bounding box of a stroke after removing anomalies. We ran an ART-ANOVA on stroke bounding box areas. The analysis revealed a significant main effect for surface type ($F_{2,85} = 10.88$, $p < 0.001$, $\eta^2_G = 0.174$), but not for pose ($F_{1,85} = 0.16$, $p = 0.694$, $\eta^2_G = 0.002$) or pose $\times$ surface interaction ($F_{2,85} = 0.51$, $p = 0.601$, $\eta^2_G = 0.010$) (see Figure 9a). The main effect of surface indicated that stroke size varied significantly across surface conditions. Post-hoc pairwise comparisons with Tukey correction revealed that the NoSurface condition produced significantly larger stroke areas ($M = 13.72$ cm^2, $SD = 4.92$ cm^2) compared to both Physical ($M = 11.44$ cm^2, $SD = 3.50$ cm^2; $p = 0.002$) and Virtual conditions ($M = 10.83$ cm^2, $SD = 2.17$ cm^2; $p < 0.001$). No significant difference was found between Physical and Virtual surface conditions ($p = 0.664$).

3.5.6 Depth Range

As only the NoSurface condition allowed for deviation on the z-axis, we only consider depth variation for strokes performed in the NoSurface condition. Depth range was calculated as the difference between maximum and minimum z-coordinates within user produced strokes (after removing start and end anomalies). So that our results are scale-independent, we calculated range as the percentage of the diagonal of the strokes 2D bounding box. Results are presented in Figure 9b. A one-way ANOVA revealed a significant effect of pose on depth range variation ($F_{1,17} = 17.39$, $p < 0.001$, $\eta^2_G = 0.241$). Point poses showed significantly greater depth range variation ($M = 11.84\%$, $SD = 3.69\%$) compared to Pinch poses ($M = 8.65\%$, $SD = 1.85\%$). Overall, participants deviated on average 10.25% of their stroke extent in the depth dimension when drawing with NoSurface ($SD = 3.30\%$).

3.5.7 Average Speed

We ran an ART-ANOVA on average stroke speed after anomaly removal. The analysis revealed no significant main effects for pose ($F_{1,85} = 0.05$, $p = 0.831$, $\eta^2_G < 0.001$) or pose $\times$ surface interaction ($F_{2,85} = 1.12$, $p = 0.332$, $\eta^2_G = 0.021$). Surface type was near significant ($F_{2,85} = 3.09$, $p = 0.051$, $\eta^2_G = 0.057$).

3.5.8 Subjective Data

Our subjective results were obtained using the NASA TLX rating sheet [19], shown in Figure 10. We ran an ART-ANOVA on each rating sheet subscale. For brevity, only statistically significant effects are reported.

Main effects on pose were revealed for both **Mental Demand** ($F_{1,85} = 5.01$, $p = 0.028$, $\eta^2_G = 0.046$) and **Physical Demand** ($F_{1,85} = 4.66$, $p = 0.034$, $\eta^2_G = 0.043$), with Pinch resulting in lower demand than Point for both measures. **Effort** showed a main effect of surface ($F_{2,85} = 9.01$, $p < 0.001$, $\eta^2_G = 0.149$), with Virtual surface

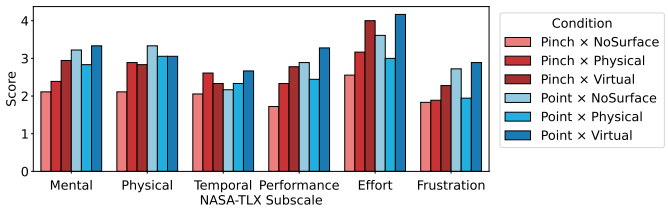


Figure 10: NASA TLX Rating Sheet Scores per sub scale across pose and surfaces in experiment 1.

requiring significantly more effort than both NoSurface ($p = 0.001$) and Physical surface ($p = 0.001$). **Performance** showed main effects of both pose ($F_{1,85} = 7.79$, $p = 0.006$, $\eta^2_G = 0.070$) and surface ($F_{2,85} = 3.60$, $p = 0.032$, $\eta^2_G = 0.065$), with Pinch resulting in better perceived performance than Point, and Virtual surface resulting in worse perceived performance than NoSurface ($p = 0.035$).

We also asked participants their preference for pose and surface for drawing in VR. For pose, 9 out of 18 participants preferred Pinch, 7 preferred Point, and 2 noted that it depended on the surface: Pinch for no surface and Point if a surface was present. Participants provided some insight for their preferences, for instance P3 noted that they preferred Point as they did not like the feeling of the pinch gesture on the physical board. Some participants had opposite opinions, for instance, P0, P4, and P11 felt pointing was easier whereas P1, P5, P7, and P15 felt pinching was easier. In terms of overall surface preference, results were also mixed: 3 participants preferred the Virtual surface, 3 preferred Physical surface, 7 preferred NoSurface, and 5 felt the Physical and NoSurface were a tie.

4 Experiment 2

4.1 Experimental Protocol Updates

While experiment 1 used identical strokes to those in Tu et al.'s study [43], these are not representative of shapes people commonly draw. To validate our findings with more realistic stimuli, we conducted a follow-up experiment using eight familiar single strokes. We suspected that user error incorrect strokes in the first experiment may have stemmed from participants misremembering complex stroke patterns or misinterpreting the required trajectories, so we selected simpler, more intuitive shapes.

Our stroke set consisted of eight strokes: figure-8, square, triangle, circle, zig-zag, pigtail, star, and wave, as illustrated in Figure 11. Five strokes (square, triangle, circle, pigtail, and star) were drawn from the unistroke gesture set proposed by Wobbrock et al. [51], while the remaining three were designed to complete the set, ensuring representation of the stroke properties examined in both Experiment 1 and prior work [43] (e.g., curved vs. straight lines, sharp corners, closed vs. open shapes, varying complexity).

4.1.1 Participants

We recruited 18 new participants for the second experiment. Demographic information is reported for 16 participants, as two declined to respond to our questionnaire. The mean age was 24.4 years (SD = 4.27), with one participant preferring not to disclose their age. 6 participants were women and the remaining 10 were men. Most participants self-reported being novice users of VR systems: 4 participants never used VR before, 9 had used VR once or twice, and 3 use VR occasionally (a few times a year). All participants reported normal or corrected-to-normal vision. The study was reviewed by our institution's research ethics board and participants provided informed consent prior to beginning the study. Participants were compensated with $20 CAD for their participation.

4.1.2 Experimental Design

As in experiment 1, our second experimental design used a 2 $\times$ 3 within-subjects design, with a different smaller set of strokes. In total, we collected:

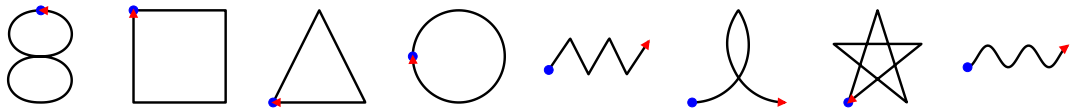


Figure 11: Target strokes in experiment 2.

18 participants ×
2 hand poses ×
3 surfaces ×
3 blocks ×
8 strokes =
2592 trials user drawn strokes

4.2 Results

Given space constraints, we report only objective measures, as the primary goal of this experiment was to understand whether we could replicate findings with a simpler stroke set.

4.2.1 Data Preparation

Eight strokes had less than 3 points and were excluded, leaving 2584 valid strokes for analysis. After parsing, aligning, and manually annotating strokes (first by one author, then validated by a second author), out of the remaining 2584 valid strokes, 435 were marked as incorrect strokes (16.83%), leaving 2149 (83.17%) correct strokes to analyze. Agreement between the two annotators was very high (Cohen's $\kappa = 0.980$, 15 disagreements out of 2584 strokes).

4.2.2 Incorrect Strokes

An ART-ANOVA revealed a significant main effect of surface type ($F_{2,85} = 10.14$, $p < 0.001$, $\eta^2_G = 0.165$) on incorrect stroke rate. Pose was not significant ($F_{1,85} = 1.80$, $p = 0.183$, $\eta^2_G = 0.017$) and the interaction between pose and surface was also non-significant ($F_{2,85} = 1.18$, $p = 0.313$, $\eta^2_G = 0.022$). Post-hoc pairwise comparisons using Tukey's method for the surface factor indicated that the VIRTUAL condition resulted in significantly higher incorrect stroke rates compared to both the NOSURFACE condition ($p < 0.001$) and the PHYSICAL condition ($p = 0.005$). No significant difference was found between the NOSURFACE and PHYSICAL surface conditions ($p = 0.522$).

Overall incorrect stroke rates were lowest in the NOSURFACE conditions ($M = 8.81\%$ for PINCH, $M = 15.46\%$ for POINT) and PHYSICAL conditions ($M = 15.51\%$ for PINCH, $M = 16.93\%$ for POINT), while VIRTUAL conditions showed higher incorrect stroke rates ($M = 22.92\%$ for PINCH, $M = 21.34\%$ for POINT) (Figure 12). The overall study incorrect stroke rate was $M = 16.8\%$. Surface means showed a clear pattern: NOSURFACE ($M = 12.12\%$), PHYSICAL ($M = 16.22\%$), and VIRTUAL ($M = 22.13\%$).

Incomplete strokes were most common in VIRTUAL conditions ($M = 10.42\%$ for PINCH, $M = 7.89\%$ for POINT), while being extremely rare in NOSURFACE conditions ($M = 0.46\%$ for PINCH, $M = 0.47\%$ for POINT). Wrong shape strokes remained relatively consistent across conditions ($M = 2.09$–4.87%), while unrecognizable stroke rates were consistently high across all conditions ($M = 6.26$–10.67%).

4.2.3 Start and End Anomaly Detection

We conducted a two-way ANOVA on log-transformed start anomaly percentages. The pose × surface interaction violated sphericity assumptions, so Greenhouse-Geisser corrections were applied. The analysis revealed significant main effects for pose ($F_{1,17} = 7.61$, $p = 0.013$, $\eta^2_G = 0.047$) and surface type ($F_{2,34} = 4.97$, $p = 0.013$, $\eta^2_G = 0.076$), but pose × surface interaction was not significant ($F_{2,34} = 3.49$, $p = 0.065$, $\eta^2_G = 0.058$).

The main effect of pose indicated that POINT poses ($M = 1.89\%$, $SD = 1.52\%$) produced larger start anomaly percentages than PINCH poses ($M = 1.35\%$, $SD = 1.06\%$). For surface type, post-hoc pairwise comparisons with Tukey correction revealed that VIRTUAL surfaces produced significantly larger anomalies than PHYSICAL surfaces ($p < 0.001$), while NOSURFACE did not differ significantly from either condition. Individual condition means ranged from $M = 1.25\%$ (PINCH, PHYSICAL) to $M = 2.81\%$ (POINT, NOSURFACE), see Figure 13.

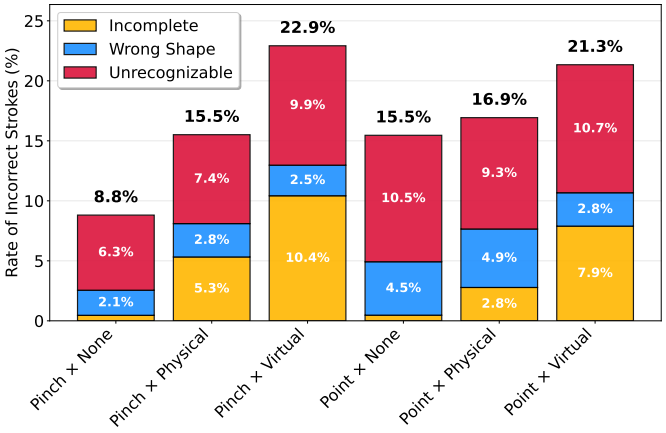


Figure 12: Incorrect stroke rates and distributions across conditions (pose × surface) in experiment 2.

We also ran a two-way within-subjects ANOVA on log-transformed end anomaly percentages. Both surface type and the pose × surface interaction violated sphericity assumptions, so Greenhouse-Geisser corrections were applied. The analysis revealed significant main effects for pose ($F_{1,17} = 23.80$, $p < 0.001$, $\eta^2_G = 0.095$) and a significant pose × surface interaction ($F_{2,34} = 4.86$, $p = 0.031$, $\eta^2_G = 0.060$), but not for surface type ($F_{2,34} = 3.83$, $p = 0.055$, $\eta^2_G = 0.045$).

The main effect of pose indicated that POINT poses ($M = 2.58\%$, $SD = 2.40\%$) produced significantly larger end anomaly percentages than PINCH poses ($M = 1.46\%$, $SD = 1.04\%$) (Figure 13). For interactions, POINT, NOSURFACE showed the highest end anomaly percentages ($M = 3.47\%$, $SD = 3.64\%$), differing significantly from several other conditions including PINCH, NOSURFACE ($p < 0.001$), PINCH, VIRTUAL ($p = 0.006$), and POINT, VIRTUAL ($p = 0.012$).

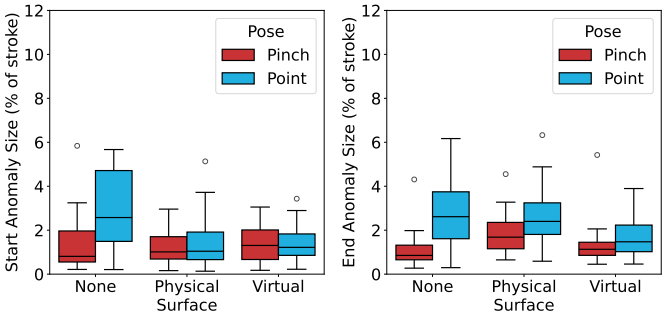


Figure 13: Box plots of start and end anomaly sizes across poses and surfaces in experiment 2.

Anomaly detection revealed that POINT poses consistently produced larger anomalies than PINCH poses at both stroke start and end anomalies. The POINT,NOSURFACE condition, again, produced the highest anomaly percentages for both start (2.81%) and end anomalies (3.47%).

4.2.4 Stroke Accuracy via Dynamic Time Warping

Our ART-ANOVA of DTW distances revealed significant main effects for pose ($F_{1,85} = 13.98$, $p < 0.001$, $\eta_G^2 = 0.119$), surface type ($F_{2,85} = 22.10$, $p < 0.001$, $\eta_G^2 = 0.300$), and pose × surface interaction ($F_{2,85} = 4.34$, $p = 0.016$, $\eta_G^2 = 0.078$).

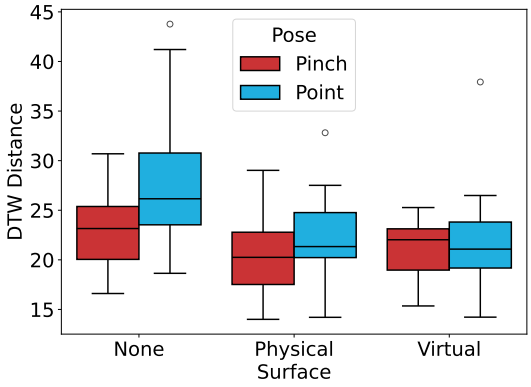


Figure 14: DTW distances from target strokes across poses and surfaces in experiment 2.

The main effect of pose indicated that POINT poses ($M = 23.84$, $SD = 6.01$) produced significantly higher DTW distances than PINCH poses ($M = 21.52$, $SD = 3.76$). For surface type, the NOSURFACE condition ($M = 25.35$, $SD = 5.90$) showed significantly higher DTW distances compared to both PHYSICAL ($M = 21.14$, $SD = 4.17$, $p < 0.001$) and VIRTUAL conditions ($M = 21.55$, $SD = 4.12$, $p < 0.001$), see Figure 14. No significant difference was found between PHYSICAL and VIRTUAL surface conditions ($p = 0.834$).

Interactions revealed that the effect of pose was dependent on surface condition. The largest difference occurred between POINT, NOSURFACE ($M = 27.60$, $SD = 6.70$) and PINCH, NOSURFACE conditions ($M = 23.10$, $SD = 3.99$; $p = 0.010$). Additionally, POINT poses showed significant improvement when performed on PHYSICAL ($M = 21.93$, $SD = 4.48$; $p < 0.001$) and VIRTUAL ($M = 21.99$, $SD = 5.03$; $p < 0.001$) surfaces compared to the NOSURFACE condition. The surface conditions showed more similar DTW distances for PINCH poses, though PHYSICAL surfaces still showed significantly lower distances than NOSURFACE ($p = 0.021$).

4.2.5 Size

We ran an ART-ANOVA on stroke 2D bounding box areas. The analysis revealed a significant main effect for surface type ($F_{2,85} = 16.77$, $p < 0.001$, $\eta_G^2 = 0.246$), but not for pose ($F_{1,85} = 0.004$, $p = 0.953$, $\eta_G^2 = 0.000$) or pose × surface interaction ($F_{2,85} = 0.14$, $p = 0.870$, $\eta_G^2 = 0.003$).

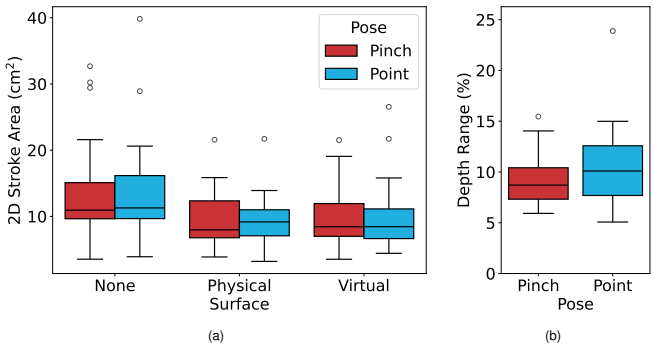


Figure 15: Dimensions of produced strokes in experiment 2 in terms of (a) Bounding box sizes across conditions (cm^2) and (b) Depth range across poses in NOSURFACE.

The main effect of surface indicated that stroke size varied significantly across surface conditions. The NOSURFACE condition produced significantly larger stroke areas ($M = 14.07$ cm^2, $SD = 8.54$ cm^2) compared to both PHYSICAL ($M = 9.59$ cm^2, $SD = 4.29$ cm^2; $p < 0.001$) and VIRTUAL conditions ($M = 10.15$ cm^2, $SD = 5.28$ cm^2; $p < 0.001$) (Figure 15a). No significant difference was found between PHYSICAL and VIRTUAL surface conditions ($p = 0.987$).

4.2.6 Depth Range

A one-way ANOVA revealed no significant effect of pose on depth range variation ($F_{1,17} = 3.48$, $p = 0.079$, $\eta_G^2 = 0.031$). While POINT poses showed numerically greater depth range variation ($M = 10.53\%$, $SD = 4.37\%$) compared to PINCH poses ($M = 9.26\%$, $SD = 2.70\%$), this difference was not statistically significant (Figure 15b). Overall, participants deviated on average 9.90% of their stroke extent in the depth dimension when drawing with NOSURFACE ($SD = 3.64\%$).

4.2.7 Average Speed

We ran an ART-ANOVA on average stroke speed after anomaly removal. No significant main effects were found for pose ($F_{1,85} = 1.72$, $p = 0.193$, $\eta_G^2 = 0.016$), surface type ($F_{2,85} = 2.22$, $p = 0.115$, $\eta_G^2 = 0.041$), or pose × surface interaction ($F_{2,85} = 1.25$, $p = 0.293$, $\eta_G^2 = 0.024$). Average drawing speeds ranged from 12.39 cm/s (PINCH, PHYSICAL) to 13.84 cm/s (PINCH, NOSURFACE), with overall means of 12.94 cm/s for PINCH poses and 13.32 cm/s for POINT poses.

5 DISCUSSION

To synthesize the results of these two studies, we provide a systematic understanding of single-stroke production on several algebraic and geometric metrics across PINCH and POINT poses and three surfaces NOSURFACE, PHYSICAL, and VIRTUAL. Several measures were consistent across both experiments. First, drawing with a PINCH pose generally resulted in better strokes than when using POINT. PINCH reliably produced smaller start and end anomalies, less variation in depth, and better stroke accuracy (via DTW). Similarly, NOSURFACE produced larger anomalies and worse accuracy (via DTW), but upon visual inspection of Figures 7, 13, 8, and 14, this effect appears primarily driven by poor performance when pairing POINT with NOSURFACE.

In terms of objective DTW distance, both surface conditions (PHYSICAL and VIRTUAL) were found to improve performance when drawing with a POINT pose. Though not causing significant issues with DTW distance for strokes coded as correct, VIRTUAL surfaces consistently produced significantly higher incorrectly coded strokes than PHYSICAL or NOSURFACE, especially in terms of incomplete strokes—where the user unintentionally ended the stroke before completion. Finally, despite only considering a 2D bounding box, NOSURFACE strokes were consistently larger than both PHYSICAL and VIRTUAL.

Despite obvious performance differences, participant preference and subjective experience was more nuanced and largely dependent on the combination of pose and surface. Overall pose and surface preferences split, with two participants noted that their preference depended on the surface condition. Interestingly, NOSURFACE was the most preferred surface despite showing poorer objective performance when using POINT, though the consistently good performance of PINCH in this condition may have contributed to its overall preference. Subjective workload measures from the NASA TLX aligned more closely with objective findings: PINCH resulted in lower mental and physical demand than POINT, while VIRTUAL surfaces required significantly more effort and had worse perceived performance than other surface conditions.

A difference that did emerge across experiments was that start anomalies were larger in the NOSURFACE condition in experiment 1, but did not reach statistical significance in experiment 2, likely reflecting the high variability within that condition. That being said, when examin-

ing Figure 13, the POINT, NOSURFACE condition appears noticeably higher than the others.

Comparing the two experiments, experiment 2 which used more common single strokes (i.e. figure-8, square, triangle, circle, zig-zag, pigtail, star, wave), demonstrated substantial improvements over the abstract strokes adapted from Tu et al.'s study [43]. Overall rates of incorrect strokes decreased from 22.37% in experiment 1 to 16.83% to experiment 2, and DTW distance improved with distances ranging from 27.88 to 32.86 in experiment 1 compared to 21.14 to 27.60 in experiment 2. The improvements in experiment 2 suggest that errors and stroke accuracy in the first experiment may have resulted from participants misunderstanding the more abstract strokes. However, trends across both studies remain largely consistent, indicating the robustness of our findings.

5.1 Revisiting our Research Questions

Taking these findings into consideration, we address our original research questions as follows.

RQ1: How do pinch and point hand poses compare for drawing single strokes in VR? PINCH poses consistently outperformed POINT poses across both experiments on most objective measures. We expect that this is due to two explanations. First, pinch offers a more explicit natural delimiter and kinesthetic feedback, whereas point gestures are harder to delimit as precisely [17,45]. Second, the pinch pose resembles the grip used when holding a pen, the traditional method for drawing. If we consider the pinch pose representative of holding a pen, we can make some comparisons to Tu et al.'s [43] results by mapping pen to pinch and point to finger. Tu et al. found that pen outperformed finger on several measures, including corner shape distance and intersecting points deviation, and users in their study perceived the pen as more accurate than finger for drawing. Our results mirror this pattern, with PINCH similarly outperforming POINT. However, unlike this prior work, we did not find significant differences between poses on produced stroke size and speed, suggesting that while the two studies converge on accuracy-related measures, the bare-hand and spatial contexts may introduce some differences.

RQ2: How does the presence of a virtual, physical, or no surface compare for single-stroke drawing?
Surface effects were more complex than pose effects. VIRTUAL surfaces consistently produced the highest incorrect strokes and required the most effort, while PHYSICAL and NOSURFACE conditions performed similarly in most metrics. This difference may stem from the varying levels of feedback each surface provides: PHYSICAL surfaces offer passive haptic feedback to support touch, while VIRTUAL surfaces lack sufficient feedback to indicate surface intersection, leading users to accidentally leave the drawing plane. Notably, NOSURFACE produced larger stroke sizes and worse accuracy, but was the most preferred by participants, likely due to the consistently good performance of PINCH in this condition. These preference findings partially differ from prior work [12,29]. This may be explained by differences in both the task and the interaction design. Our study focused on single-stroke production requiring continuous precision, whereas Cheng et al. [12] examined more discrete input such as docking and selection. In addition, unlike Liu et al. [29], who compared two different interaction techniques, we manipulated hand pose and surface independently. This suggests that some of their findings may reflect their interaction techniques as a whole rather than the physical surface itself.

RQ3: How do surfaces and poses interact with one another for single-stroke drawing?
As reported in prior work on discrete input tasks [12], the most critical finding was that POINT poses benefited significantly from surface presence, showing improved DTW distance on both PHYSICAL and VIRTUAL surfaces compared to NOSURFACE. In contrast, PINCH performance remained more consistent across all surface conditions. The POINT, NOSURFACE combination was particularly problematic, for reasons discussed above, producing the largest anomalies and worst accuracy, while PINCH poses maintained stable performance regardless of surface type.

5.2 Limitations and Future Work

Our work has several limitations to be considered. First, the Meta Quest 3's spatial mapping [35] does not have submillimeter precision as other motion capture systems (e.g. OptiTrack or Vicon). While our results indicate the physical surface mapping worked well overall, it may not have been precise to the millimeter level, potentially causing some stroke anomalies.

Second, the generalizability of our findings is limited by both the experimental tasks and participant population. While the single strokes are representative of simple doodles, they may not be representative of typical drawing or sketching, which usually consist of multiple strokes. Future work should investigate multi-stroke drawing tasks to better understand how pose and surface effects scale with drawing complexity and whether the advantages of PINCH poses and preference for NOSURFACE persist in more realistic drawing scenarios. In addition, although recruitment was conducted through a university-wide mailing list, our participant sample consisted primarily of university students, which may not represent broader user populations with different levels of drawing or XR experience.

Third, our accuracy measures were primarily objective, relying on metrics such as DTW distances and geometric properties. Though manual annotation was a component of error tagging and detecting anomalies, drawing often requires subjective evaluation of aesthetic quality that our quantitative measures do not address. For instance, we did not ask participants to rate their drawing quality nor did we allow them to retry strokes. Future work should consider a combination of objective and more subjective approaches for a more complete understanding of drawing quality in VR.

Fourth, we compared drawing on upright surfaces to limit the number of conditions to consider. However, another area for future work would be to explore drawing on surfaces with different alignments, such as flat horizontal surfaces, or angled surfaces that more closely resemble traditional easels. Future work could examine whether surface orientation affects bare-hand drawing performance in ways similar to the controller and pen-based interactions found in prior work [3].

Finally, our work focused on hand pose as a delimiter. An interesting direction for future work would be to combine hand pose with other modalities, such as gaze, when determining intent. Similar combinations of gaze and hand input have been effective for XR interaction [38] and may help reduce stroke anomalies, particularly for point-based interaction.

6 CONCLUSION

We investigated single-stroke production in VR using different hand poses and surfaces. PINCH poses outperformed POINT poses across most metrics, though surfaces helped reduce these effects. Surface influence was more complex, with VIRTUAL surfaces consistently underperforming due to poor feedback and NOSURFACE with POINT reducing overall accuracy. However, when pairing PINCH and NOSURFACE, users were able to achieve accurate strokes, even when drawing 2D shapes in 3D, demonstrating the feasibility of drawing without a surface. For VR drawing interfaces, POINT poses work well when surfaces are available, but PINCH poses provide more consistent performance across all conditions.

REFERENCES

[1] Apple Inc. Apple vision pro. https://www.apple.com/apple-vision-pro/, 2024. 1, 3

[2] R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, and K. Singh. Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–15. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173759 2

[3] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. Fitzmaurice. Experimental evaluation of sketching on surfaces in vr. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 5643–5654. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025474 1, 2, 9

[4] R. Balakrishnan and I. S. MacKenzie. Performance differences in the fingers, wrist, and forearm in computer input control. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pp. 303–310. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/258549.258764 2

[5] A. Banerjee, J. Burstyn, A. Girouard, and R. Vertegaal. Pointable: an in-air pointing technique to manipulate out-of-reach targets on tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pp. 11–20. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/2076354.2076357 1, 2, 3

[6] M. D. Barrera Machuca, W. Stuerzlinger, and P. Asente. The effect of spatial ability on immersive 3d drawing. In *Proceedings of the 2019 Conference on Creativity and Cognition*, C&C '19, pp. 173–186. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3325480.3325489 2

[7] B. Bowers, A. Tuttle, A. Rukangu, A. Franzluebbers, C. Ball, and K. Johnsen. Comparing virtual constraints and a physical stylus for planar writing and drawing in virtual reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 220–225, 2021. doi: 10.1109/VRW52623.2021.00048 2

[8] A. Cai, C. Yao, H. Liu, C. Li, Y. Zhang, Y. Guo et al. Hpipainting: A hand-pen interaction for vr painting. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 9(3), art. no. 71, 26 pp., Sept. 2025. doi: 10.1145/3749538 2

[9] X. Cao and S. Zhai. Modeling human performance of pen stroke gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pp. 1495–1504. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1240624.1240850 2

[10] S. J. Castellucci and I. S. MacKenzie. Graffiti vs. unistrokes: an empirical comparison. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pp. 305–308. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357106 2

[11] C. Chen, M. Yarmand, Z. Xu, V. Singh, Y. Zhang, and N. Weibel. Investigating input modality and task geometry on precision-first 3d drawing in virtual reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 384–393, 2022. doi: 10.1109/ISMAR55827.2022.00054 2

[12] Y. F. Cheng, T. Luong, A. R. Fender, P. Streli, and C. Holz. Comfortable user interfaces: Surfaces reduce input error, time, and exertion for tabletop and mid-air user interfaces. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 150–159, 2022. doi: 10.1109/ISMAR55827.2022.00029 2, 9

[13] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio. Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376628 2

[14] H. Elsayed, M. D. Barrera Machuca, C. Schaarschmidt, K. Marky, F. Müller, J. Riemann et al. Vrsketchpen: Unconstrained haptic assistance for sketching in virtual 3d environments. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*, VRST '20, art. no. 3, 11 pp. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385956.3418953 2

[15] N. Fourrier, G. Moreau, M. Benaouicha, and J.-M. Norm. Handwriting for efficient text entry in industrial vr applications: Influence of board orientation and sensory feedback on performance. *IEEE Transactions on Visualization and Computer Graphics*, 29(11):4438–4448, oct 2023. doi: 10.1109/TVCG.2023.3320215 2

[16] T. Giorgino. Computing and visualizing dynamic time warping alignments in r: The dtw package. *Journal of Statistical Software*, 31(7):1–24, 2009. doi: 10.18637/jss.v031.i07 4, 5

[17] T. Grossman, D. Wigdor, and R. Balakrishnan. Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pp. 61–70, 2004. 2, 9

[18] A. Hafizi, J. Henderson, A. Neshati, W. Zhou, E. Lank, and D. Vogel. In-vehicle performance and distraction for midair and touch directional gestures. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, art. no. 316, 13 pp. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581335 2

[19] S. Hart. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human mental workload/Elsevier*, 1988. 3, 6

[20] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, pp. 213–222. Association for Computing Machinery, New York, NY, USA, 1994. doi: 10.1145/192426.192501 2

[21] M. R. Jakobsen, Y. Jansen, S. Boring, and K. Hornbæk. Should i stay or should i go? selecting between touch and mid-air gestures for large-display interaction. In *Human-Computer Interaction – INTERACT 2015*, pp. 455–473. Springer-Verlag, Berlin, Heidelberg, 2015. doi: 10.1007/978-3-319-22698-9_31 1, 2, 3

[22] Y. Jiang, C. Zhang, H. Fu, A. Cannavò, F. Lamberti, H. Y. K. Lau et al. Handpainter - 3d sketching in vr with hand-based physical proxy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, art. no. 412, 13 pp. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445302 2

[23] A. Jude, G. M. Poor, and D. Guinness. Grasp, grab or pinch? identifying user preference for in-air gestural manipulation. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, p. 219. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2983310.2989209 1, 2, 3

[24] M. Kay, L. A. Elkin, J. J. Higgins, and J. O. Wobbrock. mjskay/artool: Artool 0.11.0, 2021. doi: 10.5281/zenodo.4721941 4

[25] D. Keefe, R. Zeleznik, and D. Laidlaw. Drawing on air: Input techniques for controlled 3d line illustration. *IEEE Transactions on Visualization and Computer Graphics*, 13(5):1067–1081, 2007. doi: 10.1109/TVCG.2007.1060 2

[26] D. F. Keefe, D. A. Feliz, T. Moscovich, D. H. Laidlaw, and J. J. LaViola. Cavepainting: a fully immersive 3d artistic medium and interactive experience. In *Proceedings of the 2001 Symposium on Interactive 3D Graphics*, I3D '01, pp. 85–93. Association for Computing Machinery, New York, NY, USA, 2001. doi: 10.1145/364338.364370 2

[27] F. Kern, J. Tschanter, and M. E. Latoschik. Handwriting for text input and the impact of xr displays, surface alignments, and sentence complexities. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[28] R. W. Lindeman, J. L. Sibert, and J. K. Hahn. Towards usable vr: an empirical study of user interfaces for immersive virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 64–71, 1999. 1, 2

[29] G. Liu, H. Su, J. Wang, Q. Qi, H. Sun, Z. Zhuang et al. Towards bare-hand interaction for whiteboard collaboration in virtual reality. *Proc. ACM Hum.-Comput. Interact.*, 9(2), art. no. CSCW194, 31 pp., May 2025. doi: 10.1145/3711092 2, 9

[30] M. Liu, M. Nancel, and D. Vogel. Gunslinger: Subtle arms-down mid-air interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pp. 63–71. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2807442.2807489 1, 2, 3

[31] M. D. B. Machuca, P. Asente, W. Stuerzlinger, J. Lu, and B. Kim. Multiplanes: Assisted freehand vr sketching. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*, SUI '18, pp. 36–47. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3267782.3267786 2

[32] M. D. B. Machuca, W. Stuerzlinger, and P. Asente. Smart3dguides: Making unconstrained immersive 3d drawing more accurate. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, art. no. 37, 13 pp. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359996.3364254 2

[33] A. Markussen, M. R. Jakobsen, and K. Hornbæk. Vulture: a mid-air word-gesture keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 1073–1082. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2556964 1, 2, 3

[34] Meta. *Meta XR All-in-One SDK — Version 66.0*. Meta, 2024. 3

[35] Meta. *Mixed Reality Utility Kit Overview*. Meta, 2025. 3, 9

[36] Meta Platforms. *Hand Tracking Interactions in Unity*, 2023. 3

[37] Meta Platforms, Inc. Meta quest. `https://www.meta.com/quest/`, 2024. 1

[38] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, pp. 99–108. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132180 9

[39] A. S. Raina, V. Mone, M. Gorjian, S. Sueda, F. Quek, and V. R. Kr-

ishnamurthy. Blended physical-digital kinesthetic feedback for mixed reality-based conceptual design-in-context. In *Graphics Interface 2024 Second Deadline*, 2024. 2

[40] D. Schneider, V. Biener, A. Otte, T. Gesslein, P. Gagel, C. Campos et al. Accuracy evaluation of touch tasks in commodity virtual and augmented reality head-mounted displays. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, SUI '21, art. no. 7, 11 pp. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3485279.3485283 2

[41] H. B. Surale, F. Matulic, and D. Vogel. Experimental analysis of barehand mid-air mode-switching techniques in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300426 1, 2, 3

[42] R. J. Teather, D. Natapov, and M. Jenkin. Evaluating haptic feedback in virtual environments using iso 9241–9. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 307–308, 2010. doi: 10.1109/VR.2010.5444753 2

[43] H. Tu, X. Ren, and S. Zhai. A comparative evaluation of finger and pen stroke gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 1287–1296. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208584 2, 3, 4, 6, 9

[44] R. Venkatakrishnan, R. Venkatakrishnan, C.-H. Chung, Y.-S. Wang, and S. Babu. Investigating a combination of input modalities, canvas geometries, and inking triggers on on-air handwriting in virtual reality. *ACM Trans. Appl. Percept.*, 19(4), art. no. 15, 19 pp., nov 2022. doi: 10.1145/3560817 2

[45] D. Vogel and R. Balakrishnan. 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, pp. 33–42. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1095034.1095041 1, 2, 3, 9

[46] P. Wacker, A. Wagner, S. Voelker, and J. Borchers. Physical guides: An analysis of 3d sketching performance on physical objects in augmented reality. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*, SUI '18, pp. 25–35. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3267782.3267788 2

[47] Y. Wang and C. L. MacKenzie. The role of contextual haptic and visual constraints on object manipulation in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pp. 532–539. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/332040.332494 1, 2

[48] E. Wiese, J. H. Israel, A. Meyer, and S. Bongartz. Investigating the learnability of immersive free-hand sketching. In *Proceedings of the Seventh Sketch-Based Interfaces and Modeling Symposium*, SBIM '10, pp. 135–142. Eurographics Association, Goslar, DEU, 2010. 2

[49] A. D. Wilson. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pp. 255–258. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1166253.1166292 1, 2

[50] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. 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, pp. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 4

[51] J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pp. 159–168. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1294211.1294238 2, 6

[52] X. Xu, Y. Zhou, B. Shao, G. Feng, and C. Yu. Gesturesurface: Vr sketching through assembling scaffold surface with non-dominant hand. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2499–2507, 2023. doi: 10.1109/TVCG.2023.3247059 2

[53] X. Yi, X. Wang, J. Li, and H. Li. Examining the fine motor control ability of linear hand movement in virtual reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 427–437. IEEE, 2023. 2

[54] D. Zielasko, M. Krüger, B. Weyers, and T. W. Kuhlen. Passive haptic menus for desk-based and hmd-projected virtual reality. In *2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–6, 2019. doi: 10.1109/WEVR.2019.8809589 2